



## روش تحقیق در تعلیم و تربیت

هسته ۱۸۷

عذر ما هسته ۱۸۹

نوعی از همه نمونه‌ها یکی حاصله تا ۱۱۹۶  
 از نمونه‌های - حاصله تا ۲۱۹ - ۲۲۰ - ۲۲۱ - ۲۲۲ - ۲۲۳ - ۲۲۴ - ۲۲۵ - ۲۲۶ - ۲۲۷ - ۲۲۸ - ۲۲۹ - ۲۳۰  
 انواع نمونه‌های خود را از ۲۲۲ - ۲۲۸  
 انواع آبی از کلام هند - طبع  
 نمونه‌های غیر از کلامی | آبی - هند - هند - هند - ۲۲۹ - ۲۳۰

اندازه نمونه ۲۳۲

ظرف نمونه ۲۳۳ - ۲۳۶

آزمونه‌ها ۲۳۹

از نوع ۲۵۰

در سطح اندازه‌های دیگر که در این قسمت ۲۹۳ :  
 ۱ - آزمونها - آزمون برنت ه - ۱۹۴ - آزمون ۱۹۷  
 اندازه‌ها کیفیت : برنت ه در دو بار زدن ۲۹۹ - ۳۰۱  
 ۲ - شماره برگزین ۳۰۹ (شماره همین با آزمون)

(بسیار بکرت (مجلس شماره ۲۷) - شماره برگزین ۳۱۳  
 شماره برگزین ۳۱۰ - شماره برگزین ۳۱۲ - شماره برگزین ۳۱۳

۳ - در دو بار برگزین ۳۱۷

۴ - شماره برگزین ۳۱۸ - شماره برگزین ۳۲۱

کتابخانه از نامی دیگر در شماره برگزین ۳۱۳

کتابخانه علی - شماره برگزین ۴۹۲

کتابخانه حسن ۴۹۳

۱ - مطالبات رسد ۴۹۸ - روش طرح - روش منظمی

۲ - مطالبات شهر ۴۹۸

۳ - مجلس انبار ۴۹۹

۴ - مجلس روتر ۵۰۰

۵ - بر سر راه هسته ۵۰۱

کتابخانه با نامی ۵۲۹ - سر راه نمونه ای

کتابخانه شهر کتابت با نامی ۵۳۱

شبهه شهر ۵۳۸ - ۵۳۷

نصفه ۵۴۴

برشته ۵۴۸ - و کلامی برشته ۵۵۲

انواع سوال ۵۵۳

اعتبار - ۵۵۴ - و مجلس آبر و کتابت ۵۶۷

کتابخانه علی و کتابخانه ۵۸۰

کتابخانه استی ۵۸۷ - ۵۹۲

۱۳۰۳۳

۸۰۱۴۱۱۰



# روش تحقیق در تعلیم و تربیت

نویسندگان:

دانا آری، لوسی چسر جیکوبز، اصغر رضویه

مترجمان:

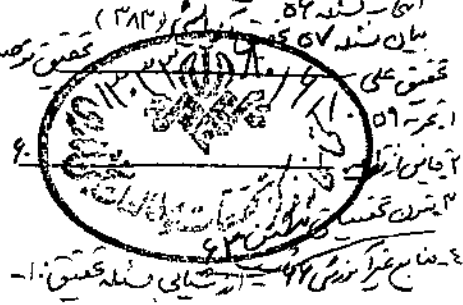
وازگن سرکیسیان، مینو نیکو، ایما سعیدیان

سروش  
تهران ۱۳۸۰

تطبیق نظریه‌ها ۱۱  
 با اکتشافات نظریه‌ها ۱۶  
 مراحل تحقیق ۳  
 مسائل نظریه‌ها ۳۶  
 سوالات کارورس ۳۷ - تحقیق سنادین تحقیق کارورس  
 نهم ۳۹  
 کورسنگر - نویسه‌نگاری و کاربرد آن

Introduction to  
**Research in Education**  
 Donald Ary, Lucy Cheser Jacobs, Asghar Razavieh  
 Forth Edition, 1990

تفسیرها ۴۲  
 انواع تفسیرها ۴۳ - مقوله‌های بیرونی  
 تفسیر فعال و غیرفعال ۴۵  
 روش آریاس - علی‌حده‌ان - روشی (روش تجزیه کلان - استاکر - همستر - ۴۶ - ۴۹۳  
 روش کیفی - تاریخی ۴۷  
 اتکا - بنده ۵۲  
 بیان بنده ۵۷  
 تحقیق علی - ۵۶  
 ۵۱ - ۵۲  
 ۲ - بیان بنده ۵۸  
 ۳ - بیان بنده ۵۸  
 ۴ - بیان بنده ۵۸  
 ۵ - بیان بنده ۵۸  
 ۶ - بیان بنده ۵۸  
 ۷ - بیان بنده ۵۸  
 ۸ - بیان بنده ۵۸  
 ۹ - بیان بنده ۵۸  
 ۱۰ - بیان بنده ۵۸  
 ۱۱ - بیان بنده ۵۸  
 ۱۲ - بیان بنده ۵۸  
 ۱۳ - بیان بنده ۵۸  
 ۱۴ - بیان بنده ۵۸  
 ۱۵ - بیان بنده ۵۸  
 ۱۶ - بیان بنده ۵۸  
 ۱۷ - بیان بنده ۵۸  
 ۱۸ - بیان بنده ۵۸  
 ۱۹ - بیان بنده ۵۸  
 ۲۰ - بیان بنده ۵۸  
 ۲۱ - بیان بنده ۵۸  
 ۲۲ - بیان بنده ۵۸  
 ۲۳ - بیان بنده ۵۸  
 ۲۴ - بیان بنده ۵۸  
 ۲۵ - بیان بنده ۵۸  
 ۲۶ - بیان بنده ۵۸  
 ۲۷ - بیان بنده ۵۸  
 ۲۸ - بیان بنده ۵۸  
 ۲۹ - بیان بنده ۵۸  
 ۳۰ - بیان بنده ۵۸  
 ۳۱ - بیان بنده ۵۸  
 ۳۲ - بیان بنده ۵۸  
 ۳۳ - بیان بنده ۵۸  
 ۳۴ - بیان بنده ۵۸  
 ۳۵ - بیان بنده ۵۸  
 ۳۶ - بیان بنده ۵۸  
 ۳۷ - بیان بنده ۵۸  
 ۳۸ - بیان بنده ۵۸  
 ۳۹ - بیان بنده ۵۸  
 ۴۰ - بیان بنده ۵۸  
 ۴۱ - بیان بنده ۵۸  
 ۴۲ - بیان بنده ۵۸  
 ۴۳ - بیان بنده ۵۸  
 ۴۴ - بیان بنده ۵۸  
 ۴۵ - بیان بنده ۵۸  
 ۴۶ - بیان بنده ۵۸  
 ۴۷ - بیان بنده ۵۸  
 ۴۸ - بیان بنده ۵۸  
 ۴۹ - بیان بنده ۵۸  
 ۵۰ - بیان بنده ۵۸  
 ۵۱ - بیان بنده ۵۸  
 ۵۲ - بیان بنده ۵۸  
 ۵۳ - بیان بنده ۵۸  
 ۵۴ - بیان بنده ۵۸  
 ۵۵ - بیان بنده ۵۸  
 ۵۶ - بیان بنده ۵۸  
 ۵۷ - بیان بنده ۵۸  
 ۵۸ - بیان بنده ۵۸  
 ۵۹ - بیان بنده ۵۸  
 ۶۰ - بیان بنده ۵۸  
 ۶۱ - بیان بنده ۵۸  
 ۶۲ - بیان بنده ۵۸  
 ۶۳ - بیان بنده ۵۸  
 ۶۴ - بیان بنده ۵۸  
 ۶۵ - بیان بنده ۵۸  
 ۶۶ - بیان بنده ۵۸  
 ۶۷ - بیان بنده ۵۸  
 ۶۸ - بیان بنده ۵۸  
 ۶۹ - بیان بنده ۵۸  
 ۷۰ - بیان بنده ۵۸  
 ۷۱ - بیان بنده ۵۸  
 ۷۲ - بیان بنده ۵۸  
 ۷۳ - بیان بنده ۵۸  
 ۷۴ - بیان بنده ۵۸  
 ۷۵ - بیان بنده ۵۸  
 ۷۶ - بیان بنده ۵۸  
 ۷۷ - بیان بنده ۵۸  
 ۷۸ - بیان بنده ۵۸  
 ۷۹ - بیان بنده ۵۸  
 ۸۰ - بیان بنده ۵۸  
 ۸۱ - بیان بنده ۵۸  
 ۸۲ - بیان بنده ۵۸  
 ۸۳ - بیان بنده ۵۸  
 ۸۴ - بیان بنده ۵۸  
 ۸۵ - بیان بنده ۵۸  
 ۸۶ - بیان بنده ۵۸  
 ۸۷ - بیان بنده ۵۸  
 ۸۸ - بیان بنده ۵۸  
 ۸۹ - بیان بنده ۵۸  
 ۹۰ - بیان بنده ۵۸  
 ۹۱ - بیان بنده ۵۸  
 ۹۲ - بیان بنده ۵۸  
 ۹۳ - بیان بنده ۵۸  
 ۹۴ - بیان بنده ۵۸  
 ۹۵ - بیان بنده ۵۸  
 ۹۶ - بیان بنده ۵۸  
 ۹۷ - بیان بنده ۵۸  
 ۹۸ - بیان بنده ۵۸  
 ۹۹ - بیان بنده ۵۸  
 ۱۰۰ - بیان بنده ۵۸



مرکز تحقیقات،  
 مطالعات و منسجم برنامه‌های  
 سازمان اسناد و کتابخانه ملی ایران

روش تحقیق در تعلیم و تربیت

نویسندگان:

دانلد آری، لوسی چسر جیکوبز، اصغر رضویه

مترجمان:

واژمن سرکیسیان، مینو نیکو، ایما سعیدیان

چاپ اول: ۱۳۸۰

تیراژ: ۲۰۰۰ نسخه

حرف‌فجینی: خدمات رایانه‌ای مرکز تحقیقات

قیمت: ۳۵۰۰ تومان

کلیه حقوق این اثر متعلق به مرکز تحقیقات می‌باشد

ISBN 964 - 92 554 - 6 - x ۹۶۴ - ۹۲۵۵۴ - ۶ - x

۱ - مقایسه‌ها اندازه‌گیری ۱۵۴ (مقاله‌های) برهمه - نا صمدی - ۱۶۱ - ۱۵۶  
 ۲ - سازمان‌ها هر دو در هر دو ۱۶۱  
 ۳ - انسان هر دو در هر دو ۱۶۳  
 ۴ - بیان - بیان ۱۶۴ - ۱۷۱  
 ۵ - بیان میانگین نازید توزیع در ۱۷۱  
 ۶ - بیان - بیان ۱۸۱  
 ۷ - بیان - بیان ۱۸۵

## فهرست مطالب

صفحه	عنوان
۱	پیشگفتار
۵	بخش اول: مبانی تحقیق
۵	فصل اول: روش علمی در تعلیم و تربیت
۶	منابع علم تجربی - مرجع بطالع - استدلال قیاسی - استدلال استقرایی - روش علمی - اسناد ارزشی
۱۹	ماهیت علم نظریه علم - اهدای نظریه - روش‌های نظریه‌پردازی - دیدی کلان‌رویی استقرایی - روش‌های نظریه‌پردازی
۳۱	ماهیت تحقیق و عمل علمی - ماهانه فرضیه - گردآمدن راه، استنتاج
۳۸	زبان تحقیق علمی
۴۶	روش‌شناسی تحقیق در تعلیم و تربیت - آزادی - سه‌ای - تدریسی
۴۷	خلاصه
۴۹	تمرینها
۵۲	پاسخها
۵۵	بخش دوم: زمینه تحقیق
۵۵	فصل دوم: موضوع تحقیق
۵۷	ماهیت مسائل
۵۸	منابع مسائل و موضوعات
۶۶	ارزشیابی مسئله تحقیق
۷۰	بیان مسئله
۷۲	تعیین جامعه مورد بررسی و متغیرها
۸۲	خلاصه
۸۲	تمرینها
۸۴	پاسخها
۸۷	فصل سوم: تحقیقات پیشین
۸۸	نقش تحقیقات پیشین در یک پروژه تحقیقی

۱۳۸۰  
 ۹  
 ۱۳۴  
 ۱۰۲۸  
 ۱۳۳

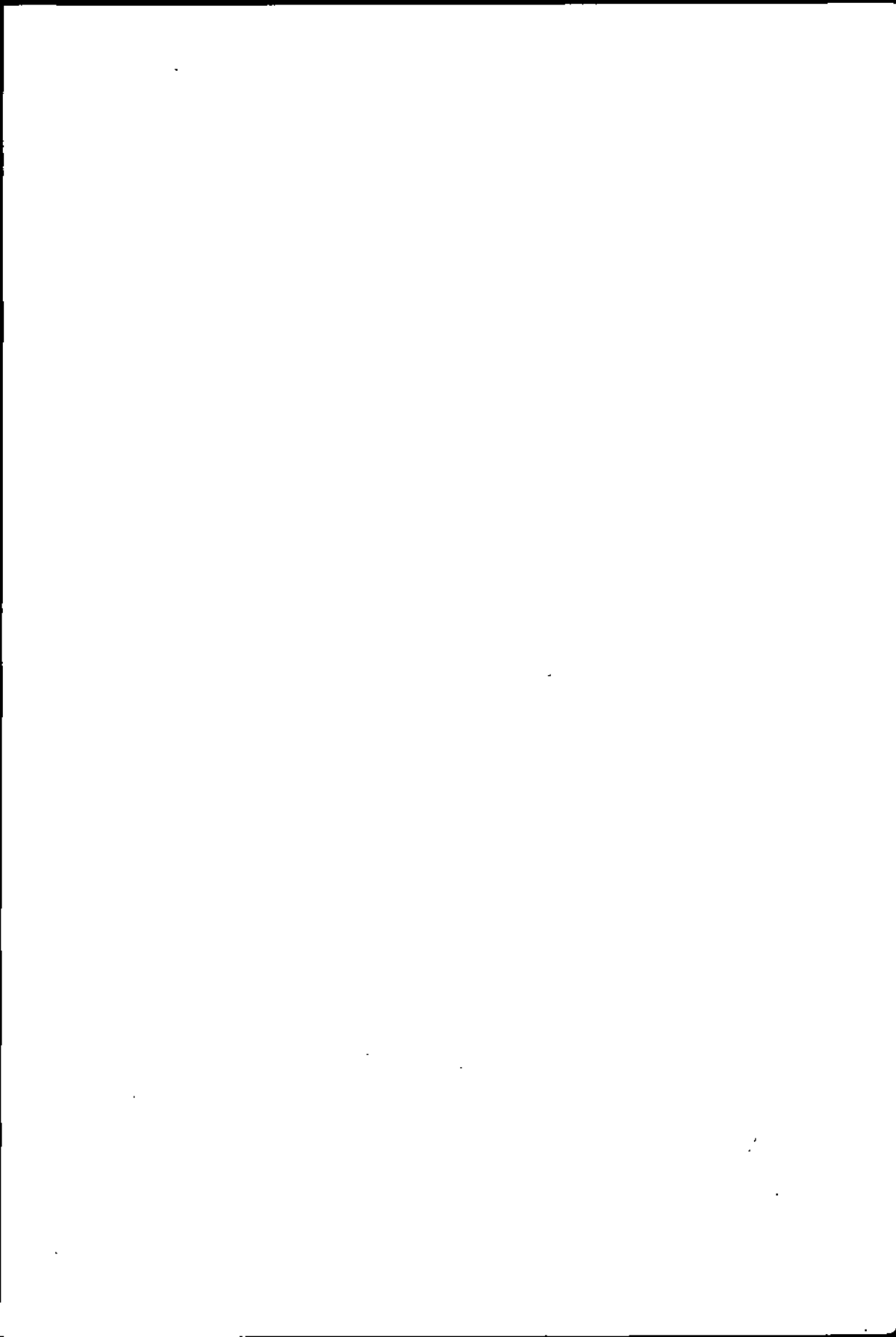
۹۰	منابع مرجع در تعلیم و تربیت
۱۱۵	سازماندهی تحقیقات پیشین
۱۱۹	خلاصه
۱۱۹	تمرینها
۱۲۰	پاسخها
۱۲۳	فصل چهارم: فرضیه
۱۲۷	پیشنهادهایی برای استنتاج فرضیه‌ها
۱۳۲	ویژگیهای فرضیه‌های قابل استفاده
۱۴۰	انواع فرضیه
۱۴۲	آزمون فرضیه
۱۴۴	برنامه تحقیق
۱۴۶	خلاصه
۱۴۷	تمرینها
۱۴۹	پاسخها
۱۵۳	بخش سوم: تحلیل آماری
۱۵۳	فصل پنجم: آمار توصیفی
۱۵۵	مقیاسهای اندازه‌گیری
۱۶۱	سازماندهی داده‌های پژوهش
۱۶۳	اندازه‌های گرایش مرکزی
۱۷۵	اندازه‌های تغییرپذیری
۱۸۱	نمره‌های استاندارد
۱۸۵	منحنی نرمال
۱۸۷	همبستگی
۱۸۹	ضرریبهای همبستگی
۲۰۵	فرا تحلیل
۲۰۶	خلاصه

۲۰۹	تمرینها
۲۱۳	پاسخها
۲۱۷	فصل ششم: نمونه گیری و آمار استنباطی
۲۱۹	نمونه گیری
۲۳۹	استراتژی آمار استنباطی
۲۴۲	سطح معنی داری
۲۴۵	آزمون t
۲۵۹	تحلیل واریانس
۲۷۳	آزمون معنی داری خی دو
۲۷۹	خلاصه
۲۸۱	تمرینها
۲۸۷	پاسخها
۲۹۱	بخش چهارم: مبانی اندازه گیری
۲۹۱	فصل هفتم: ابزار تحقیق
۲۹۴	وسایل اندازه گیری
۳۲۳	خلاصه
۳۲۴	تمرینها
۳۲۶	پاسخها
۳۲۷	فصل هشتم: اعتبار و پایایی
۳۲۹	اعتبار
۳۴۶	پایایی
۳۶۹	خلاصه
۳۷۱	تمرینها
۳۷۵	پاسخها
۳۸۱	بخش پنجم: روشهای تحقیق
۳۸۱	فصل نهم: تحقیق آزمایشی در تعلیم و تربیت
۳۸۳	ویژگیهای تحقیق آزمایشی

۳۹۵	مقایسه آزمایشی
۳۹۷	طرح آزمایشی
۴۱۲	طبقه‌بندی طرحهای آزمایشی
۴۴۹	خلاصه
۴۵۱	تمرینها
۴۵۳	پاسخها
۴۵۵	فصل دهم: تحقیقات علی - مقایسه‌ای
۴۵۷	مقایسه روش تحقیق علی - مقایسه‌ای با روش آزمایشی
۴۶۰	شرایط لازم برای استنباط روابط علی
۴۶۸	کنترل محدود در تحقیق علی - مقایسه‌ای
۴۷۹	طرح‌ریزی یک بررسی علی - مقایسه‌ای
۴۸۲	نقش تحقیق علی - مقایسه‌ای
۴۸۵	خلاصه
۴۸۶	تمرینها
۴۸۷	پاسخها
۴۹۱	فصل یازدهم: تحقیق توصیفی
۴۹۳	مراحل تحقیق توصیفی
۵۲۱	خلاصه
۵۲۲	تمرینها
۵۲۴	پاسخها
۵۲۷	فصل دوازدهم: تحقیقات پیمایشی
۵۳۰	سرشماریها و تحقیقات پیمایشی نمونه‌ای
۵۳۵	شیوه پیمایش
۵۴۴	ابزار گردآوری داده‌ها
۵۶۵	اعتبار
۵۶۶	پایایی
۵۶۷	تحلیل آماری در تحقیقات پیمایشی



۵۷۱	خلاصه
۵۷۳	تمرینها
۵۷۵	پاسخها
۵۷۹	فصل سیزدهم: تحقیقات کیفی و تاریخی
۵۸۰	بررسی کیفی
۵۹۲	تحقیق تاریخی
۵۹۵	خلاصه
۵۹۶	تمرینها
۵۹۷	پاسخها
۵۹۹	بخش ششم : تدوین گزارش تحقیق
۵۹۹	فصل چهاردهم : رهنمودهایی برای نگارش طرح پیشنهادی تحقیق
۶۰۰	نگارش یک طرح تحقیق
۶۱۶	نقد طرحهای پیشنهادی
۶۱۸	اهمیت تکمیل طرح پیشنهادی تحقیق قبل از گردآوری داده‌ها
۶۱۹	ملاحظات اخلاقی و حقوقی
۶۲۴	خلاصه
۶۲۶	تمرینها
۶۲۷	پاسخها
۶۲۹	فصل پانزدهم: تحلیل، تفسیر و گزارش نتایج
۶۳۰	پردازش داده‌ها
۶۳۳	اصول تفسیر
۶۳۹	گزارش تحقیق
۶۵۰	راهنماهای نگارش
۶۵۰	خلاصه
۶۵۱	تمرینها
۶۵۲	پاسخها
۶۵۵	



### /// پیشگفتار

در سال ۱۹۷۲، هدف ما از تدارک اولین چاپ کتاب «مقدمه‌ای بر تحقیق در تعلیم و تربیت»، تهیه کتابی بود که خوانندگان را قادر سازد تا به قابلیت‌های اولیه و ضروری برای (۱) درک و ارزشیابی تحقیقاتی که دیگران انجام داده‌اند و (۲) طراحی و اجرای تحقیق خود با حداقل مساعدت، دست یابند. تأیید مداوم این کتاب طی سه چاپ مکرر، نشان می‌دهد که ما در این امر موفق بوده‌ایم.

با تکوین و تکامل تحقیق در تعلیم و تربیت، ما نیز چاپ جدید کتاب را به منظور تجهیز خوانندگان به دانش مورد نیاز برای هماهنگ شدن با این تغییرات، تهیه کرده‌ایم.

سلسله موضوعات مورد بحث در این کتاب، از تبیین روش علمی و ارتباط این روش با کندوکاو برای دستیابی به دانش در زمینه تعلیم و تربیت آغاز می‌شود. فرض ما بر این است که خوانندگان ما با مفاهیم، فرضیات، و اصطلاحات روش علمی آشنایی ندارند، بنابراین به توضیح آنها پرداخته‌ایم. مطالب کتاب با پیشنهادهایی برای تبدیل مسائل کلی به سؤالات مشخص دنبال می‌شود که با بررسی علمی از طریق جامعه و متغیرهای مورد علاقه انطباق دارد. سپس نقش تحقیقات و نظریه‌های پیشین را در طراحی پروژه تحقیق توضیح می‌دهیم. در این کتاب سعی کرده‌ایم که منابع مربوط به تحقیقات پیشین را به روز درآوریم و به این منظور تأکید ویژه‌ای بر نحوه استفاده از بانکهای اطلاعاتی داشته‌ایم، که شرایط مناسبی را برای دسترسی به تحقیقات و نظریه‌های مربوط فراهم می‌آورند. بعد از آن به بررسی راههایی می‌پردازیم که از آن طریق نظریه، تجربه، مشاهدات و تحقیقات پیشین به سمت صورتبندی فرضیه حرکت می‌کنند.

شیوه‌های نمونه‌گیری و شیوه‌های رایج‌تر آمار توصیفی و آمار استنباطی با تأکید بر نقش آنها در فرایند پژوهش و نحوه تفسیر آنها، تشریح می‌شوند. نقش مشاهده نظام‌مند و اندازه‌گیری نیز مورد بحث قرار گرفته، و نمونه‌هایی از شیوه‌های مناسب اندازه‌گیری ضمیمه شده است. فصل مربوط به اعتبار و پایایی مورد تجدید نظر قرار گرفته، با آخرین اطلاعات همراه شده و گسترش یافته است.

در مرحله بعد، در مورد انواع گوناگون تحقیق، که سودمندی آنها در تعلیم و تربیت به اثبات رسیده است، بحث شده و نقاط قوت و ضعف انواع پژوهشها، بدون حمایت از نوع خاصی به عنوان برترین تحقیق، بیان شده است. به چاپ اخیر، بخشی در مورد تحقیق کیفی اضافه کرده‌ایم و بخش مربوط به تحقیق همبستگی، به یک فصل مجزا تبدیل شده است.

بالاخره با ارائه قوانین عام برای تحلیل نتایج تحقیق و شیوه‌های مورد قبول در گزارش این نتایج، کتاب را به پایان می‌بریم. همچنین بخش مربوط به کاربرد کامپیوتر را با آخرین دانش موجود در این زمینه همراه ساخته‌ایم.

تمرکز اصلی چاپ اخیر نیز تدارک یک متن درسی بوده که برای یک واحد درسی مقدماتی تحقیق در تعلیم و تربیت طراحی شده است. هدف این کتاب آشنا کردن محققان تازه کار با شیوه‌های انجام یک پروژه تحقیق بدیع است. تأکید ما بر مسائل شاخص و کاربردی است که رویاروی امر تحقیق قرار می‌گیرد، و در این مسیر از صورتبندی مسئله تحقیق آغاز می‌کنیم و تا تهیه گزارش نهایی پیش می‌رویم.

در پایان هر فصل علاوه بر پرسشهای درسی، سوالات چند جوابی نیز در قالب راهنمای آموزشی گنجانده شده است.

هر چند کتاب مقدمه‌ای بر تحقیق در تعلیم و تربیت برای دانشجویان مبتدی در رشته تعلیم و تربیت تهیه شده است، ولی امیدواریم که دیگرانی هم که آروزمند یادگیری بیشتر در مورد فلسفه، مسائل و شیوه‌های پژوهش علمی در قلمرو تعلیم و تربیت هستند، آن را سودمند بیابند. معیار اصلی ما برای انتخاب مطالب کتاب، سودمندی بالقوه جنبه‌های مختلف تحقیق در تعلیم و تربیت برای دست‌اندرکاران تحقیق بوده است.

از همه پژوهشگرانی که از سه چاپ اولیه این کتاب بهره برده‌اند و پیشنهادهای بسیار ارزشمندی برای بهبود و روزآمد کردن چهارمین چاپ آن، ارائه کرده‌اند، عمیقاً قدردانی می‌کنیم.

از جان دی. میلر برای همکاری در تهیه بخش ملاحظات اخلاقی و حقوقی و از توماس شوآنت (Schwandt) برای شرکت در تهیه بخش پژوهش کیفی سپاسگزاریم. از ساموئل هوآنگ (Samuel Huang) برای تلاش وسیعش در تهیه یک فصل جامع در زمینه پیشینه تحقیق، شامل آخرین تحقیقات موجود در این زمینه، سپاسگزاری می‌کنیم. همچنین از وسلی کووالت (Westley Covalt) و جوناتان ریچ برای مساعدت در تهیه راهنمای آموزشی تشکر می‌کنیم.

ما سپاس خود را به ویراستار نهایی کتاب، سر رونالد فیشر F.R.S.، و به دکتر فرانک بیس F.R.S. تقدیم می‌کنیم و از گروه انتشاراتی لانگ من لندن سپاسگزاریم که به ما اجازه دادند جداول ۸۲، ۸۴، ۸۶ را از کتاب جداول آماری برای تحقیقات زیست‌شناسی کشاورزی و پزشکی (چاپ ششم ۱۹۷۴) در این کتاب چاپ کنیم.

در چاپ جدید این کتاب، از نظرات اساتید زیر استفاده شده است:

کوین کرهان (Kevin Crehan)، از دانشگاه نوادا، لاس وگاس، (NV)

آلفردو کوئلار (Alfredo Cuellar)، از دانشگاه سان دیه گو، سان دیه گو، (CA)

گیلبرت کوئواز (Gilbert Cuevas)، از دانشگاه میامی و کورال گیبلز، (FL)

شارلوت فار (Charlotte Farr)، از دانشگاه وایومینگ، لارامین، دلیو وای (WY)

رن جاکوبز (Ron Jacobs)، دانشگاه ایالتی سان دیه گو، سان دیه گو، (CA)

بن لین (Ben Layne)، دانشگاه ایالتی جورجیا، آتلانتا، (GA)

ادوین نوواک (Edwin Novak)، دانشگاه ایالتی اوهایو، کلمبوس، (OH)

رابرت پرزی بیژوسکی (Robert Przybyszewski)، دانشگاه نیآگارا، (NY)

دیوید کوئیست (David Quist)، کالج ایالتی ورستر، ورستر، (MA)

بروس راجرز (Bruce Rogers)، دانشگاه ایوای شمالی، سوارفالز، (IA)

دیل شاو (Dale Shaw)، دانشگاه کلرادوی شمالی، گریلی، (CO)

ویلیام ویر (William ware)، دانشگاه کارولینای شمالی، چپل هیل، (NC)

کینارد وایت (Kinnard White)، دانشگاه کارولینای شمالی، چپل هیل، (NC)

ریچارد ویلیامز (Richard Williams)، دانشگاه میامی، کورال گیبلز، (FL)

# بخش اول

## مبانی تحقیق

### فصل اول

#### روش علمی در تعلیم و تربیت

##### هدفهای آموزشی

دانشجو پس از مطالعه این فصل، قادر خواهد بود:

- ۱- پنج منبع دانش را فهرست و در مورد هر یک از آنها اظهار نظر کند.
- ۲- ویژگیهای روش علمی را تبیین کند.
- ۳- فرضیاتی مبتنی بر علم و نگرش ویژه دانشمندان اقامه کند.
- ۴- هدف و ویژگیهای نظریه علمی را مشخص کنند و بین شیوه‌های گوناگون ارائه یک نظریه در زمینه علوم رفتاری، تمیز قائل شود.
- ۵- محدودیتهای موجود در بکارگیری روشهای علمی را در علوم اجتماعی بیان کند.
- ۶- بین ویژگیهای تحقیق پایه‌ای و کاربردی تمیز قائل شود.
- ۷- اصطلاحات مفهوم، سازه و متغیر را تشریح کند.
- ۸- بین انواع گوناگون متغیرها تمایز قائل شود (مقوله‌ای در مقابل پیوسته، مستقل در مقابل وابسته، فعال در مقابل خصیصه‌ای).
- ۹- انواع گوناگون تعاریف را بیان کند (تعاریف ساختاری در مقابل تعاریف کاربردی، و تعاریف کاربردی قابل اندازه‌گیری در مقابل تعاریف کاربردی تجربی).
- ۱۰- پنج روش مورد استفاده در بررسیهای حوزه تعلیم و تربیت را تشریح کند.

کارشناسان تعلیم و تربیت الزاماً نقش تصمیم گیرندگان را نیز ایفا می کنند؛ زیرا در کار روزانه خود در زمینه فرآیند تعلیم و تربیت، با وظیفه تصمیم گیری در مورد چگونگی برنامه ریزی تجربیات یادگیری، روش تدریس، راهنمایی دانشجویان و سازماندهی یک نظام درسی در ابعاد بشمار دیگر روبرو هستند.

یک فرد متخصص، برخلاف افراد غیر متخصص که به آنها توصیه می شود چه کاری را چگونه انجام دهند، باید برای کار خود برنامه ریزی کند؛ چرا که فرض بر این است که دانش و مهارت لازم را برای تصمیم گیری درست در مورد اینکه چه می باید بکند و چگونه باید انجام دهد، دارد. اما یک کارشناس تعلیم و تربیت چگونه تشخیص می دهد که در شرایط خاص تصمیم درست کدام است؟ هر چند منابع دانش دیگری از جمله تجربه، مرجع مطلع و سنت وجود دارند، ولی این دانش علمی در زمینه فرآیند تعلیم و تربیت است که بیشترین سهم را در تصمیم گیری در این مورد به عهده دارد و کارشناسان تعلیم و تربیت می توانند برای کسب اطلاعات و ارائه پیشنهادهای معتبری که در شرایط تصمیم گیری به کار می آیند، به آن روی آورند. این گنجینه دانش به عنوان نتیجه پژوهشهای علمی در مسائل تعلیم و تربیت قابل دسترس بوده است. جریان تعلیم و تربیت همواره از نتایج این گونه تحقیقات علمی و نظام یافته تأثیر نمی پذیرد. در واقع باید متذکر شد که علم تعلیم و تربیت در مقایسه با سایر زمینه ها هنوز در مراحل اولیه پیشرفت قرار دارد.

### /// منابع علم

قبل از اینکه بیشتر به نقش تحقیق علمی در تعلیم و تربیت پردازیم، راههایی را که انسان در سراسر تاریخ برای پاسخگویی به سؤالاتش در نوردیده است، مرور می کنیم. سرچشمه های دانش آدمی را می توان به پنج دسته تقسیم کرد: ۱- تجربه ۲- مرجع مطلع و آگاه ۳- استدلال قیاسی ۴- استدلال استقرایی و ۵- روش علمی.

### // تجربه

تجربه، یکی از منابع آشنا و متداول علم است. برای مثال یک فرد بعد از آزمودن مسیرهای متعدد از خانه به محل کار، می آموزد که کدام مسیر در زمان کوتاهتری طی می شود و یا کمترین



تراکم را در رفت و آمد دارد و یا دارای زیباترین مناظر است هر فرد بر اساس تجربه شخصی می تواند پاسخهای بسیاری از سؤالاتی را که با آن روبرو است، بیابد. بخش عمده معرفت که از نسلی به نسل دیگر منتقل می شود، نتیجه تجربیات انسان است. اگر انسان قادر به بهره برداری از تجربیاتش نبود، پیشرفت وی به طور جدی به تاخیر می افتاد. در واقع، توانایی یادگیری از تجربه، عموماً به عنوان شاخص رفتار هوشمندانه مطرح می شود.

با وجود تمامی این فواید، تجربه به عنوان یکی از سرچشمه های حقیقت، دارای محدودیت های خاص خود نیز هست؛ زیرا هر فرد بر اساس ویژگیهای خاص خود تحت تاثیر حوادث قرار می گیرد. دو فرد در شرایط مشابه می توانند تجربیات بسیار متفاوتی داشته باشند. جنگلی که برای یکی بهشتی دلپذیر است، برای دیگری می تواند فضایی وحشت آور و تهدید کننده باشد. دو مسئول که یک کلاس را همزمان مورد مشاهده قرار می دهند، در صورتی که یکی بر موارد درست متمرکز شود و دیگری به موارد نادرست توجه کند، در واقع می توانند دو گزارش کاملاً متفاوت تهیه کنند.

نقش دیگر تجربه آن است که گرچه فرد نیاز به دانستن چیزهایی را به کرات در می یابد، ولی به تنهایی قادر به یاد گرفتن آنها از طریق تجربه شخصی نخواهد بود. ممکن است کودکی که آزاد گذاشته شده است تا حساب را نزد خود بیاموزد، روش عمل جمع را درک کند، ولی بسیار بعید به نظر می رسد که بتواند روشی کارآ برای حل مسائل جذر بیابد. یک آموزگار می تواند از طریق تجربه، جمعیت کلاس خود را در یک روز خاص برآورد کند، ولی نمی تواند جمعیت آمریکا را به تنهایی شمارش کند.

## // مرجع مطلع

برای دانستن چیزهایی که درک آن از طریق تجربه مشکل و یا غیر ممکن است، شخص اغلب به مرجعی مطلع روی می آورد، یعنی پاسخ سؤالاتش را نزد کسی جستجو می کند که در مورد مشکل خاص، تجربه دارد و یا از گونه ای آگاهی و صلاحیت بهره مند است. ما کلام کسانی را که به عنوان مرجع شناخته شده اند، به مثابه حقیقت می پذیریم. یک فرد برای دانستن تعداد جمعیت آمریکا به دفتر گزارشهای آماری ایالات متحده مراجعه می کند. یک دانش آموز تلفظ صحیح یک کلمه را در فرهنگ لغات جستجو می کند. مسئول یک دبستان در مورد یک مشکل

حقوقی و قانونی در دبستان، با یک مشاور حقوقی مشورت می‌کند. یک معلم تازه کار هم از آموزگاری با تجربه اظهار نظر می‌خواهد. یک معلم تازه کار ممکن است تکنیک و یاروش معینی را برای تدریس روخوانی آزمایش کند؛ چرا که سرپرستش آن را به دلیل موثرتر بودن به او پیشنهاد کرده است.

ما حتی در تاریخ هم می‌توانیم مثالهایی دربارهٔ اعتماد به نظر مراجع مطلع برای درک حقیقت بیابیم. بویژه در قرون وسطی، که مراجعی مانند افلاطون<sup>(۱)</sup> و ارسطو<sup>(۲)</sup> و کشیشان متقدم کلیسا به عنوان منابعی از حقیقت، حتی از مشاهده مستقیم و یا تجربه ارزشمندتر شمرده می‌شدند. اگر چه مرجع مطلع در تاریخ همواره یکی از مفیدترین منابع علم به شمار رفته است، ولی این سؤال که «مرجع از کجا می‌داند» نیز نباید به فراموشی سپرده شود. در روزگاران پیشین از آنجا که مرجع مطلع موقعیتی همانند شاه، رئیس و یا اسقف اعظم داشته، همیشه بر حق فرض می‌شده است. امروزه کسی صرفاً به علت موقعیت مرجع مطلع به استناد و یا اعتماد به او تمایل ندارد. ما اظهارات این مراجع را در صورتی می‌پذیریم که براساس تجربه و یا سایر منابع شناخته شده علمی بنا شده باشند.

مراجع با رسوم و سنتها، که بشر برای پاسخگویی به بسیاری از سؤالات خود به آنها متکی است، رابطه نزدیکی دارند. این امر هم در مورد مسائل حرفه‌ای و هم در مورد مسائل روزمره صادق است. به عبارت دیگر انسان اغلب از خود می‌پرسد با این مسئله در گذشته چگونه برخورد می‌شده است. و سپس از این سؤال به عنوان راهنمایی برای حرکت خود استفاده می‌کند. عرف و سنت، بویژه در محافل آموزشی که مریبان به تجربیات گذشته به عنوان چراغ راه تکیه می‌کنند، تأثیرات قابل توجهی دارند، ولی با این حال در تاریخ تعلیم و تربیت با بسیاری از سنتها مواجه می‌شویم که طی سالیان مسلط بوده‌اند و بعداً مشخص شده است که از اساس نادرست بوده‌اند و باید کنار گذاشته شوند. روش خردمدانه این است که سنتها و رسوم را قبل از پذیرفتن به عنوان حقیقت به دقت ارزیابی کنیم.

مرجع مطلع به عنوان یکی از ذخایر دانش، دارای کاستی‌هایی است که باید آنها را مورد تأمل قرار داد. نخست آنکه مرجع مطلع می‌تواند در اشتباه باشد؛ هیچ دلیلی بر اشتباه ناپذیری او وجود ندارد. همچنین در می‌یابیم که در میان مراجع مطلع نسبت به اصول، اختلاف نظر وجود

دارد. دلیل این امر آن است که اظهارات موثق آنان بیش از آنکه بر واقعیت متکی باشد، به عقاید شخصی بستگی دارد.

### // استدلال قیاسی

شاید اولین گام اساسی در تکامل یک رویکرد نظام‌مند جهت کشف حقیقت را فلاسفه یونان برداشتند. ارسطو و پیروانش استفاده از استدلال قیاسی را، که فرآیندی فکری است و طی آن فرد با بهره‌جویی از قوانین معین منطق، از احکام کلی به احکام جزئی می‌رسد، مطرح کردند. این روش یک سیستم برای سازماندهی واقعیات شناخته شده جهت دستیابی به یک نتیجه است. این جریان با بکارگیری استدلال منطقی انجام می‌شود. یک استدلال مبتنی بر چند حکم است که با یکدیگر مرتبط هستند. حکم پایانی، نتیجه و بقیه صغری و کبری خوانده می‌شوند، که شامل شواهد اثبات شده هستند. یکی از مهمترین انواع استدلال قیاسی، قیاس صوری است. یک قیاس صوری مبتنی بر یک فرض اصلی (کبری) و یک فرض فرعی (صغری) است که به یک نتیجه ختم می‌شود. یک نمونه از قیاس صوری عبارت است از: همه مردان فنا پذیرند (فرض کبرای اصلی)؛ پادشاه یک مرد است (فرض صغرای فرعی)؛ بنابراین، پادشاه فنا پذیر است (نتیجه).

در استدلال قیاسی، اگر صغری و کبری درست باشند، ضرورتاً نتیجه نیز درست خواهد بود. استدلال قیاسی این توانایی را به فرد می‌دهد که صغری و کبری‌ها را در طرحی سازماندهی کند که شواهدی قطعی را برای اعتبار نتیجه فراهم می‌آورد. همانطور که خوانندگان کتب جنایی به یاد می‌آورند، شرلوک هلمز دائماً می‌گفت: «من قیاس می‌کنم...» وی حقایق گسسته و ناپیوسته اولیه را به نحوی ترکیب می‌کرد که نتایج تردیدناپذیر حاصل شوند.

با این حال، استدلال قیاسی محدودیت‌هایی خاص خود دارد. نتیجه یک قیاس صوری نمی‌تواند از محتوای صغری و کبری فراتر رود. بنابراین برای رسیدن به نتایج درست می‌باید از صغری و کبری درست شروع کنیم. از آنجا که نتیجه‌گیریهای قیاسی ضرورتاً از شرح و بسط معرفت موجود حاصل می‌شوند، نمی‌توان به دلیل دشواری در تبیین حقیقت جهان شمول بسیاری از احکام مرتبط با پدیده‌های علمی، بررسی علمی را تنها از طریق استدلال قیاسی انجام داد. استدلال قیاسی اگر چه می‌تواند آنچه را که تاکنون شناخته شده است، سازمان دهد و

ارتباطات نوینی را به شکل جریانی از کل به جزء میان آنها نشان دهد، ولی به عنوان ابزاری برای درک واقعیت‌های تازه کافی نیست.

به رغم این محدودیتها، استدلال قیاسی در فرآیند پژوهش مفید است. این استدلال، ابزاری پیوستن نظریه را به مشاهده فراهم می‌آورد و پژوهشگران را قادر می‌سازد که از طریق نظریه‌های موجود، قیاس کنند که چه پدیده‌ای می‌باید مورد مشاهده قرار گیرد. قیاس از نظریه می‌تواند فرضیه‌ها را، که بخش حیاتی یک بررسی علمی هستند، فراهم آورد.

### // استدلال استقرایی

نتایج ناشی از استدلال قیاسی تنها زمانی درست است که صغری و کبری آن به شیوه‌ای صحیح بنا شده باشد، اما چگونه می‌توان درستی صغری و کبری را دریافت؟ در قرون وسطی عقاید تعصب آمیز یا نتایجی که از طریق یافته‌های نامعتبر به دست آمده بودند، اغلب جانشین مفروضات درست می‌شدند. برای نخستین بار فرانسیس بیکن (۱۶۲۶-۱۵۶۱)، خواهان رویکرد جدیدی برای کسب دانش شد. او اظهار داشت که انسان نباید خود را اسیر مفروضات پذیرفته شده‌ای کند که به منزله حقایق مسلم و به وسیله مراجع مطلع سینه به سینه نقل شده است. او عقیده داشت که یک محقق باید نتایج عام را بر اساس واقعیاتی بنیان نهد که از طریق مشاهده مستقیم جمع‌آوری شده است. بیکن به جویندگان حقیقت توصیه می‌کرد تعصبات و عقاید قبلی را که «بت»<sup>(۱)</sup> می‌نامید، به کناری بگذارند و طبیعت را مستقیماً مورد مشاهده قرار دهند. از نظر بیکن کسب علم ایجاب می‌کند که انسان شخصاً طبیعت را مورد مشاهده قرار دهد و واقعیات معین را جمع‌آوری کند و از این یافته‌ها به یک تعمیم دست یابد. اهمیت مشاهده در مثالی که منتسب به فرانسیس بیکن است، نشان داده شده است.<sup>(۲)</sup>

در سال ۱۴۳۲ میلادی در مورد تعداد دندانهای اسب بحثی دائمی در بین کشیشان در گرفته بود. این مجادله ۱۳ روز بدون وقفه در جریان بود. تمام کتب قدیمی و تاریخ‌نگاریها بیرون کشیده شدند و افاضات عجیب و بر حجم و ثقلی که تا قبل از آن هرگز شنیده نشده بود، به صورت بیانیه ارائه شد. در روز چهاردهم، راهب جوانی اجازه خواست برای ختم این

1- Idol

2- Mess, C.E K (1934). Scientific thought and social reconstruction. General Electric Review, 37, 113 - 119.

مجادلات حیرت‌انگیز که اندیشمندان ژرف نگر را چون زخمی می‌آزرد؛ کلامی بر زبان راند و راهی مستقیم پیشنهاد کند. این راهب جوان از آنان نقاضا کرد که به روشهای مهجور تسلیم نشوند و به دهان باز اسب نگاه کنند تا پاسخی برای سؤالات خود بیابند. این سخن بر کشیشان گران آمد و آن را توهین مستقیم به مقام و شأن خود تلقی کردند. غوغای عظیمی به پاخاست. بر سر راهب جوان ریختند و استخوان رانش را شکستند و بی‌درنگ او را از جمع خود بیرون و اظهار کردند که شیطان، این راهب گستاخ را اغوا کرده است تا با این بیانات کفرآمیز و بی‌سابقه، با میراث کشیشان به مخالفت برخیزد. چندین روز بعد از این نزاع اندوهناک، انجمن صلح‌آمیزی تشکیل یافت و کشیشان یکپارچه اعلام داشتند که این مسئله به دلیل کمبود فاحش مدارک تاریخی و دینی به صورت یک معمای همیشگی باقی خواهد ماند و به این ترتیب، اظهار یک چنین احکامی نیز ممنوع شد.

در این حکایت، راهب جوان در طلب راه جدیدی برای جستجوی واقعیات، به جای وابستگی و اتکا به مراجع مطلع و دیدگاههای مطلق بود. این تلاش به اصل بنیادین تمامی علوم تبدیل شد.

در نظام فکری بیکن، بر رویدادهای خاص یک طبقه (گروه)، مشاهده صورت می‌گیرد و سپس بر مبنای رویدادهای مشاهده شده، استنتاجهایی در مورد کل طبقه (گروه) انجام می‌پذیرد. این رویکرد که به استدلال استقرایی معروف است، خلاف رویکردهای مورد استفاده در روشهای قیاسی است. تفاوت بین استدلال قیاسی و استقرایی در مثال زیر دیده می‌شود.

الف. قیاس: هر پستانداری شیشی دارد.

همه خرگوشها پستاندارند.

بنابراین، هر خرگوشی شش دارد.

ب. استقرا: هر خرگوشی که تا به حال مشاهده شده است، شش دارد.

بنابراین همه خرگوشها شش دارند.

توجه کنید که در استدلال قیاسی صغری و کبری قبل از دستیابی به نتیجه باید شناخته شده و معین باشد، اما در استدلال استقرایی نتیجه از طریق موارد مشاهده شده و تعمیم آنها به کل طبقه حاصل می‌شود. برای اطمینان کامل از نتایج استقرا تمامی موارد می‌باید مورد مشاهده قرار گیرند. در نظام فلسفی بیکن این مورد به عنوان استقرای تام ذکر شده است؛ این مسئله ایجاب

می‌کند که محقق تمام نمونه‌های یک پدیده را آزمایش کند. در مثال بالا، اگر محقق می‌خواست مطمئن باشد که هر خرگوشی شش دارد، می‌باید تمامی خرگوشهای گذشته، حال و آینده را مورد مشاهده قرار داده باشد. این آزمایش معمولاً در عمل امکان‌پذیر نیست، بنابراین محقق در مجموع باید به استقرای ناقص و مبتنی بر مشاهده ناکامل اعتماد کند.

نتایج حاصل از استقراء، تنها زمانی می‌تواند کامل باشد که طبقه و یا گروه مطالعه کوچک باشد. برای مثال، یک آموزگار ممکن است مشاهده کند که تمام شاگردان سرخ‌موی کلاس، در هجی کردن، نمراتی بالای متوسط دارند. او طبیعاً می‌تواند ادعا کند که دانش‌آموزان سرخ‌مویی که در این کلاس ثبت نام کرده‌اند، نمراتی بالاتر از حد متوسط دارند، ولی مجاز نیست که این نتایج را به نمرات هجی کردن بچه‌های سرخ‌موی سایر کلاسها و یا کلاسهای آتی خود تعمیم دهد.

از آنجا که تنها در یک گروه نمونه کوچک می‌توان استقرای تام انجام داد، معمولاً از استقرای ناقص استفاده می‌کنیم؛ به این معنا که نمونه‌ای از یک گروه را مورد مشاهده قرار می‌دهیم و از این طریق خصوصیت مشترک کل گروه را استنتاج می‌کنیم. یک مثال از نتیجه‌گیری مبتنی بر استقرای ناقص، طرز تفکر کنونی در مورد خصوصیات جسمانی بسیاری از کودکان تیزهوش است. سالهای متمادی این مسئله به طور عام پذیرفته شده بود که کودکان تیزهوش، مستعد داشتن جسمی ضعیف هستند. حتی امروز هم نقاشان تصاویر متحرک، اغلب، پسر بچه‌های باهوش را به صورت کودکانی استخوانی با عینک ته استکانی به تصویر می‌کشند. لویس ترمن، یکی از پیشگامان روان‌آزمایی به خصوصیات ویژه کودکان تیزهوش علاقه‌مند بود<sup>(۱)</sup> و مطالعه‌ای متمرکز بر روی ۱۰۰۰ کودک کالیفرنایی که هوشبهر آنها در آزمون هوش استانفورد - بینه بالایی ۱۴۰ نمره بود، صورت داد. او دریافت که متوسط قد و وزن و سلامت جسمی و عمومی این دسته از کودکان کمی از متوسط سایر کودکان بالاتر است. از این مطالعه می‌توان نتیجه گرفت که کودکان تیزهوش، برخلاف تصور سنتی، از نظر رشد جسمانی نسبت به کودکان با هوشبهر متوسط، کمی بالاترند.

توجه کنید که این نتیجه به طور دقیق به اثبات نرسیده است و تنها با احتمال بالایی به نظر درست می‌رسد. برای اینکه بتوان به طور یقین از این نتایج مطمئن بود، باید اندازه‌های فیزیکی تمامی کودکانی که در آزمون استانفورد - بینه هوشبهر ۱۴۰ و بالاتر دارند، گردآوری کرد. در این

صورت می‌توان نسبت به خصوصیات این دسته از کودکان در این مقطع زمانی مطمئن بود، ولی نمی‌توان ۱۰۰ درصد اطمینان داشت که این نتایج در مورد یک چنین کودکانی در آینده نیز صدق خواهد کرد. اگر چه استقرای ناقص، امکان دستیابی به نتایج مسلم و مصون از خطا را برای ما فراهم نمی‌آورد، ولی اطلاعات معتبری را به دست می‌دهد که از آن طریق انسان می‌تواند تصمیماتی منطقی اتخاذ کند.

### // روش علمی

استفاده انحصاری از روش استقراء اگر چه به انباشت آگاهیها و اطلاعات ازهم‌گسته انجامید، اما سهم کمی در پیشرفت علم داشت. به علاوه این نتیجه به دست آمد که بسیاری از مسائل از طریق استقرای صرف قابل حل نیستند.

در نهایت چاره‌ای نبود جز آنکه دانش پژوهان مهمترین جنبه‌های روشهای استقرایی و قیاسی را با هم ترکیب کنند و روش جدیدی به نام روش استقراء - قیاس یا روش علمی به دست آورند. از نظر عموم، چارلز داروین در بسط نظریه تکامل خود، اولین کسی است که در پی جویی علم، از این روش استفاده کرده است.

داروین می‌گوید که مدت مدیدی را با امید دستیابی به قوانین کلی تکامل، صرف مشاهدات زیست شناسانه کرده است. او تشریح می‌کند که چگونه به رویکردی جدید دست پیدا کرد:

اولین دفترچه یادداشتهای من (در باب نظریه تکامل) در جولای ۱۸۳۷ تقریر شد. من بر مبنای اصول واقعی بیکن کار می‌کردم و بدون آن که نظریه‌ای در دست داشته باشم، واقعیات را در مقیاس نامحدود و مشخصاً با مراجعه به گونه‌های اهلی شده حیوانات، تحقیقات مدون و مصاحبه با پرورش دهندگان حیوانات و باغبانان ماهر و همچنین مطالعه وسیع جمع‌آوری کردم. وقتی فهرست کتابهایی را که خوانده و خلاصه کرده‌ام و شامل کلیه جراید و استاد و مدارک می‌شود، نگاه می‌کنم از مجاهدت و پشتکار خود در حیرت فرو می‌روم. خیلی زود دریافتم که انتخاب<sup>(۱)</sup>، کلید موفقیت بشر در بهره‌برداری از نژادهای مختلف حیوانات و گیاهان است، ولی برای من این مسئله همچنان به عنوان یک معما باقی بود که چگونه در مورد یک موجود خاص، که در بخشی از طبیعت زندگی می‌کند، این انتخاب صورت می‌گیرد.

در اکتبر ۱۸۳۸، بانزده ماه بعد از شروع این تحقیقات نظام‌مند، هنگامی که تصادفاً برای وقت گذرانی، مشغول خواندن «نظریه جمعیت مالتوس» بودم، زمینه درک پدیده مبارزه برای بقا، که در تمامی عادات حیوانات و گیاهان به چشم می‌خورد، در من فراهم شده بود، ناگهان این نکته به ذهنم رسید که در چنین شرایطی گونه‌های مطلوب حفظ و انواع نامطلوب منقرض می‌شوند و نتیجه، شکل‌گیری گونه‌های جدید خواهد بود. چنین شد که به نظریه‌ای دست یافتم که از طریق آن...<sup>(۱)</sup>

روش داروین، یعنی مشاهده صرف، تا پیش از این که مطالعات وی و فراتر از آن تفکر او این امکان را فراهم آورد که فرضیه‌ای آزمایشی برای تبیین واقعیات به دست آمده از طریق مشاهده صورت‌بندی کند، بی‌حاصل بود.

سپس، او با انجام قیاسهایی از این فرضیه و گردآوری داده‌های اضافی به مرحله آزمون فرضیه رسید تا تعیین کند که آیا این داده‌ها فرضیه را تأیید می‌کنند؟ یا استفاده از این روش بررسی، داروین توانست نظریه تکامل خود را ارائه کند. بکارگیری توأمان استدلال قیاسی و استقرایی، از خصوصیات بارز روش‌نویین تحقیق است که به عنوان معتبرترین روش کسب علم مورد توجه قرار گرفته است.

روش علمی مجموعاً به عنوان روشی تبیین می‌شود که از طریق آن محقق به طور استقرایی از مشاهده به فرضیه و سپس به طور قیاسی از فرضیه به استنباط منطقی می‌رسد و محقق، پیامدهایی را که در صورت درستی رابطه فرض شده حادث خواهند شد، قیاس می‌کند. اگر استنباطهای قیاسی با پیکره سازمان یافته دانش مورد قبول سازگار باشند، پس با گردآوری داده‌های تجربی باز هم مورد آزمون قرار می‌گیرند و بر مبنای شواهد و فرضیه‌ها رد و یا تأیید می‌شوند.

استفاده از فرضیه یکی از تفاوت‌های اصولی بین رویکرد علمی و استدلال استقرایی است. در استدلال استقرایی، انسان در وهله اول مشاهده می‌کند و سپس اطلاعات به دست آمده را سازمان می‌دهد. در رویکرد علمی، شخص چنین استدلال می‌کند که اگر این فرضیه درست باشد، چه به دست خواهد آمد و سپس به طور نظام‌مند برای آنکه بتواند فرضیه را رد و یا قبول کند، مشاهده را آغاز می‌کند.



## // نمونه‌ای از یک رویکرد علمی

رابرت پیرزیگ با تشبیه رویکرد علمی به فرآیند نگهداری یک موتور سیکلت در یک وضعیت خوب (از نظر کار موتور)، تعریف روشن و موجزی از آن ارائه می‌کند:

دو نوع منطقی به کار گرفته می‌شود: استقرا و قیاس. استدلال استقرایی با مشاهده موتور آغاز می‌کند و به یک نتیجه‌گیری کلی می‌رسد. برای مثال، اگر یک موتور سیکلت در دست‌اندازی بیفتد و موتور آن خاموش شود و دوباره در دست‌انداز دیگری بیفتد و موتور خاموش شود، بار دیگر نیز این مورد تکرار شود و سپس در جاده صاف و هموار بدون دست‌اندازی بی‌آنکه موتور خاموش شود، حرکت کند، راننده می‌تواند منطقیاً نتیجه بگیرد که خاموشی موتور ناشی از دست‌انداز بوده است. این یک استقرا است: استنتاج از تجربیات خاص به حقایق کلی.

استنتاجات قیاسی خلاف این عمل می‌کنند. آنها از دانش کلی شروع و یک مشاهده خاص را پیش‌بینی می‌کنند. برای نمونه، اگر یک تعمیرکار با خواندن سلسله مراتب مربوط به موتور ماشین در یابد که محور چرخ منحصرأ با نیروی برق تولید شده از طریق باطری به حرکت در می‌آید، سپس می‌تواند منطقیاً استدلال کند که اگر باطری تمام شود، محور چرخ کار نخواهد کرد. این یک قیاس است.

راه حل مسائلی که حل آنها برای یک ذهن عادی بسیار پیچیده است، به وسیله رشته‌های طولانی حاصل می‌شود که آمیزه‌ای هستند از استنتاجهای استقرایی و قیاسی که بین ماشین مورد مشاهده و سلسله مراتبی که بر اساس دفترچه‌های راهنمای ماشین در ذهن ایجاد می‌شود، در نوسان است. یک برنامه صحیح برای در هم تابیدن این رشته‌ها، به عنوان روش علمی رسمیت می‌یابد.

در واقع من هرگز فکر نکرده‌ام که مسئله تعمیر و نگهداری موتور سیکلت به قدری پیچیده است که به یک روش علمی رسمی نیاز دارد. مشکلات تعمیر موتور سیکلت چندان هم سخت نیست. هنگامی که به یک روش علمی رسمی می‌اندیشیم، اغلب تصویری از یک تریلی غول یکر یا یک بولدوزر عظیم‌الجثه، که به آرامی و به شیوه‌ای کسل‌کننده، بر سر و صدا و پر تلاش ولی خستگی ناپذیر کار می‌کند، در ذهن مجسم می‌شود. کار شما ممکن است دو برابر، پنج برابر و یا دوازده برابر زمان کار تعمیرکاری که از تکنیکهای معمولی بهره می‌گیرد،

به طول انجامد، ولی شما می‌دانید که در نهایت از پس آن بر خواهید آمد. در تعمیر و نگهداری موتور سیکلت هیچ عیب و اشکال مکانیکی وجود ندارد که از طریق روش علمی نتوان آنرا تشخیص داد. هر گاه به یک مشکل واقعی برخورد کردید، همه راهها را امتحان کردید، به مغز خود فشار آوردید، ولی مشکل حل نشد و شما دانستید که اکنون طبیعت تصمیم گرفته است که کار را بر شما دشوار سازد، به خود خواهید گفت: خوب ای طبیعت، اکنون دیگر وقت ملایمت نیست، باید کاری دیگر کرد، آن گاه به روش علمی رسمی روی خواهید آورد. به این منظور یک دفترچه یادداشت فراهم می‌کنید و همه چیز را به طور منظم در آن ثبت می‌کنید، بنابراین شما می‌دانید که در هر لحظه در کجا قرار دارید و می‌خواهید به کجا بروید. این امر در کار علمی و تکنولوژی الکترونیک ضروری است: چون مشکلات به قدری پیچیده می‌شوند که شما در حل آنها احساس ضعف می‌کنید، سردرگم می‌شوید و فراموش می‌کنید که چه می‌خواهید و یا طالب چه چیزی نیستید و باید آنرا را رها کنید. در تعمیر و نگهداری موتور سیکلت، مسائل در حدی نیست که انسان را گرفتار کند؛ ولی وقتی سردرگمی آغاز می‌شود، بهترین کار آن است که با منظم و دقیق کردن هر چیز بر آن فائق آید. گاهی اوقات، تنها نوشتن مشکلات، ذهن شما را به سمت شناخت واقعی آنها هدایت می‌کند.

بیانات منطقی ثبت شده در دفترچه یادداشت به ۶ طبقه تقسیم می‌شوند: (۱) بیان مسئله، (۲) فرضیه‌ها؛ به عنوان هدف مسئله؛ (۳) طرح آزمایشی برای آزمون هر فرضیه، (۴) نتایج پیشگویانه آزمایشها؛ (۵) نتایج مشاهده شده در جریان آزمایشها؛ و (۶) نتیجه‌گیری از نتایج آزمایشها. این شیوه، با نظم و ترتیب موجود در بسیاری از دفترچه‌های یادداشت آزمایشگاهی در دبیرستانها و کالج‌ها تفاوتی ندارد، ولی در اینجا هدف تنها به کارگل واداشتن نیست. اکنون هدف هدایت دقیق افکار است که اگر درست نباشند، به ثمر نخواهند رسید.

هدف واقعی روش علمی، حصول اطمینان از این است که طبیعت شما را به گونه‌ای گمراه نسازد که نا دانسته‌های خود را دانسته بپندارید. هیچ تعمیرکار یا دانشمند یا تکنسینی را نمی‌توان یافت که از این مسئله رنج برده و در نتیجه به طور غریزی محتاط نشده باشد. این، پاسخ اصلی این سؤال است که چرا اطلاعات علمی و فنی تا این حد خسته‌کننده و احتیاط آمیز به نظر می‌آیند. اگر دقت نکنید، اگر با اطلاعات علمی به طور احساسی برخورد کنید، اگر اطلاعات علمی را بی حساب اینجا و آنجا به کار ببرید، طبیعت خیلی زود از شما یک احمق

کامل خواهد ساخت. به هر حال، حتی وقتی که به طبیعت فرصت چنین کاری را نمی‌دهید، باز هم این بلا را بر سر شما می‌آورد. کسی که با طبیعت سروکار دارد، باید بی‌اندازه دقیق و کاملاً منطقی باشد؛ یک لغزش در استدلال منطقی کافی است تا تمام کاخ علم درهم فروریزد. یک قیاس نادرست در مورد ماشین کافی است تا شمارا مدت‌های مدید گرفتار سازد.

در مرحله اول روش علمی رسمی، که بیان مسئله است، مهارت اصلی آن است که چیزی را بیان کنید که مطمئن هستید می‌دانید، نه چیزی بیش از آن. بسیار بهتر است که با این بیان آغاز کنید. «مسئله را حل کنید: چرا موتور سیکلت کار نمی‌کند؟» که هر چند به نظر گنگ می‌آید، ولی درست است؛ تا این که بگویید «مسئله را حل کنید: سیستم برق موتور سیکلت چه اشکالی دارد؟» در حالی که اصلاً نمی‌دانید که اشکال از سیستم برق آن است، آنچه که شما باید بیان کنید این است «مسئله را حل کنید: موتور سیکلت چه اشکالی پیدا کرده است؟» و سپس برای ورود به مرحله دوم با این بیان ادامه می‌دهید «فرضیه شماره یک: اشکال در سیستم برق است». شما در مورد هر یک از فرضیه‌های ممکن فکر می‌کنید و سپس برای آزمون آنکه کدام یک درست و کدام یک غلط هستند، آزمایشهایی را طراحی می‌کنید.

این رویکرد دقیق در طرح پرسش آغازین، شما را از ارتکاب یک اشتباه بزرگ، که می‌تواند به چند هفته کار فوق‌العاده منجر شود و یا شما را کاملاً گرفتار کند، باز می‌دارد. به این دلیل سوالات علمی اغلب دارای ظاهری گنگ هستند. این سؤالات برای آن مطرح می‌شوند که از اشتباهات خسته‌کننده بعدی جلوگیری کنند.

مرحله سوم، یعنی مرحله‌ای از روش علمی رسمی که آزمایشگری نامیده می‌شود و برخی از اوقات، افراد خیالی‌باف آن را به مثابه همه مراحل روش علمی می‌پندارند، چون این تنها مرحله‌ای است که جذابیت بصری دارد. آنها به این مرحله نه عنوان تعداد بشمارای از لوله‌های آزمایش و وسایل عجیب و غریب نگاه می‌کنند که افراد جهت اکتشاف به دور آنها حلقه زده‌اند. آنها آزمایش را به عنوان بخشی از یک فرایند وسیع‌تر عقلانی نگاه نمی‌کنند و از این رو اغلب این آزمایشها را با نمایش عملی آزمایشهای درسی، که شبیه هم به نظر می‌رسند، به یک چشم نگاه می‌کنند. فردی که یک برنامه تدرستی علمی را با ابزار و تجهیزات محیرالعقول ۵۰ هزار دلاری اجرا می‌کند، اگر از قبل بداند که تلاشهایش به کجا ختم می‌شود، در واقع هیچ نوع کار علمی انجام نداده است. از طرف دیگر، یک تعمیرکار که برای آگاهی از

کار باطری بوق موتورسیکلت را به صدا در می‌آورد، به طور غیر رسمی یک تجربه علمی واقعی را انجام می‌دهد. او با فرار دادن سؤالی در مقابل طبیعت، یک فرضیه را مورد آزمایش قرار می‌دهد. در یک نمایش تلویزیونی، نجوای اندوهگین دانشمندی که می‌گوید: «آزمایش به نتیجه نمی‌رسد، ما شکست خوردیم»، سخنی نادرست است. اساساً یک آزمایش، هرگز به دلیل این که به نتایج پیش‌بینی شده دست نمی‌یابد، شکست خورده محسوب نمی‌شود. یک آزمایش تنها هنگامی یک شکست است که موفق نشود فرضیه مورد سؤال را به طرز شایسته آزمون کند، یعنی داده‌های آن، هیچ چیزی را اثبات نکنند.

مهارت در این مرحله شامل استفاده از آزمایشهایی است که صرفاً فرضیه مورد سؤال را آزمون کند، نه چیزی بیشتر و نه چیزی کمتر. اگر بوق کار کند و موتور هم روشن شود، این بدان معنی است که کل سیستم برقی موتور سیکلت کار می‌کند، پس تعمیرکار دچار دردسر عمیقی شده و به یک نتیجه غیر منطقی رسیده است. کار کردن بوق صرفاً به او می‌گوید که باطری و بوق کار می‌کنند. برای طراحی یک آزمایش مناسب او باید خیلی دقیق در این معنا بیندیشد که چه چیزی مستقیماً چه چیز را موجب می‌شود. این مطلب را از طریق سلسله مراتب خواهید یافت. بوق باعث حرکت موتور سیکلت نمی‌شود. باطری نیز جز در یک طریق بسیار غیر مستقیم موتور را به کار نمی‌اندازد. نکته اصلی این است که سیستم برق مستقیماً باعث روشن شدن موتور از طریق شمع می‌شود و اگر شما این مسئله را در این مرحله، یعنی در خروجی سیستم برق، آزمون نکنید، هرگز به واقع نخواهید دانست که نقص از سیستم برقی است یا نه. برای آزمون مناسب، تعمیرکار سر شمع را می‌کشد و آن را در جوار بدنه موتور نگه می‌دارد تا پایه سر شمع با زمین اتصال پیدا کند. سپس استارت می‌زند و دهانه شمع را از نظر وجود جرقه آبی بررسی می‌کند. اگر جرقه‌ای وجود نداشته باشد، او می‌تواند به یکی از این دو نتیجه برسد: (الف) در اینجا یک نقص برقی وجود دارد یا (ب) آزمایش او دقیق نیست. اگر فرد با تجربه‌ای باشد، سعی خواهد کرد که مدت بیشتری روی آن کار کند، اتصالات را بررسی و در مورد همه روشهای ممکن برای گرفتن برق از آن برای سر شمع فکر کند. سپس اگر نتوانست از آن برق بگیرد، در نهایت نتیجه می‌گیرد که فرضیه الف درست است. در اینجا یک نقص برقی وجود دارد و آزمایش پایان یافته است. او ثابت کرده است که این فرضیه درست است.

در مرحله آخر، یعنی نتیجه‌گیری، مهارت عبارت است از بیان نتیجه‌ای که آزمایش آن را

ثابت کرده است و نه بیشتر. این تعمیرکار معلوم نکرده است که وقتی سر شمع را سر جای خود قرار می‌دهد، موتور شروع به کار خواهد کرد یا نه. امکان دارد اشکالات دیگری نیز وجود داشته باشد، ولی او می‌داند مادامی که سیستم برقی کار نکند، موتورسیکلت به حرکت در نخواهد آمد. بنابراین سؤال رسمی بعدی را مطرح می‌کند. «مسئله را حل کنید: سیستم برقی موتور سیکلت چه اشکالی دارد؟»

او سپس برای این اشکالات فرضیه‌هایی اقامه و آنها را آزمون می‌کند. تعمیرکار با طرح سؤالات مناسب و انتخاب آزمونهای درست و کسب نتایج صحیح، به ترتیب تمام قطعات موتور را واریسی می‌کند تا دقیقاً علت این نقص را بیابد و سپس آنرا رفع کند تا دیگر باعث از کارافتادن موتورسیکلت نشود.

مشاهده‌گری که آموزش ندیده باشد، تنها به کار جسمانی توجه دارد و اغلب تصور می‌کند که کار جسمانی اساساً همه کاری است که یک تعمیرکار انجام می‌دهد. در واقع کارهای جسمانی کوچکترین و سهل‌ترین بخش از کارهایی هستند که یک تعمیرکار انجام می‌دهد. مهمترین بخش از کار او را تا حد زیادی مشاهده و تفکر دقیق تشکیل می‌دهد و به این دلیل است که تعمیرکاران اغلب اوقات به هنگام انجام این آزمایشها کم حرف و در خود فرو رفته به نظر می‌رسند. آنها ضمن کار دوست ندارند با کسی صحبت کنند؛ چون آنها بر تصورات ذهنی و سلسله مراتب موجود در موتور سیکلت متمرکز هستند و به واقع اصلاً به شما و یا بدنه موتور سیکلت نگاه نمی‌کنند. آنها از آزمایش به عنوان بخشی از یک برنامه برای افزایش سلسله مراتب دانش خود در مورد نقص موتور سیکلت استفاده می‌کنند و آن را با سلسله مراتب صحیح در مغز خود مقایسه می‌کنند. آنها به شکل زیربنایی کار، نظر دارند.<sup>(۱)</sup>

### /// ماهیت علم

در این جا می‌باید به این مسئله نیز اشاره کرد که اگر چه ممکن است تمام علوم از نظر اجزای تشکیل دهنده و یا روشهای ویژه با یکدیگر متفاوت باشند، ولی این روش کلی برای دستیابی به دانش معتبر در آنها مشترک است. همین روش بررسی تعیین می‌کند که چه چیزی علم است یا علم نیست. شاید علم در بهترین تعریفش، نوعی روش بررسی باشد که به محقق امکان آزمودن

1- Pirsig R.M. (1974). Zen and the Art Of Motorcycle: An Inquiry Into Values (PP.107111). New - york. (Reprinted by Permission of William Morrow & Company).

پدیده مورد علاقه‌اش را می‌دهد. علاوه بر روشی که دانشمندان در جستجوی حقیقت از آن پیروی می‌کنند، مفهوم علم جنبه‌های دیگری نیز در بردارد که آنها را به اختصار خواهیم آزمود: (۱) اقامه فرضیه از سوی دانشمندان، (۲) نگرش‌های دانشمندان، و (۳) صورتبندی نظریه علمی.

### // فرضیات دانشمندان

یک فرض اساسی دانشمندان این است که رویدادهایی که آنها بررسی می‌کنند، قانونمند و منظم هستند. هیچ رویدادی به خودی خود اتفاق نمی‌افتد. علم بر این باور بنا شده است که تمام پدیده‌های طبیعی دارای عوامل متقدم هستند. از این فرض غالباً به عنوان جبرگرایی (علت‌گرایی) جهانشمول یاد می‌شود. بشر اولیه برای بسیاری از رویدادهایی که مشاهده می‌کرد، دلایل ماورالطبیعه قائل می‌شد. چنانچه انسان به فراتر از تعبیرات ماورالطبیعه نظر نمی‌افکند و برای یافتن پاسخ سؤالات خود طبیعت را مورد مشاهده قرار نمی‌داد، علم جدید نمی‌توانست پیشرفت کند.

این فرض زمینه‌ساز این گفته است که تحت شرایط خاص رویدادهای معینی روی می‌دهند. برای مثال، یک شیمی‌دان می‌تواند اظهار کند که وقتی مخلوط کلرات پتاسیم و دی اکسید منگنز را گرم کنیم، اکسیژن تولید می‌شود. متخصصان علوم رفتاری نیز فرض می‌کنند که رفتار موجود زنده قانونمند و قابل کشف است.

این فرض نخست با این باور مرتبط است که رویدادها در طبیعت نهایتاً بر میزان، نظم و قاعده استوار هستند، که این نظم و قاعده طبیعی را می‌توان از طریق روش علمی کشف کرد. فرض دوم آن است که حقیقت، در نهایت از طریق مشاهده مستقیم قابل دریافت است. اتکاء به مشاهده مبتنی بر تجربه، علم را از غیر علم متفاوت می‌سازد. دانشمندان به منبع آگاه و مطلع به عنوان منبعی از حقیقت متکی نیست و در عوض به جمع‌آوری اطلاعات مرتبط معتقد است. در تاریخ علم به نمونه‌های زیادی از دانشمندان برخورد می‌کنیم که افکار غالب و چیره‌عصر خود را رد کرده و بر اساس مشاهدات و تجربیات خود به پیش رفته‌اند. تجربه‌های اولیه گالیله در زمینه سقوط اجسام دانش جدیدی به وجود آورد که با باورهای مسلم زمان خود در تضاد بود. یکی از نتایج این فرض این اعتقاد است که تنها پدیده‌هایی که در عالم هستی به طور واقعی قابل مشاهده هستند، در حیطه تحقیق علمی قرار می‌گیرند.

## // نگرش دانشمندان

دانشمندان ضمن کار خود به نگرشهای خاصی دست پیدا می‌کنند و آنها را مورد قبول قرار می‌دهند.

۱- یک دانشمند اساساً فردی شکاک است که در مورد داده‌های علمی قائل به شک می‌گردد. دانشمندان یافته‌های خود را آزمایشی تلقی می‌کنند و تا هنگامی که نتوان آنها را اثبات کرد، مورد پذیرش قرار نمی‌دهند. اثبات نتایج ایجاب می‌کند که سایرین نیز قادر به تکرار آزمایش و دستیابی به نتایج مشابه باشند. دانشمندان مایلند که تمامی عقاید مربوط به روابط بین پدیده‌های طبیعی را آزمایش کنند. بعلاوه، دانشمندان جریان آزمایش را به طریقی تشریح می‌کنند که دیگران هم بتوانند به همان ترتیب نتایج به دست آمده را اثبات کنند.

۲- دانشمند فردی طرفدار حقایق و بی‌غرض است و در استنتاج مشاهدات و اطلاعات مربوطه سعی در پافشاری بر موردی خاص و یا اثبات آن ندارد. او تمهیدات خاصی به کار می‌بندد تا اغراض شخصی احتمالی در مشاهدات او نفوذ نکنند. او واقعیت را جستجو می‌کند و آن را حتی اگر با عقایدش نیز متضاد باشد، می‌پذیرد. اگر دلایل و مدارک جمع‌آوری شده، نظریه مورد علاقه او را دگرگون سازد، او یا نظریه خود را رد می‌کند و یا آن را در جهت انطباق با داده‌های واقعی تغییر می‌دهد.

۳- دانشمند درگیر واقعیت‌هاست و نه ارزشها. او در یافته‌های خود هیچ استنباط اخلاقی بالقوه‌ای را آشکار نمی‌کند. او برای ما تصمیم نمی‌گیرد که چه چیز خوب و چه چیز بد است. دانشمند داده‌های مربوط به روابط موجود بین رویدادها را فراهم می‌آورد، اما اگر بخواهیم درباره مطلوبیت پیامد خاصی تصمیم بگیریم، باید از این داده‌های علمی فراتر رویم. بنابراین، هر چند که یافته‌های علمی شاید در حل یک مشکل مرتبط با تصمیم‌گیری شخصی اهمیت اساسی داشته باشند، اما داده‌ها، خود، آن قضاوت شخصی را فراهم نمی‌آورند.

۴- دانشمند با واقعیت‌های منفرد قانع نمی‌شود، بلکه جستجو می‌کند تا یافته‌هایش را جامع و نظام‌مند کند. او می‌خواهد موارد شناخته شده را در سازمانی منظم قرار دهد. بنابراین دانشمند در پی نظریه‌هایی است که این یافته‌های تجربی را در طرحی معنادار در کنار هم قرار دهد. به هر حال، او این نظریه‌ها را آزمایشی و موقتی تلقی می‌کند؛ به طوری که با یافتن شواهد جدید مورد تجدید نظر قرار می‌گیرند.

## // نظریه علمی

آخرین جنبه از مفهوم علم که مورد توجه قرار می‌گیرد، ساخت نظریه است. دانشمندان طی تحقیقات تجربی خود، بسیاری از واقعیات را جمع‌آوری می‌کنند، اما وقتی که این واقعیات جمع‌آوری شدند، برای آنکه از یافته‌های منفرد به اطلاعاتی با معنا تبدیل شوند، نیاز به ادغام، سازماندهی و طبقه‌بندی دارند. روابط مهم در داده‌ها باید مشخص و تعریف شوند، به عبارت دیگر نظریه‌ها می‌باید صورت‌بندی شوند. یک نظریه را این‌گونه تعریف می‌کنند: مجموعه‌ای از ساختهای (مفاهیم) دارای پیوند درونی، تعاریف و قضایا که از طریق مشخص کردن روابط موجود بین متغیرها و به منظور تشریح و پیش‌بینی پدیده‌ها، دید نظام‌مندی از آنها ارائه می‌کند.<sup>(۱)</sup>

نظریه‌ها نتایج مشاهدات را به هم پیوند می‌دهند و دانشمندان را قادر می‌سازند که یک اظهار نظر عمومی در مورد متغیرها و روابط بین آنها ارائه کنند. برای مثال، زمانی که حرارت گاز هیدروژن از ۲۰ درجه به ۴۰ درجه سانتیگراد افزایش یابد و فشار ثابت نگه داشته شود، مشاهده می‌شود که گاز هیدروژن منبسط می‌گردد. همچنین در شرایطی که فشار ثابت نگه داشته شود و حرارت گاز اکسیژن از ۶۰ درجه به ۵۰ درجه سانتیگراد کاهش یابد، گاز اکسیژن منقبض می‌شود. یک نظریه معروف، یعنی قانون چارلز، اثرات مشاهده شده در تغییرات درجه حرارت بر روی حجم تمامی گازها را این‌گونه بیان می‌کند: در شرایطی که فشار ثابت است، همانگونه که حرارت گاز افزایش می‌یابد، بر حجم آن نیز افزوده می‌شود و بالعکس زمانی که حرارت گاز کاهش می‌یابد، از حجم آن نیز کاسته می‌شود. این نظریه تنها اطلاعات موجود در این زمینه را خلاصه نمی‌کند، بلکه پدیده‌های دیگری را با تبیین این مسئله که باید از هر گازی، در هر درجه حرارت، انتظار چه نوع تغییری را داشت، پیش‌بینی می‌کند.

هدف نهایی و اصلی علم، ساختن نظریه است. این نظریه برای کسانی که یک نظریه را همچون گمانی پوچ و مبهم و یا اندیشه‌ای غیر عملی می‌پندارند، به نظر عجیب خواهد آمد؛ هر چند یک نظریه علمی تعریفی مبتنی بر تجربه از یک پدیده به دست می‌دهد. از یک چنین تعریفی می‌توانیم به سمت پیش‌بینی و در نهایت کنترل آن پدیده حرکت کنیم. به مجرد ساخته



شدن یک نظریه، فی‌المثل در مورد رابطه بین پشه آنوفل و بیماری مالاریا در انسان، دانشمند می‌تواند (۱) توضیح دهد که چرا بیماری مالاریا در منطقه‌ای همه‌گیر است و در منطقه دیگری نیست، (۲) پیش‌بینی کند که چگونه تغییرات محیط با تغییرات در شیوع مالاریا همراه خواهد بود و (۳) به وسیله ایجاد تغییراتی در محیط، مالاریا را کنترل کند.

### / اهداف نظریه‌ها

در جریان پیشرفت هر علم، نظریه‌ها در نیل به اهداف متعددی به کار گرفته شده‌اند و در اولین وهله، نظریه، دانش موجود در زمینه خاصی را خلاصه می‌کند و به نظم در می‌آورد. برای مثال یک نظریه یادگیری، نتایج بسیاری از بررسیهای پراکنده در فرآیند یادگیری را به شیوه‌ای منسجم به یکدیگر پیوند می‌دهد. افزون بر این، نظریه مفاهیمی را در رابطه با یافته‌های تجربی و منفرد گذشته روشن و ساخته و پرداخته می‌کند.

نظریه، توصیفی برای رخدادهای مشاهده شده و روابط آن فراهم می‌آورد. همچنین نشان می‌دهد که چه متغیرهایی و چگونه با هم رابطه دارند. یک نظریه یادگیری روابط بین سرعت و کفایت یادگیری و یا عواملی از قبیل انگیزه، پاداش، تمرین و امثال آن را تشریح می‌کند. بر اساس اصول توصیفی ارائه شده از طریق نظریه‌ها، استنتاجات قیاسی امکان پیش‌بینی وقوع پدیده‌هایی را نیز که تاکنون مشاهده نشده‌اند، فراهم می‌آورند. برای نمونه، با کمک نظریه، ستاره شناسان وجود سیارات را در دورترین نقاط، قبل از آنکه به طور واقعی مورد مشاهده قرار گیرند، پیش‌بینی می‌کنند.

نظریه با فراهم آوردن راهبردهایی برای تحقیقات بیشتر جهت پیشرفت علوم جدید، ایجاد انگیزه می‌کند. از طریق نظریه، دانشمندان استنتاجات مربوط به این مسئله را که در موقعیت خاص و تحت شرایطی ویژه چه واقعه‌ای اتفاق خواهد افتاد، انجام می‌دهند. از طریق نظریه‌های تقویت، فرد می‌تواند استنتاجی در مورد تأثیرات پاداش در ایجاد یک پاسخ منظم، در مقابل پاسخهای تصادفی و اتفاقی، به عمل آورد. این استنتاجات قیاسی فرضیاتی را برای تحقیقات علمی فراهم می‌آورند، که اطلاعات حاصل از آن به پذیرش، رد یا تغییر نظریه‌هایی که برای آزمایش طراحی شده‌اند، منجر می‌شود. به این سیاق دانشمندان از دانسته‌های خود به عنوان سکوی پرشی برای دستیابی به پیشرفتهای تازه بهره‌برداری می‌کنند.

## / روشهای ساخت نظریه

ملوین مارکس<sup>(۱)</sup> چهار رویکرد اصلی را در ساخت نظریه شناسایی کرده است: (۱) مدل که تأکید آن عمدتاً بر مفهوم سازی است، (۲) رویکرد قیاسی، که با تأکید اولیه بر مفهوم سازی رسمی هم عناصر مفهومی و هم عناصر تجربی را در بر می گیرد، (۳) رویکرد کارکردی، که در آن بر عناصر مفهومی و تجربی به طور یکسان تأکید می شود، و (۴) رویکرد استقرایی که در آن تلاش مطلقاً بر متحدالشکل کردن دانش تجربی است.

## مدل

مارکس مدل را به عنوان یک «آنالوگ مفهومی که به طور کلی ماهیت فیزیکی یا ریاضی دارد، تعریف می کند، که برای پیشنهاد تحقیق تجربی مورد استفاده قرار می گیرد.» مدلها شیوه های مفیدی برای تفکر در مورد پدیده های پیچیده هستند. آنها پیچیدگی ها را به شیوه ای ساده ارائه می کنند و قابل فهم تر می سازند. مدلها با جمع آوری داده های تجربی تغییر نمی یابند؛ آنها در صورتی که داده ها آنها را تأیید کنند، باقی می مانند و در صورت عدم تأیید داده ها کنار گذاشته می شوند.

در علوم رفتاری نمونه های بیشماری از کاربرد آنالوژی را در صورت بندی مدلها می توان یافت. یک مثال قدیمی نظریه معروف باز پیدایی رشد انسان جی. استانلی هال است،<sup>(۲)</sup> که می گوید: رشد اجتماعی در فرد، به گونه ای، تکرار تاریخ بشر است. بر این اساس، کودکان در سنین اولیه رشد خود مانند انسانهای اولیه رفتار می کنند و در راه خود برای متمدن شدن، به تدریج از حالت بربریت و مراحل نیمه متمدن عبور می کنند. هر چند این تئوری در حال حاضر مطرود شده است، ولی در دهه های نخست این قرن نقش مهمی در برانگیختن تحقیق در حیطه رشد کودکان و بزرگسالان ایفا کرده است. در حال حاضر از مدلهای کامپیوتری برای کمک به محققان در تکمیل نظریه های دقیق تر در زمینه چگونگی یادگیری، مسئله گشایی، تفکر و امثالهم، وسیعاً استفاده می شود. مدلهای کامپیوتری را می توان به گونه ای طراحی کرد که کم و بیش به طریق مشابهی به کودکان «بیاموزند»، به این ترتیب درک ما از این که کودکان چگونه و چه چیزی را می توانند بیاموزند، ارتقا پیدا می کند.

1- Marx, M. (1965). *Theories In Contemporary Psychology* (P.43). New York: Macmillan

2- Hall, G. S. (1904). *Adolescence* (vols. 1 and 2). New York: Appleton.

### رویکرد قیاسی

برای ساختن نظریه، از واقعتهای شناخته شده آغاز می‌کند و از طریق استدلال قیاسی به یک سیستم مفهومی کاملاً تدوین یافته می‌رسد. رویکرد قیاسی مبتنی بر این فرض است که فرد برای تبیین یک پدیده در یک حوزه خاص باید از دانسته‌های موجود فزاینده‌تر رود تا یک نظام منسجم منطقی ایجاد کند؛ حتی اگر در مورد اعتبار بسیاری از مفروضه‌ها و قیاسهای درون سیستم تردیدهایی وجود داشته باشد.

در رویکرد قیاسی تأکید بسیار زیادی بر مفهوم سازی وجود دارد، یعنی ابتدا یک سیستم نظریه‌ای کاملاً تدوین شده تهیه می‌شود، سپس از این نظریه، فرضیه‌هایی برای آزمون کردن استخراج می‌شود.

اطلاعات واقعی جدید که از طریق آزمونهای تجربی به دست می‌آید، برای جرح و تعدیل نظریه مورد استفاده قرار می‌گیرند.

رویکرد قیاسی، هر چند به دلیل ذهنی بودنش به لحاظ ساختن یک سیستم نظریه‌ای دقیق بر اساس اطلاعات واقعی محدود مورد انتقاد قرار دارد، ولی در پیشرفت علوم رفتاری از طریق برانگیختن بررسیهای تحقیقاتی بیشمار، نقش مهمی ایفا کرده است.

### رویکرد کارکردی

بسیاری از دانشمندان بر این باورند که اشتغال ذهنی به سیستم‌های نظریه‌ای کاملاً تدوین شده، در مرحله‌ای که داده‌های تجربی در دسترس نیستند یا محدود هستند، شاید موجب شود که فرد بر روی برخی تبیین‌های ممکن تثبیت شود و بنابراین جستجو برای تبیین‌های جایگزین آن پدیده را به مخاطره افکند. آنها رویکرد کارکردی را، که مشخصه آن غیر رسمی بودن و انعطاف پذیری به درجات زیاد است، ترجیح می‌دهند. نظریه‌های آزمایشی و اغلب ناقص برای تبیین پدیده‌ها پیشنهاد می‌شوند. فرضیه‌های مبتنی بر این نظریه‌های آزمایشی، سپس، صورتبندی و به طور تجربی آزمون می‌شوند.

اگر نتایج این آزمونها نظریه را کاملاً تأیید نکند، آن نظریه مورد تجدید نظر قرار می‌گیرد و فرضیه‌های جدیدی از نظریه‌های تجدید نظر شده استنتاج می‌شود. یک رویکرد کارکردی می‌تواند به عنوان مرحله‌ای از تأثیر متقابل بین مفهوم‌سازی و آزمون تجربی فرضیه ترسیم شود.

برخی عقیده دارند که ماهیت التقاطی رویکرد کارکردی منجر به محدود شدن حوزه تحقیقات جدید می‌شود. با این حال رویکرد کارکردی نقش مهمی در پیشرفت علوم رفتاری ایفا کرده است.

### رویکرد استقرایی

در رویکرد استقرایی سعی می‌شود که تنها گفتارهایی ارائه شود که روابط تجربی بین پدیده‌ها را منسجم و خلاصه کند. استدلال قیاسی و منطق رسمی کمترین نقش را در رویکرد استقرایی دارند. حتی صورتبندی یک فرضیه تحقیقاتی قبل از انجام آزمایش در نظریه‌های استقرایی مطلوب نیز به دلیل ایجاد سوگیری در تلاشهای تحقیقاتی، غیرضروری محسوب می‌شود. برای این دانشمندان، نظریه، غایت و نتیجه همه تلاشهای علمی است و باید تنها از طریق استقرا به دست آید.

کارهای بی. اف. اسکینر و پیروانش در روان‌شناسی آزمایشی نمونه‌هایی از این رویکرد در ساختن فرضیه است. این دانشمندان از بکارگیری هر گونه قیاس و منطق رسمی در تحقیقات علمی خود پرهیز می‌کنند. نظریه رفتاری بی. اف. اسکینر فرض می‌کند که تنها از رفتار قابل مشاهده در پایه‌ریزی علم رفتار انسان می‌توان سود جست و به طرح حالت‌های درونی انسان، از قبیل انگیزه‌ها، ارزشها و دانش شناختی، نیازی نیست.

این دیدگاه افراطی به دلیل رد فرایندهایی که بسیاری، آنها را در کارهای علمی اجتناب‌ناپذیر می‌دانند، مورد انتقاد قرار گرفته است. منتقدان رویکرد استقرایی صرف همچنین عقیده دارند که انتخاب مسئله تحقیق، طرح تحقیقاتی و شیوه تحقیق، در نهایت یک نظریه‌سازی غیررسمی و استدلال قیاسی را در بر دارد.

### / ویژگیهای نظریه

اگر هدف یک نظریه خدمت به علم باشد، باید دارای معیارهای چندی باشد. اینها مواردی از صفات مشخصه یک نظریه صحیح و منطقی است:

۱- یک نظریه باید قادر به توصیف واقعیات مشاهده شده در مورد مسئله‌ای خاص باشد؛ همچنین باید توانایی ارائه «چرایی» پدیده مورد بررسی را داشته باشد. یک چنین توصیفی از

رخدادهای می‌باید در ساده‌ترین شکل ممکن انجام شود. یک نظریه با پیچیدگی و مفروضات کمتر، نسبت به نوع پیچیده‌تر آن مطلوب‌تر است. این امر را قانون ایجاز می‌نامند.

۲- یک فرضیه باید با واقعیات مشاهده شده و با پیکره دانش موجود همساز باشد. ما در جستجوی نظریه‌ای هستیم که محتمل‌ترین و کاراترین شیوه توجیه واقعیتهای انباشته‌شده را فراهم آورد.

۳- یک نظریه باید برای اثبات خود راههایی را ارائه کند؛ یعنی یک نظریه باید امکان انجام قیاسهایی را به شکل فرضیه‌ها فراهم سازد و این فرضیه‌ها پیامدهایی را که محقق، در صورت درستی نظریه، انتظار مشاهده آنها را دارد، تعیین کنند. دانشمندان سپس می‌توانند این فرضیات را به طور تجربی در جهت تعیین آنکه آیا داده‌ها، نظریه را تأیید می‌کنند یا نه، مورد بررسی قرار دهند. این نکته می‌باید مورد تأکید قرار گیرد که صحبت از درستی و یا بطلان یک نظریه نایجابست. پذیرش و یا رد یک نظریه ابتدا بستگی به سودمندی آن دارد. سودمندی یک نظریه بسته به این است که با چه کفایتی پیامدهای قابل مشاهده‌ای را که بعداً از طریق گردآوری داده‌های تجربی تأیید خواهند شد، پیش‌بینی می‌کند. پس بنابراین هر نظریه‌ای آزمایشی تلقی می‌شود و با اثبات شواهد جدید مورد تجدید نظر قرار می‌گیرد.

۴- یک نظریه باید کشفیات تازه‌ای را برانگیزد و پهنه‌های بیشتری را در امر تحقیق بنمایاند. نظریه‌ها در زمینه علوم فیزیکی به نسبت علوم اجتماعی، به‌طور بسیار وسیعتری به اهداف خود دست پیدا کرده‌اند. این مسئله شگفت‌انگیزی نیست؛ چرا که آنها علوم قدیمی تری هستند. در آغاز هر علمی، تأکید ویژه بر تجربه‌گرایی است. دانشمندان درگیر گردآوری واقعیات در مورد مسائل ویژه هستند. انسجام دانشهای مجزا در قالب چهارچوب نظریه‌ای، تنها با بلوغ علم آغاز می‌شود.

تعلیم و تربیت به طور اخص از فقدان سوگیرهای<sup>(۱)</sup> نظریه‌ای در رنج است؛ چرا که تأکید اصلی در آن بر تجربه‌گرایی بوده است. کارشناسان تعلیم و تربیت به دلیل آنکه به «کسب واقعیات» بیش از یافتن «چرایی» آنها علاقه‌مندند، مورد انتقاد قرار گرفته‌اند. این علاقه در شواهد متعددی که در جریان بررسیهای مربوط به تعلیم و تربیت به دست آمده و هیچگاه به طور همبسته و به صورت یک نظریه برای توصیف مسائل تعلیم و تربیت به کار گرفته نشده است،

منعکس است. تعلیم و تربیت همانگونه که به هدایت تلاشهایش در سطحی تجربی نیاز دارد، به همان ترتیب نیز نیازمند تمرکز توجهات بیشتر در ارائه نظریه‌ها، جهت کسب دور نماهای گسترده‌تر در زمینه مسائل تعلیم و تربیت است.

به این ترتیب با آنکه در تعداد و قدرت نظریه‌هایی که در علوم فیزیکی و اجتماعی تبیین شده است، تفاوت‌های بارزی وجود دارد، ولی نظریه، در جریان پیشرفت هر علمی نقش مشابهی ایفا می‌کند. به عبارت دیگر نظریه، بدون در نظر گرفتن موضوع آن، در همه علوم به طریقی مشابه عمل می‌کند. نظریه، دانش موجود را به منظور تبیین رخدادهای مشاهده شده و روابط بین آنها و پیش‌بینی وقوع رخدادهای مشاهده نشده و روابط بین آنها، خلاصه می‌کند. می‌توان گفت که نظریه‌ها بهترین و ارزنده‌ترین کوششهای بشری را در فهم ساختمان اصلی جهانی که در آن زندگی می‌کنیم، ارائه می‌کنند.

### // محدودیتهای روش علمی در علوم اجتماعی

به رغم فایده روش علمی و گردآوری مقدار معتناهی از دانش معتبر در تعلیم و تربیت و دیگر علوم اجتماعی، این زمینه‌ها همانند علوم طبیعی به پیکره‌ای علمی دست پیدا نکرده‌اند. علوم اجتماعی از نظر قدرت تبیین و ظرفیت ارائه پیش‌بینی‌های دقیق قادر به تبیین تعمیم‌هایی همپای نظریه‌های علوم طبیعی نبوده است. اغلب، در بین محققان علوم اجتماعی درباره این که واقعیتهای مسلم (پذیرفته شده) چیست یا چه توصیف قانع‌کننده‌ای برای واقعیت‌های فرضی وجود دارد، اتفاق نظر دیده نمی‌شود. شاید علوم اجتماعی هرگز نتوانند همانند علوم طبیعی به اهداف علم به طور کامل دست یابند. این مسئله باید به صراحت مورد تاکید قرار گیرد که کاربرد روش علمی به خودی خود شرایط موثر را برای پیشرفتهای علمی مهیا نمی‌کند. محدودیتهای متعددی در بکارگیری رویکرد علمی در تعلیم و تربیت و سایر علوم اجتماعی وجود دارد.

### / پیچیدگی موضوعات مورد نظر

یکی از محدودیتهای اصلی، پیچیدگی ذاتی موضوعات مطرح شده در علوم اجتماعی است. علوم طبیعی با پدیده‌های فیزیکی و بیولوژیکی سرو کار دارند. در این علوم تعداد محدودی از متغیرها که قابلیت اندازه‌گیری دقیق دارند، در توصیف بسیاری از پدیده‌ها دخیل هستند و نیز

تبین قوانین جهانشمول امکان پذیر است. برای مثال، قانون بویل<sup>(۱)</sup> درباره اثر فشار در حجم گازها با متغیرهای نسبتاً ساده‌ای سروکار دارد و روابط بین پدیده‌هایی را که آشکارا در جهان تغییرناپذیرند، صورت‌بندی می‌کند.

از طرف دیگر، کارشناس علوم اجتماعی با موضوعات انسانی درگیر است. او علاقه‌مند به رفتار فرد و رشد او به عنوان عضوی از گروه است. متغیرهای بسیار زیادی وجود دارند که به طور مستقل و یا درکنش با متغیرهای دیگر عمل می‌کنند و می‌باید در هر تلاش برای فهم پیچیدگی رفتار انسانی در نظر گرفته شوند. هر فرد در جریان رشد خود، در دستگاه ذهنی، در رفتار اجتماعی و عاطفی و در شکل کلی شخصیتش انسانی منحصر به فرد است. رفتار فرد در گروه و اثر رفتار اعضای گروه بر فرد نیز از طرف کارشناس علوم اجتماعی می‌باید مورد نظر قرار گیرد. گروه‌های مختلف دانش‌آموزان دبستانی در موقعیتهای یکسان مانند هم رفتار نخواهند کرد. فراگیران، آموزگاران و محیطهای مختلف، هر یک با متغیرهایی در بروز یک پدیده رفتاری دخیل هستند. از اینرو محقق باید در امر تعمیم فوق‌العاده محتاط باشد؛ چراکه اطلاعات گردآوری شده در یک شرایط گروهی ممکن است برای گروه دیگر و شرایط دیگر دارای اعتبار نباشند.

### / دشواریهای مشاهده

انجام مشاهده، یعنی جزء لاینفک<sup>(۲)</sup> علوم، در علوم اجتماعی مشکل‌تر و پرمخاطره‌تر از علوم طبیعی است. مشاهده در علوم اجتماعی بیشتر ذهنی و وابسته به طرز تفکر شخصی است؛ چراکه این علوم اغلب درگیر تعبیرات مشاهده‌گران هستند. برای مثال، موضوع مورد نظر برای مطالعه، اغلب پاسخهای انسان به رفتارهای دیگران است. انگیزه‌ها و ارزشها و نگرشها قابل واریسی نیستند. زمانی که مشاهده‌گران می‌خواهند تعیین کنند که رفتار مشاهده شده نشانه‌ای از یک انگیزه، ارزش یا نگرش خاص است، باید به تعبیر ذهنی خود اتکا کنند. این مشکل دانشمندان علوم اجتماعی است که ارزشها و نگرشهایش، هم در مشاهداتش در پدیده اجتماعی و هم در روش ارزیابی یافته‌هایش به منظور پایه‌ریزی نتایج، اثر می‌گذارد. دانشمندان علوم طبیعی پدیده‌هایی را بررسی می‌کنند که کمتر به تعبیرات ذهنی نیاز دارد.

## / دشواریهای تکرار تجربه

یک شیمیدان می‌تواند به طور عینی واکنش بین دو ماده شیمیایی را در لوله آزمایش بررسی کند. او می‌تواند یافته‌هایش را گزارش کند و مشاهداتش نیز می‌توانند به راحتی از طرف دیگران تکرار شوند. انجام این امور در علوم اجتماعی به مراتب مشکل‌تر است. یک کارشناس تعلیم و تربیت آمریکایی نمی‌تواند کاملاً شرایط روش تجربی تدریس یک کارشناس تعلیم و تربیت روسی را بازسازی کند. حتی در داخل ساختمان مدرسه‌ای واحد نیز فرد نمی‌تواند یک شرایط معین را به طور کامل و دقیق باز آفرینی کند. پدیده‌های اجتماعی رخدادهای منحصر به فردی هستند که نمی‌توانند به منظور مشاهده تکرار شوند.

## / تعامل مشاهده‌گر و آزمودنی‌ها

مشکل دیگر آن است که صرف مشاهده پدیده اجتماعی ممکن است در آن تغییراتی ایجاد کند که در غیر این صورت امکان وقوع آن میسر نمی‌بود. محقق ممکن است فکر کند که  $X$  علت  $Y$  است، ولی در واقع محتمل است که مشاهده  $X$  باعث  $Y$  شده باشد. این مسئله به وسیله تجربیات مشهور «هاوثرن»<sup>(۱)</sup> روشن شده است که تغییراتی در بازدهی کارگران نه به دلیل شرایط متنوع کاری، بلکه در واقع تنها به این دلیل بوده است که کارگران می‌دانستند برای انجام بررسی انتخاب شده‌اند. محقق یک انسان است و طرز تفکرش به عنوان یک مشاهده‌گر در شرایطی خاص ممکن است رفتار فرد مورد مطالعه‌اش را تغییر دهد. استفاده از دوربین مخفی و ضبط صوت شاید به تخفیف این مسئله در بعضی موارد کمک کند، ولی اکثر تحقیقات در علوم اجتماعی متضمن پاسخگویی مستقیم فرد مورد مطالعه به مشاهده‌گر است.

## / دشواریهای کنترل

میزان و حدود امکانات کنترل تجربیات در موضوعات مورد نظر در علوم اجتماعی نسبت به علوم طبیعی بسیار محدودتر است. پیچیدگی موجود در پژوهش موضوعات انسانی، نمایشگر دشواریهای نابرابر کنترل نسبت به علوم طبیعی است؛ چرا که در نهایت، کنترل دقیق شرایط تجربی در آزمایشگاه امکان‌پذیر است. یک چنین کنترلی در مورد مسائل انسانی



امکان پذیر نیست؛ دانشمندان علوم اجتماعی باید همزمان با متغیرهای زیادی درگیر باشند و اجباراً تحت شرایطی کار کنند که از دقت بسیار کمتری برخوردار است. او سعی می‌کند که بسیاری از این متغیرها را تا آنجا که ممکن است شناسایی و کنترل کند، ولی انجام این کار گاهی اوقات بسیار مشکل است.

### / مسائل اندازه‌گیری

آزمایش باید شرایط لازم را برای اندازه‌گیری عوامل مورد نظر فراهم آورد. وسایل اندازه‌گیری در علوم اجتماعی، نسبت به علوم طبیعی، از دقت و خطاناپذیری بسیار کمتری برخوردارند. در علوم اجتماعی ما هیچ وسیله‌ای نداریم که با دقت یک خط کش، یک ترمومتر و یا وسایل بی‌شمار آزمایشگاهی قابل مقایسه باشد. قبلاً اشاره کردیم که فهم یک رفتار انسانی به دلیل دخالت متغیرهای تعیین‌کننده زیادی که به طور مستقل و در تعامل با یکدیگر عمل می‌کنند، پیچیده است. روشهای چند متغیری موجود به طور نسبی ناظر بر تعداد کمی از عوامل هستند که به طور وضوح در رابطه متقابل با یکدیگر قرار دارند. به علاوه این روشها به فرد امکان می‌دهند که واریانس را تنها به عواملی که هنگام اندازه‌گیری عمل می‌کنند، تخصیص دهد. عواملی که در گذشته بر روی جریان رشد مؤثر بوده‌اند و هنوز هم در جریان آن اثر قابل توجهی دارند، در حال حاضر قابل اندازه‌گیری نیستند.

از آنجا که تحقیق در علوم اجتماعی به واسطه این عوامل، پیچیده است، محقق باید در تعمیم نتایج جانب احتیاط را رعایت کند. اغلب ضروری است که قبل از تعمیم نتایج، بررسیهای متعددی را در زمینه‌ای خاص هدایت کرد. تنها اگر یافته‌های اولیه به طور مداوم مورد تأیید قرار گیرند، محقق می‌تواند با اطمینان بیشتر، تعمیم‌های گسترده‌تری انجام دهد.

به رغم محدودیتها، تعلیم و تربیت و علوم اجتماعی پیشرفت مهمی کرده‌اند و انتظار می‌رود که پیکره علمی آنها همانند بررسیهای علمی و روش‌شناسی، نظام‌مندتر و منسجم‌تر رشد یابند.

### /// ماهیت تحقیق

نخست، شاید بهتر باشد که این سؤال را مطرح کنیم که تحقیق چیست؟ تحقیق را شاید بتوان به عنوان «بکارگیری روش علمی به منظور بررسی یک مسئله» تعریف کرد. این روشی است

برای دستیابی به اطلاعات قابل اعتماد و مفید و هدف آن کشف پاسخهایی برای سؤالات پرمعنا از طریق به کارگیری روشهای علمی است.

یک بررسی می‌باید با روشهای علمی تعریف شده سر و کار داشته باشد، به طریقی که در دسته بندیهای اساسی بتواند به عنوان یک تحقیق طبقه‌بندی شود. اگر چه این بررسی ممکن است در شرایط متفاوت و با بکارگیری روشهای متفاوت صورت گیرد، ولی تحقیق به طور عام، تلاشی نظام‌مند و هدفمند برای کسب دانش معتبر و قابل اطمینان است.

### // تحقیق در تعلیم و تربیت

نتیجه به کار بردن روش علمی در مسائل تربیت و تعلیم، تحقیق در تعلیم و تربیت است. تحقیق در تعلیم و تربیت جریانی است که از طریق آن فرد به اطلاعات معتبر و مفیدی در زمینه فرآیند تعلیم و تربیت دست می‌یابد. هدف آن کشف اصول و یا تفاسیر کلی درباره رفتار است که می‌توان از آنها در توصیف، پیش‌بینی و کنترل رویدادها در زمینه تعلیم و تربیت و به عبارت دیگر در نظریه علمی استفاده به عمل آورد.

پذیرش روش علمی در تعلیم و تربیت و سایر علوم اجتماعی نسبت به پذیرش آن در علوم تجربی دیرتر صورت پذیرفت. در سال ۱۸۹۷ جی ام رایس<sup>(۱)</sup> پیشرو در تحقیق مسائل تعلیم و تربیت، خود را در شرایطی همانند شرایطی که به وسیله بیکن پیش از این تشریح شده بود، یافت. رایس، از متخصصان تعلیم و تربیت، در اجلاس سالیانه مدیران، این سؤال را مطرح کرد که: «آیا می‌توان تعیین کرد که دانش‌آموزانی که در روز ۴۰ دقیقه از وقت خود را به هجی کردن اختصاص می‌دهند، از دانش‌آموزانی که ۱۰ دقیقه در روز به این کار می‌پردازند، بیشتر یاد می‌گیرند؟» او گزارش می‌کند:

«... با کمال تعجب این سؤال جمع را مبهوت ساخت. اولین پاسخ دهنده، پروفیسور محبوبی در زمینه روان‌شناسی بود که در غرب به تربیت مدرس اشتغال داشت. او گفت در واقع، این سؤالی است که انسان هرگز قادر به پاسخگویی به آن نخواهد بود؛ و مرا نیز به خاطر طرح یک چنین سؤالات مضحکی که تنها باعث اتلاف وقت یک چنین مجمع پراهمیتی شده بود، بسختی سرزنش کرد.»<sup>(۲)</sup>

1- J. M. Rice

2- Joseph M Rice Scientific Management in Education (New York): Hinds, Noble and Eldredge, (1912) PP.17. 18.

ولی رایس این کار را انجام داد. در واقع برای پاسخگویی به سؤالش مدارکی تجربی جمع‌آوری کرد و دریافت که در پیشرفت تحصیلی کسانی که ۱۰ دقیقه در روز را صرف خواندن می‌کنند و آنهایی که چهل دقیقه در روز به این کار می‌پردازند، تفاوتی جزئی موجود است. او همچنین یادآور می‌شود که بسیاری از کلماتی که کودکان باید هجی کردن آن را بیاموزند، دارای ارزش کاربردی ناچیزی هستند. کار او به سایر محققان، از جمله ای. ال. ثورندایک، این امکان را داد که برای تعیین فراوانی استفاده از کلمات در زبان ما، از تجزیه و تحلیل اسنادی<sup>(۱)</sup> استفاده کنند. کار آنها به همین ترتیب امکان اصلاحاتی را در متون هنری و مطالب درسی فراهم آورد. بنابراین، اگر چه تحقیق در تعلیم و تربیت دانش جدیدی است، ولی از زمان آغاز آن در اواخر قرن نوزدهم تاکنون پیشرفت بسیاری کرده است.

تحقیق در تعلیم و تربیت به عنوان یک علم، از روشهای تحقیقاتی بهره می‌گیرد که با جریانها و روشهای اساسی و مفاهیم کاربردی علم انطباق دارند. با این توضیح، تحقیق دارای مراحل مختلفی است.

### // مراحل متداول در تحقیق / انتخاب یک مسئله

محقق بررسی را با سؤالی آغاز می‌کند که از نظر او به آن اندازه اهمیت دارد که جریان تحقیق را توجیه کند؛ و آن باید سؤالی باشد که پاسخ آن اگر چه هنوز به دست نیامده، ولی از طریق روشهایی چون مشاهده و تجربه قابل دستیابی است. به این سؤال توجه کنید: «آیا کودکانی که از طریق روش تدریس بنیادی الفبا<sup>(۲)</sup> با خواندن آشنا شده‌اند، نسبت به کودکانی که خواندن را با روش دیداری آموخته‌اند، از نظر مهارت در خواندن نمره بالاتری می‌گیرند؟» این مسئله را می‌توان به طور تجربی، با مقایسه نمره‌های دو گروه همسان در آزمون پیشرفت خواندن که یکی از طریق روش الفبایی و دیگری با روش دیداری با خواندن آشنا شده‌اند، مورد بررسی قرار داد.<sup>(۳)</sup>

1- Documentary Analysis

۲- روش تدریس بنیادی الفبا (Initial Teaching Alphabet) روش تدریس خواندن در کلاس اول ابتدایی است که در آن دانش‌آموزان اسامی حروف کلمه را با صدای بلند هجی می‌کنند.

3- Goodman, K. (1986). What's Whole in whole Language? Portsmouth, NH: Heinemann.

۱۲۰۲۴۳۲۱  
۸۰۴۱۸



سوالات پرمعنایی وجود دارند که از طریق به کارگیری روش علمی نیز قابل پاسخگویی نیستند. برای نمونه این سؤال که آیا تدریس از طریق روش الفبایی برای دانش‌آموزان مفید است، مادامی که منظور از مفید بودن برای دانش‌آموز روشن نباشد، یا مشاهده و یا اندازه‌گیری آن امکان‌پذیر نباشد، قابل بررسی علمی نخواهد بود. بنابراین کلماتی که متضمن قضاوت‌های ارزشی هستند، نباید در تعریف یک مسئله به کار گرفته شوند.

### / اقامه فرضیه

بعد از آنکه مسئله مشخص شد، بررسی جامعی از تحقیقاتی که احتمالاً در مورد مسئله انجام شده است، صورت می‌گیرد. مرور تحقیقات قبلی برای ایجاد بینش نسبت به مسئله و ایجاد زمینه برای صورتبندی فرضیه بررسی، ضروری است. بهترین راهنما برای ارائه فرضیه‌ای هوشمندانه، تجزیه و تحلیل محتاطانه اطلاعات موجود و مربوط به مسئله است. اگر به مثال قبل برگردیم، محقق پس از بررسی دقیق تحقیقات پیشین ممکن است این فرضیه را اقامه کند: «برای یادگیری در امر خواندن، روش بنیادی الفبا کاراتر از روش دیداری است.» فرضیه، در تحقیقاتی که روابط بین متغیرها را بررسی می‌کنند، به کار گرفته می‌شود. در انواع دیگر تحقیق، پژوهشگر ممکن است بررسی را با طرح یک سؤال آغاز کند. این امر بویژه در بیشتر تحقیقات پیمایشی مصداق دارد.

به غیر از صورتبندی یک فرضیه، در این مرحله محقق همچنین تعریف اصطلاحاتی را که در بررسی مورد استفاده قرار خواهد داد، روشن می‌سازد.

### / انتخاب راهبرد تحقیق و تهیه ابزار آن

در فرایند استنتاج قیاسی، مفاهیم ضمنی فرضیه پیشنهادی - یعنی چه پیش خواهد آمد اگر فرضیه درست باشد - تعیین می‌شود. اگر این فرضیه درست باشد که تدریس خواندن با روش بنیادی الفبا باعث پیشرفت بیشتری نسبت به تدریس خواندن از طریق روش دیداری می‌شود، محقق می‌باید مشاهده کند که نمرات پیشرفت خواندن در بین دانش‌آموزانی که خواندن را با روش الفبایی آموخته‌اند، نسبت به دانش‌آموزان همطرازشان، که آن را با روش دیداری آموخته‌اند، بیشتر است.

این امر با انتخاب یک روش تحقیقاتی معین دنبال می‌شود. برخی از مسائل، مستلزم انجام آزمایش هستند؛ مسائل دیگر نیز ممکن است نیازمند یک راهبرد توصیفی باشند. پس انتخاب یک روش تحقیقاتی، جزئیات طرح تحقیق و روشهای اندازه‌گیری متغیرها را تحت تأثیر قرار می‌دهد. ابزار اندازه‌گیری متغیرها به عنوان وسایلی استاندارد ممکن است از قبل وجود داشته باشد، و یا می‌باید توسط محققان ساخته شود.

### /گردآوری و تفسیر داده‌ها

نتایج قیاسهای حاصل از فرضیه‌های بررسی باید آزمون شوند. بنابراین در این مرحله گردآوری داده‌ها صورت می‌گیرد که شامل جنبه‌های عادی به اجرا گذاردن ابزار تحقیق، ثبت اطلاعات، تهیه جداول و غیره است. بر خلاف باور عمومی، این مرحله معمولاً نسبت به مراحل برنامه ریزی قبلی طرح، زمان کمتری را به خود اختصاص می‌دهد.

بعد از گردآوری داده‌ها، می‌باید آنها را به طور معمول به شیوه آماری تجزیه و تحلیل کرد و تفسیرهای مناسبی از یافته‌های تحقیق به عمل آورد. در رویکرد علمی، فرد مدعی اثبات فرضیه نیست؛ زیرا این امر مستلزم پی‌جویی حقیقت مطلق است که از ویژگیهای این رویکرد به شمار نمی‌رود. محقق صرفاً نتیجه می‌گیرد که آیا شواهد فرضیه را حمایت می‌کنند یا خیر.

### /گزارش نتایج

محقق می‌باید روشها، یافته‌ها و نتایج را به شکلی قابل فهم در دسترس دیگرانی که مشتاق آن هستند، قرار دهد. این امر ارائه روشن و مختصر مراحل تحقیق را با جزئیات کافی شامل می‌شود؛ به طوری که محقق دیگری نیز می‌تواند آن را تکرار کند.

هر یک از مراحل گفته شده در مطالعه تحقیقاتی به طور مبسوط در فصول بعدی این کتاب بررسی شده است. شاید ضروری باشد که مجدداً این نکته را به خواننده یادآوری کنیم که احتمال بسیار کمی وجود دارد که یک محقق بتواند سلسله مراحل مشخصی را، که در فرازهای قبل توصیف شد، دنبال کند. این مراحل مداوماً با هم تداخل دارند و ممکن است که جای خود را با هم عوض کنند.

## // سوالات مطروحه از سوی محققان تعلیم و تربیت

سؤال ویژه‌ای که محقق انتخاب می‌کند، مسلماً به حیطه مورد علاقه، زمینه قبلی و مسئله رویاروی او بستگی خواهد داشت. به هر حال، ما می‌توانیم مسائل مربوط به تحقیق در تعلیم و تربیت را به مسائل نظری، که با اصول اساسی سر و کار دارد، یا مسائل عملی، که برای حل مشکلات عاجل در شرایط روزمره طراحی شده‌اند، تقسیم کنیم.

## / سوالات نظریه‌ای

سؤالهایی که ماهیت نظری دارند، از این دست هستند: «این چیست؟» یا «این امر چگونه اتفاق می‌افتد؟» یا «چرا این اتفاق می‌افتد؟» در تحقیق در تعلیم و تربیت، سؤالانی که با «چه» مرتبط هستند، به طور مشخص این گونه صورت‌بندی می‌شوند: «هوش چیست؟» یا «اخلاقیات چیست؟» سوالات نوعی درباره «چگونگی» عبارتند از: «کودکان چگونه یاد می‌گیرند؟» یا «شخصیت چگونه شکل می‌گیرد؟» سؤالهای «چرایی» می‌توانند از این قبیل باشند: «انسان چرا فراموش می‌کند؟» یا «چرا بعضی از کودکان بیش از دیگران پیشرفت‌گرا هستند؟»

تحقیقی که سوی‌گیری نظری داشته باشد، می‌تواند هم به سمت ارائه نظریه و هم به سوی آزمون نظریه‌های موجود هدایت شود. حالت اول مطالعه‌ای است که محقق ضمن آن تلاش می‌کند با هدف روشن کردن ماهیت رابطه‌های موجود بین متغیرها، به تعمیم‌هایی درباره رفتار دست یابد. او ممکن است بر این باور باشد که متغیرهای خاصی با یکدیگر رابطه دارند، بنابراین تحقیق را برای توصیف ماهیت رابطه انجام می‌دهد. او می‌تواند بر اساس یافته‌هایش نظریه‌ای را در مورد پدیده‌ای خاص صورت‌بندی کند. نظریه‌های یادگیری بعد از آن که محققان قادر به نشان دادن رابطه بین روشهای خاص، فرد و متغیرهای محیطی، و کارآیی فرآیند یادگیری شدند، تکامل یافتند.

بررسیهای متداولتر در تحقیقات آموزشی احتمالاً بررسیهایی با هدف آزمون نظریه‌های موجود هستند. ارائه یک نظریه به عنوان هدف یک بررسی، بویژه برای محققان تازه‌کار، هدفی بلندپروازانه است و معمولاً کوشش برای قیاس فرضیه‌هایی از نظریه‌های موجود در مباحث یادگیری، شخصیت، انگیزش و غیره و نیز آزمون آنها واقع‌بینانه‌تر است. اگر فرضیه‌ها به طور منطقی از نظریه قیاس شده باشند و آزمایشهای تجربی شواهدی در تأیید آنها ارائه کنند، این شواهد به نوبه خود از نظریه نیز حمایت خواهند کرد.

### / سؤالات کاربردی

بسیاری از سؤالات در تحقیق مسائل تعلیم و تربیت، دارای ماهیتی کاربردی و عملی هستند و هدفشان حل مسائلی مشخصی است که کارشناسان تعلیم و تربیت در فعالیتهای روزانه خود با آن دست به گریبان هستند. آنها به این علت سؤالاتی مرتبط با تحقیق در زمینه تعلیم و تربیت هستند که با مسائلی واقعی و مبتنی بر تجربه سر و کار دارند. این سؤالات از این گونه‌اند:

آموزش برنامه‌ای در تدریس هجی کردن به دانش‌آموزان کلاس دوم چقدر مؤثر است؟  
بکارگیری روش تدریس بنیادی الفبا در پیشرفت خواندن دانش‌آموزان کلاس اول چه اثری دارد؟  
در تدریس علوم اجتماعی دبیرستان، تاثیر نسبی روش گفتگو پیرامون مسائل، در مقایسه با روش سخنرانی به چه میزان است؟ پاسخگویی به این سؤالات ممکن است در یاری رساندن به معلمان برای تصمیم‌گیرهای معقول بسیار ارزشمند باشد.

### // تحقیقات بنیادین و کاربردی

گاهی از اوقات یک سیستم طبقه‌بندی دیگر نیز در رابطه با این دو نوع سؤال مورد استفاده قرار می‌گیرد که مبنای آن هدف تحقیق است. اولین نوع تحقیق، که هدفش دستیابی به داده‌های تجربی برای صورت‌بندی، بسط و ارزشیابی نظریه است، تحقیق بنیادین خوانده می‌شود. سوی‌گیری این نوع مطالعه، چه در طرح تحقیق و چه از نظر هدف، به سمت حل مسائل عملی نیست. مقصود اصلی آن گشودن مرزهای دانش بدون توجه به کاربرد عملی آن است. البته ممکن است یافته‌های تحقیق به طور تصادفی در مسائل عملی که ارزش اجتماعی دارند، به کار گرفته شوند. برای مثال، پیشرفت در تجربیات پزشکی به تحقیق بنیادین در بیوشیمی و میکروبیولوژی وابسته است؛ همچنانکه پیشرفت در امر تعلیم و تربیت به پیشرفت در کشف قوانین عمومی از طریق انجام تحقیق بنیادین در زمینه روان‌شناسی، تعلیم و تربیت و علوم اجتماعی وابسته است.

کار اصلی تحقیق بنیادین کسب علم برای علم است. تأمل در فواید اجتماعی یافته‌های اینگونه تحقیقات، مانع از طرح ریزی آنها نمی‌شود.

دومین نوع تحقیق، که هدف آن حل یک مسئله عملی فوری است، تحقیق کاربردی خوانده می‌شود. تحقیق کاربردی گونه‌ای از تحقیق است که در رابطه با مسائل واقعی و در شرایطی که آن

مسائل در عمل بروز می‌کنند، صورت می‌گیرد. از طریق تحقیق کاربردی کارشناس تعلیم و تربیت اغلب قادر خواهد بود که مسائلی را در سطح مناسبی از پیچیدگی، یعنی در موقعیت آموزش - یادگیری در کلاس، حل کند. محقق ممکن است برای کشف قوانین کلی‌تر یادگیری به تحقیق بنیادین نیاز داشته باشد، ولی برای تعیین نحوه عملکرد این قوانین در کلاس درس باید از تحقیق کاربردی کمک بگیرد. اگر قرار باشد تغییرات علمی در کار تدریس جامعه عمل بپوشد، این رویکرد ضرورتی اساسی خواهد داشت. بجز کارشناسان تعلیم و تربیت فرد دیگری عهده‌دار حل مسائل کاربردی از این دست نخواهد بود. این نکته باید تذکر داده شود که تحقیق کاربردی نیز از روش تحقیق علمی بهره می‌گیرد. ما اذعان داریم که خط مشخصی بین تحقیق بنیادین و کاربردی وجود ندارد. یقیناً کاربردهای حاصله از نظریه به حل مسائل عملی کمک می‌کند. ما سعی می‌کنیم که نظریه‌های یادگیری را در کلاسهای درس به کار بریم. از سوی دیگر، ممکن است تحقیق بنیادین برای تکمیل صورتبندی نظریه‌ای خود به یافته‌های تحقیق کاربردی وابسته باشد. یک تجربه یادگیری در کلاس درس می‌تواند پرتو جدیدی بر نظریه یادگیری بیفکند. بعلاوه، مشاهده در یک موقعیت عملی، به آزمون نظریه‌ها کمک می‌کند و ممکن است به صورتبندی نظریه‌های جدید منجر شود.

هیچیک از روشهای گفته شده لزوماً بر دیگری رجحان ندارد. روش مورد استفاده در یک مطالعه تحقیقی را ماهیت مسئله و نوع داده‌های مورد نیاز تعیین می‌کند. در این مورد اغلب ترتیبی منطقی وجود دارد که از طریق آن نوعی از تحقیق، نوع دیگر را دنبال می‌کند. گاهی اوقات ممکن است یک محقق جهت روشننگری اتفاقات گذشته، بررسی را با مطالعه‌ای تاریخی آغاز کند و در مرحله بعدی، با استفاده از روش توصیفی اطلاعات مربوط به وضعیت کنونی مسئله را در تعلیم و تربیت تهیه کند. با این زمینه، محقق می‌تواند برای تعیین رابطه بین متغیرهایی که ممکن است از طریق انواع دیگر تحقیق پیشنهاد شده باشد، به آزمایش دست زند.

### /// زبان تحقیق

هر رشته علمی برای تشریح و خلاصه کردن مشاهدات زمینه خاص خود به یک زبان علمی نیازمند است. دانشمند در سطح کاربردی به اصطلاحاتی نیاز دارد که با کمک آنها مشاهدات ویژه را توصیف کند؛ او همچنین در سطح نظری برای اشاره به فرآیندهای فرضی که مستقیماً



نمی‌توانند مورد مشاهده قرار گیرند نیز به اصطلاحاتی نیاز دارد. دانشمندی می‌تواند از کلماتی برگرفته از زبان روزمره بهره جوید، ولی اغلب به این کلمات معانی تازه و بخصوصی نسبت می‌دهد که به طور معمول در کاربرد رایج آن یافت نمی‌شود و یا ممکن است مجبور شود اصطلاحات فنی جدیدی را که جزو زبان روزمره و رایج نیستند، خلق کند. هدف این بخش معرفی بعضی از اصطلاحات عمومی مورد استفاده در تحقیق آموزشی است.

اصطلاحاتی که یک دانشمند در دو سطح توصیفی و نظری به کار می‌برد، به ترتیب، مفاهیم و سازه‌ها نامیده می‌شوند.

### // مفاهیم و سازه‌ها

یک مفهوم، شکل تجربیدی رویدادهای مشاهده شده است؛ کلمه‌ای است که تشابهات و جنبه‌های مشترک اشیا و یا رویدادهایی را که بجز آن تشابهات و جنبه‌های مشترک، کاملاً با هم تفاوت دارند، معرفی می‌کند. کلماتی مانند *صندلی*، *سگ*، *درخت*، *ماهی* و *هزاران کلمه* در زبان ما مفاهیم مشترک چیزهای گوناگونی را ارائه می‌کنند. هدف هر مفهوم، ساده کردن تفکر از طریق گردآوری شماری از رویدادها در زیر عنوانی عمومی است. بعضی از مفاهیم کاملاً به رویدادهایی که معرف آنها هستند، نزدیک می‌باشند. بنابراین به عنوان مثال، مفهوم *درخت* به راحتی می‌تواند با اشاره به درختهای بخصوصی روشن شود. مفهوم *درخت*، تجرید کلیه خصوصیات مشترک درختان است؛ خصوصیات که مستقیماً قابل مشاهده هستند.

به این ترتیب، اصطلاحاتی از قبیل *انگیزه*، *قضاوت* و *توانایی* حل مسئله از طریق اشاره به موضوعات و رویدادهای ویژه براحتی قابل شرح نیستند. این تجریدهای سطح بالا به عنوان سازه‌ها تلقی می‌شوند. انسان از مفاهیم موجود، تجریدهای پیچیده‌تری خلق کرده و در کنار هم قرار داده است. همانگونه که انسان از کنار هم قرار دادن چوب و سایر مواد در طراحی هدفمند موفق به ساختن خانه شده است، با ادغام مفاهیم و سازه‌هایی ساده‌تر در یک طرح هدفمند، سازه‌های پیچیده‌تری را خلق کرده است. برای مثال، مفاهیمی مانند *قدرت بینایی*، *تفکیک*، *نمادها*، *جهت یابی*، *چپ و راست*، *گنجینه لغات شنیداری* و *غیره*، در طراحی هدفمند به منظور خلق سازه آمادگی برای خواندن ادغام شده‌اند. سازه‌ها در تفسیر داده‌های تجربی و در ساختن نظریه مفید هستند. آنها برای توصیف قواعد و روابط مشاهده شده مورد استفاده قرار می‌گیرند.

سازه‌ها برای این خلق می‌شوند که مشاهدات را خلاصه کنند و تبیین‌هایی را فراهم سازند. یک سازه زمانی کنار گذاشته می‌شود که راه بهتری برای تبیین و خلاصه کردن مشاهدات جایگزین آن شود. برای مثال این مشاهدات که (۱) بعضی از مواد می‌سوزند و بعضی نمی‌سوزند (۲) بعضی از مواد سریعتر از دیگر مواد می‌سوزند، برای اولین بار با سازه و یا مفهوم قابلیت اشتعال<sup>(۱)</sup> خلاصه شد و این اعتقاد را در برداشت که این امر جزء لاینفک همه مواد قابل اشتعال است که در معرض فرایند سوختن قرار می‌گیرند. این سازه هنگامی که تبیینهای مفیدتر از فرایند اشتعال ارائه شد، کنار گذاشته شد.

### // مشخص کردن معنی

هر چه مفاهیم و یا سازه‌های (ذهنی) فرد از واقعیات علمی یا پدیده‌هایی که آن مفاهیم و سازه‌ها معرف آنها هستند، دورتر باشد، امکان سوء تفاهم بیشتر خواهد بود و در نتیجه نیاز بیشتری به تعریف دقیق آنها احساس می‌شود. معنی کلمات موجود در گنجینه لغات دانشمندان باید تعریف شود. مفاهیم باید هم با اصطلاحات تجربیدی، که در واقع معنی کلی فرض شده برای این مفاهیم را ارائه می‌دهند و هم بر حسب عملیاتی که در یک مطالعه خاص بر آن مبنا اندازه‌گیری خواهند شد، تعریف شوند. کرلینگر<sup>(۲)</sup> نوع اول تعریف را یک تعریف ترکیبی و نوع دوم آنرا تعریف کاربردی می‌خواند.

### / تعریف ترکیبی

به نوع قراردادی‌تر تعریف، که در آن یک اصطلاح با استفاده از اصطلاحات دیگر تعریف می‌شود، تعریف ترکیبی اطلاق می‌شود. برای مثال، هوش به عنوان توانایی تفکر تجربیدی تعریف شده است. این نوع تعریف به انتقال ماهیت کلی پدیده‌ای که محقق به آن جلب شده است و همچنین نشان دادن رابطه آن با نظریه و سایر مطالعاتی که از مفاهیم مشابه استفاده کرده‌اند، کمک می‌کند. یک تعریف ترکیبی، یک اصطلاح را توضیح می‌دهد و شاید درباره پدیده موصوف، بینشی را در دیگران ایجاد کند. به هر حال، اگر فرد بخواهد تحقیقی انجام دهد، باید مفاهیم مورد نظر خود را به رویدادهای قابل مشاهده تبدیل کند.

### / تعریف کاربردی

یک تعریف کاربردی تعریفی است که با مشخص کردن عملیاتی که انجام آنها برای اندازه‌گیری یک مفهوم یا سازه ضروری است، معنایی را به آن سازه یا مفهوم نسبت می‌دهد، این نوع تعریف در تحقیق اساسی است؛ زیرا داده‌ها باید برحسب رویدادهای قابل مشاهده گردآوری شوند. یک دانشمند ممکن است در سطح نظریه‌ای با سازه‌هایی از قبیل یادگیری، انگیزه، هیجان یا پیشرفت سر و کار داشته باشد، ولی قبل از آنکه بتواند آنها را به طور تجربی مطالعه کند، باید بر روی چند نوع رویداد قابل مشاهده برای معرفی آن سازه‌ها تصمیم‌گیری کند. هنگامی که کسی یک مفهوم یا سازه را از نظر عملیاتی تعریف می‌کند، رویدادهای متمایزی را به عنوان نشانگرهای آن مفهوم مجرد انتخاب و عملیاتی را برای دستیابی به داده‌های مرتبط با آن مفهوم ابداع می‌کند.

دو نوع تعریف کاربردی وجود دارد: اندازه‌گیری شده و آزمایشی. یک تعریف اندازه‌گیری شده کاربردی به عملیاتی اطلاق می‌شود که از طریق آن محققان می‌توانند یک مفهوم را بسنجند. برای مثال، هوشبهر را به طور کاربردی می‌توان به عنوان نمرات به دست آمده از آزمون هوش استانفورد - بینه، و یا خلاقیت را به طور کاربردی می‌توان به عنوان نمرات به دست آمده از آزمون خلاقیت مینه‌سوتا، تعریف کرد. یک تعریف کاربردی آزمایشی به مراحل طی شده از طرف محقق برای خلق شرایط دقیق آزمایش اطلاق می‌شود. برای مثال، تعریف کاربردی از ناکامی، در یک بررسی تحقیقی می‌تواند در قالب باز داشتن آزمودنی‌ها از رسیدن به هدف، بیان شود، و یا ناهماهنگی شناختی را به طور کاربردی می‌توان به درخواست از آزمودنی‌های یک بررسی برای اظهار بیانات عمومی برخلاف باورهای خصوصی و فردی آنها اطلاق کرد.

اگر چه محقق بر اساس تجربه و دانش خود و با استفاده از گزارشهای دیگر محققان هدایت می‌شود، ولی تعریف کاربردی از یک مفهوم در بعضی از پهنه‌ها یک روش اختیاری است. محقق اغلب از بین تعاریف کاربردی ممکن، تعریفی را انتخاب می‌کند که بهترین وجه رویکرد او را نسبت به مسئله بیان کند. یک تعریف کاربردی مطمئناً نمی‌تواند معنای کامل هر مفهوم را در برگیرد. این تعریف در واقع بسیار مشخص است، هدف آن نیز محدود کردن یک اصطلاح، برای اطمینان از این امر است که هر کس با آن سروکار داشته باشد، بتواند شیوه خاص استفاده از آن اصطلاح را در یابد. تعاریف کاربردی در صورتی مکفی تلقی می‌شوند که بر مبنای آنها بتوان

داده‌هایی را گردآوری کرد که نشانگرهای قابل پذیرش مفاهیمی باشند که باید معرفی شوند. اغلب دست یافتن و یا عدم دستیابی به این نتیجه، موضوعی است که به عقیده شخصی فرد مربوط می‌شود.

تعاریف کاربردی به این دلیل در تحقیق اساسی هستند که به محققان اجازه اندازه‌گیری مفاهیم و سازه‌های مجرد را می‌دهند و همچنین برای آنها این امکان را فراهم می‌آورند که بر مبنای علم از سطح ذهنیات و نظریه به سطح مشاهده حرکت کنند. با استفاده از تعاریف کاربردی است که محققان مبادرت به تحقیقاتی می‌کنند که در فقدان این تعاریف انجام آنها غیر ممکن خواهد بود. یادآوری این مسئله مهم است که محقق با اینکه یافته‌های خود را بر اساس سازه‌های تجریدی گزارش می‌کند و سپس آنها را به نظریه و تحقیقات دیگر مرتبط می‌کند، اما در واقع آنچه که در عمل یافته است، رابطه بین دو مجموعه داده‌های قابل مشاهده و اندازه‌گیری است که برای معرفی سازه‌ها انتخاب شده بودند. برای مثال، یک تحقیق دربارهٔ رابطه بین سازه خلاقیت و سازه هوش، در عمل نمرات یک آزمون هوش و نمرات یک مقیاس خلاقیت را با یکدیگر مرتبط خواهد کرد.

### // متغیرها

یک متغیر، چند مفهوم و یا سازه را منعکس و یا تشریح می‌کند. یک متغیر مفهومی است که می‌تواند مقادیر متفاوتی را به خود بگیرد. بلندی قد، مثالی از یک متغیر است، در مورد یک شخص، قد از زمانی به زمان دیگر تغییر می‌کند، در یک زمان بین افراد مختلف فرق می‌کند، بین متوسط گروه‌ها، متفاوت است و الی آخر. طبقه اجتماعی، جنسیت، سطح گنجینه لغات، هوش‌بهر و نمرات آزمون هجی کردن مثالهای دیگری از متغیرها هستند. محققان تعلیم و تربیت به تعیین رابطه متغیرها با یکدیگر علاقه‌مند هستند. در یک مطالعه در مورد ارتباط سطح گنجینه لغات و پیشرفت علمی در بین دانش‌آموزان کلاس هشتم، متغیرهای مورد توجه، اندازه‌های سطح گنجینه لغات و اندازه‌های پیشرفت علمی خواهد بود. ما متغیرها را با مقادیر ثابت، یعنی مقادیری که ارزش عددی آنها ثابت هستند، مقایسه می‌کنیم. مفهوم سطح کلاس، اگر چه بنا به تعریف یک متغیر است، اما در این مطالعه، یک متغیر محسوب نمی‌شود؛ زیرا همه دانش‌آموزان در کلاس هشتم هستند.

### انواع متغیرها /

برای طبقه‌بندی متغیرها راههای متعددی وجود دارد. بعضی از متغیرها می‌توانند مقوله‌ای<sup>(۱)</sup> و بعضی دیگر پیوسته<sup>(۲)</sup> باشند. زمانی که آزمودنی‌ها از طریق دسته‌بندی به گروه‌هایی تقسیم شوند، آن ویژگی که تقسیم‌بندی بر مبنای آن صورت گرفته است، یک متغیر مقوله‌ای نامیده می‌شود. زبان مادری، محل سکونت، شغل اصلی پدر، و مدرسه‌ای که در آن ثبت نام کرده‌اید، مثالهایی از متغیرهای مقوله‌ای هستند. ساده‌ترین نوع متغیر مقوله‌ای، تنها دو وجه دارد و متغیر دو وجهی<sup>(۳)</sup> یا دو مقوله‌ای نامیده می‌شود. زن - مرد، شهری - روستایی، قبول - مردود از جمله متغیرهای دو مقوله‌ای هستند. بعضی از متغیرهای مقوله‌ای دارای بیش از دو وجه هستند، به عنوان مثال سطح آموزش، وابستگی مذهبی، و محل تولد از جمله این متغیرها هستند. هنگامی که یک ویژگی مقادیر عددی نامحدودی را در یک دامنه دارا باشد، آن را یک متغیر پیوسته می‌نامند. هنگامی که قد یک کودک برای مثال از ۴۰ اینچ به ۴۱ اینچ می‌رسد، او در واقع بی‌نهایت واحد اندازه‌گیری قد را طی می‌کند. قد، وزن، سن و نمرات آزمون پیشرفت تحصیلی، نمونه‌هایی از متغیرهای پیوسته هستند.

مهمترین طبقه‌بندی متغیرها بر مبنای نوع استفاده از آنها در جریان تحقیق و با توجه به این مسئله است که در چه زمانی به عنوان متغیر مستقل<sup>(۴)</sup> و در چه زمانی به عنوان متغیر وابسته<sup>(۵)</sup> طبقه‌بندی می‌شوند.

بعضی از متغیرها نسبت به دیگر متغیرها مقدم هستند. ما ممکن است این مسئله را بر اساس تجربه و یا بر اساس فرضیاتی بر مبنای نظریه دریابیم که یک متغیر نسبت به دیگر متغیرها مقدم است. به عنوان مثال، فرض بر این است که فرد برای خواندن، نیازمند سطحی از هوش است. به این ترتیب، متغیر هوش است که بر متغیر خواندن تقدم دارد. در زمینه‌ای خاص، توانایی خواندن، یکی از نتایج منطقی متغیر هوش است که به سطح هوش فرد بستگی دارد. در تحقیق، متغیرهایی که نتیجه یک متغیر متقدم و یا وابسته به آن هستند، متغیرهای وابسته نامیده می‌شوند. متغیرهایی که نسبت به متغیرهای وابسته مقدم هستند، متغیرهای مستقل نامیده می‌شوند. برای مثال: بلندی قد یک کودک (متغیر وابسته) در حد مشخصی به سن او (متغیر

1- Categorical  
3- Dichotomous  
5- Dependent

2- Continuous  
4- Independent

مستقل) بستگی دارد. اغلب این اصطلاحات را، حتی در فقدان دلایل تجربی و یا نظریه‌ای برای داشتن تقدم و تأخر متغیرها، به کار می‌بریم. این متغیرها برای روشن کردن سیر پیش‌بینی - از وضعیت افراد در متغیر مستقل تا وضعیت آنها در متغیر وابسته - نیز به کار برده می‌شوند.

یک متغیر وابسته پدیده‌ای است که موضوع یک مطالعه و یا تحقیق به حساب می‌آید. متغیر مستقل پارامتری است که به شیوه‌ای قابل محاسبه از متغیر وابسته قابل تفکیک و تشخیص است، ولی ممکن است با متغیرهای وابسته نیز مرتبط باشد. بسیاری از پارامترها می‌توانند به عنوان متغیرهایی مستقل، که از جنبه‌های مختلف محیطی قابل تفکیک هستند، عمل کنند، مانند سطح طبقاتی، محیط خانه و شرایط کلاس درس. بعلاوه، خصیصه‌های فردی، مانند، سن، جنس، هوش و انگیزه نیز خود به خود می‌توانند متغیرهای مستقلی باشند که با متغیر وابسته رابطه دارند.

از این به بعد در هر کجا از روشهای آزمایشی در تحقیق بحث می‌کنیم، متغیر مستقل را به عنوان پارامتری تعریف خواهیم کرد که از طرف آزمایشگر مورد دستکاری<sup>(۱)</sup> قرار گرفته است یا تغییر داده می‌شود. هر متغیر بر اساس آنکه کدامیک از اثرات دستکاری شده آن مورد مشاهده قرار گیرد، متغیر وابسته به شمار می‌رود. این متغیر را به آن دلیل وابسته می‌نامند که مقدارش وابسته به متغیر مستقل است و با مقدار آن تغییر می‌کند. برای مثال، به منظور بررسی اثر حل تمرینات با کمک کامپیوتر در پیشرفت ریاضی، محقق روش آموزش، متغیر مستقل را دستکاری می‌کند، و سپس اثر آن را در پیشرفت ریاضی، متغیر وابسته، مورد مشاهده قرار می‌دهد. بعد از آنکه رابطه بین متغیرها در جریان تحقیق معین شد، محقق می‌تواند از یک متغیر مستقل، متغیر وابسته را پیش‌بینی کند. در تحقیق در تعلیم و تربیت، روشهای تدریس و رویه‌های مختلف احتمالاً بیش از همه عوامل به عنوان متغیرهای مستقل مورد استفاده قرار گرفته‌اند. عوامل دیگری از این دست را سن، جنس، طبقه اجتماعی، نگرشها، هوش و انگیزش تشکیل می‌دهند. یکی از متداولترین متغیرهای وابسته پیشرفت تحصیلی و یا یادگیری است.

ممکن است متغیری در یک مطالعه متغیر مستقل و در مطالعه دیگر متغیر وابسته باشد؛ چرا که یک متغیر بر مبنای هدف مطالعه به عنوان متغیر مستقل و یا وابسته مورد توجه قرار می‌گیرد.

اگر اثر انگیزش بر پیشرفت مورد بررسی قرار گیرد، آنگاه انگیزش به عنوان متغیر مستقل به حساب خواهد آمد. به هر حال، اگر کسی به تعیین اثر روشهای آزمون کردن، گروه‌بندی کلاس، یا روشهای نمره‌گذاری بر انگیزه دانش‌آموزان راغب باشد، آن وقت است که انگیزش، یک متغیر وابسته خواهد بود و به طور کلی زمانی که خواهان بررسی اثر هوش بر یادگیری به عنوان متغیری وابسته هستیم، هوش متغیر مستقل قلمداد می‌شود و بر همین منوال در مطالعات تحقیقاتی در مورد اثر رفتن به کودکستان در رشد قوای ذهنی کودکان، هوش یک متغیر وابسته است.

دسته‌بندی دیگری از متغیر مستقل می‌تواند بر اساس قابلیت دستکاری شدن آن متغیرها از طرف محقق صورت گیرد. بر طبق این تقسیم‌بندی، دو نوع متغیر وجود دارد: فعال<sup>(۱)</sup> و خصیصه‌ای<sup>(۲)</sup>. یک متغیر فعال به عنوان متغیری تعریف می‌شود که محقق می‌تواند آنرا مستقیماً دستکاری کند. برای مثال، روش تدریس، روش گروه‌بندی و روشهای تقویت، همگی متغیرهایی هستند که می‌توان آنها را مورد دستکاری قرار داد و بر اساس متغیرهای فعال خوانده می‌شوند.

یک متغیر خصیصه‌ای متغیری است که محقق نمی‌تواند آنرا مورد دستکاری قرار دهد. اینگونه متغیرها، که گاهی اوقات متغیرهای مقرر<sup>(۳)</sup> نیز خوانده می‌شوند، جزو خصوصیات افراد محسوب می‌شوند و به خواست محقق قابل دستکاری نیستند. برای مثال، استعداد، جنس، نژاد، سن و طبقه اجتماعی نمونه‌های بارزی از متغیرهای مقرر هستند. یک محقق با تخصیص آزمودنی‌ها به گروههایی مبتنی بر این گونه متغیرهای از پیش تعیین شده، متغیرهای مقرر را در تحقیق خود به کار می‌گیرد.

در برخی شرایط گفتن آنکه کدام متغیر بر دیگری اثر می‌گذارد، ممکن نیست. برای مثال، هنگامی که رابطه بین جزم اندیشی<sup>(۴)</sup> و نگرش سیاسی مورد بررسی قرار می‌گیرد، محقق نمی‌تواند تعیین کند کدامیک از متغیرها مستقل و کدامیک وابسته هستند. این گونه موقعیتهای برای انجام آزمایش مساعد نیستند. رابطه بین چنین متغیرهایی اغلب در بررسیهای همبستگی مورد تحقیق قرار می‌گیرند.

1- Active

2- Attribute

3- Assigned

4-Dogmatism

### /// روش شناسی تحقیق در تعلیم و تربیت

روش تحقیق به راهبرد کلی، که گردآوری و تجزیه و تحلیل داده‌های ضروری برای پاسخگویی به سؤال مورد نظر را دنبال می‌کند، اطلاق می‌شود. روش تحقیق، برنامه‌ریزی و یارویی با مسئله مورد بررسی است. پنج طبقه‌بندی متداول از تحقیقات آموزشی عبارتند از:

۱- آزمایشی: یک نوع بررسی علمی است که در جریان آن، محقق یک یا چند متغیر مستقل را دستکاری و کنترل می‌کند و تغییرات حاصله از این دستکاری را در متغیر یا متغیرهای وابسته مورد مشاهده قرار می‌دهد.

#### مثالها

اثرات تمرین برنامه‌نویسی کامپیوتری لوگو در حل مسئله و توانایی تجسم روابط فضایی اثر رتوس مطالب و سرفصلها در یادآوری متن از طرف خواننده

۲- علی - مقایسه‌ای: (۱) این طبقه‌بندی مشابه روش آزمایشی است، بجز آنکه در این روش محقق نمی‌تواند به طور مستقیم متغیرهای مستقل را دستکاری کند.

#### مثالها

رابطه اعتماد به نفس و عملکرد تحصیلی با پشتکار

رابطه بلوغ جنسی با توانایی تجسم فضایی و مهارت کارکردی در دو نیمکره مغز در زنان بزرگسال

۳- توصیفی: هدف اصلی در این روش بیان چگونگی وضعیت است. در تحقیق توصیفی زیرتقسیمات متعددی وجود دارد: (الف) پیمایشها (ب) بررسیهای رشد (ج) بررسیهای پیگیری (د) تجزیه و تحلیل اسنادی (۲) (ه) مطالعه روند (۳) و (و) بررسیهای همبستگی (۴) (۵)

#### مثالها

یک پیمایش راجع به نیازهای یک معلم علوم دوره دوم

1- Causal - Comparative  
3- Documentary Analysis  
5- Correlational Studies

2- Developmental Studies  
4- Trend Studies



الگوهای تغییر در نمرات آزمونهای استعداد کلاسی در بین نوجوانان دارای استعداد تحصیلی ۴- کیفی: در این نوع بررسی، محقق اشخاص یا رویدادها را در محیط طبیعی آنها مورد مشاهده قرار می‌دهد. هدف اصلی اینگونه بررسی پی‌بردن به اثر شرایط ویژه بر رویدادهاست. این روش شناسی شامل مورد پژوهی‌ها و انواع دیگر روشهای طبیعت‌گرایانه است.

### مثالها

یک بررسی در مورد کلاس چهارم موسیقی با استفاده از روش مشاهده. آمادگی حاصل از رفتن به کودکستان، و به خاطر سپاری: یک بررسی کیفی در مورد باورها و شیوه کار آموزگاران.

۵- تاریخی: این مقوله شامل گردآوری داده‌های دوران گذشته برای درک رویدادهای آن زمان و معانی ضمنی آنها برای رویدادهای این زمان است. هدف اصلی آن، بیان آنچه که رخ داده است، می‌باشد.

### مثالها

تعلیم و تربیت فرقه مذهبی کویکرها در آمریکای قرن نوزدهم  
سیاستگذاریهای آموزشی در دوره زمامداری نیکسون

### /// خلاصه

بشر تلاش کرده است تا از طریق تجربه، منبع آگاه و مطلع، استدلال قیاسی، استدلال استقرایی، و روش علمی به سؤالات خود پاسخ گوید. هر روشی فرضیات معینی را طلب می‌کند. صحت پاسخها بستگی به فرضیات زیربنایی روشهای مورد استفاده دارد.

روش علمی به دو فرض بنیادین متکی است: (۱) حقیقت می‌تواند از مشاهده حاصل شود، و (۲) هر پدیده‌ای از روابط قانونمند پیروی می‌کند.

تحقیقات علمی حقایق مطلق را جستجو نمی‌کنند، بلکه به دنبال نظریه‌هایی هستند که پدیده را طبق روشی معتبر تبیین و پیش‌بینی کنند. این روشها نظریه‌هایی را جستجو می‌کنند که نه تنها موجز، قابل آزمون و پایدار باشند، بلکه به توبه خود محرکی برای تحقیقات بعدی

به شمار آیند. رویکرد علمی متضمن ویژگی خوداصلاحی<sup>(۱)</sup> است؛ زیرا هر نظریه‌ای آزمایشی تلقی می‌شود و احتمالاً می‌تواند با نظریه‌ای جدیدتر و درخورتر با معیارهای موجود، تعویض شود.

رویکرد علمی طی قرون متمادی جهت توضیح، پیش‌بینی، و کنترل پدیده‌های فیزیکی به کار برده شده است، اما در تعلیم و تربیت به تازگی مورد استفاده قرار گرفته است. پیچیدگی متغیرهای آموزشی و مشکلات موجود بر سر راه مشاهدات معتبر، مانع انجام تحقیق علمی در تعلیم و تربیت بوده است؛ هر چند با شروع جنبش در آغاز قرن اخیر، جستجوی علمی در تعلیم و تربیت، هم در تحقیقات نظریه‌ای و هم در بررسی‌های عملی، از پذیرش عام برخوردار شد و به موفقیت‌های بیشتری دست یافت.

مراحل ویژه تحقیق در تعلیم و تربیت عبارتند از: (۱) انتخاب یک مسئله، (۲) صورت‌بندی یک فرضیه، (۳) انتخاب راهبرد تحقیقاتی و تهیه ابزار تحقیق، (۴) گردآوری، تجزیه و تحلیل، و تفسیر داده‌ها و (۵) ارائه یافته‌ها از طریق گزارش نتایج بررسی.

تحقیق در زمینه تعلیم و تربیت را می‌توان به دو دسته اصلی تقسیم کرد: بنیادی و کاربردی. گرایش اصلی تحقیق بنیادی گسترش مرزهای دانش و کشف قوانین عام است. هدف اصلی تحقیق کاربردی حل فوری مسائل عملی است.

در سطح نظریه‌ای، صاحب‌نظران حیطه تعلیم و تربیت از اصطلاحاتی مانند هوش، خلاقیت، توانایی حل مسئله، و انگیزه استفاده می‌کنند که تجربه‌های حاصل از مشاهده رفتارهای معین می‌باشند. این اصطلاحات به عنوان سازه‌ها تلقی می‌شوند. در هر بررسی سازه‌ها کمی شده هستند و مقادیر متنوعی به خود می‌گیرند. به این ترتیب آنها در زمره متغیرها تلقی می‌شوند. دو نوع متغیر اصلی وجود دارد: وابسته و مستقل. اگر یک متغیر بر دیگری مقدم باشد، متغیر مستقل خوانده می‌شود، ولی اگر پیامد متغیر دیگر باشد، متغیر وابسته قلمداد می‌شود.

در تحقیق، این موضوع که متغیرها به طور عملی تعریف شوند، امری اساسی به‌شمار می‌رود. دو راه برای تعریف عملی متغیرها وجود دارد: (۱) به کار بردن بعضی از انواع اندازه‌گیری یا (۲) فهرست کردن مراحل انجام یک آزمایش برای خلق شرایط مطمئن. تحقیق نوع اول را تعریف کاربردی اندازه‌گیری شده و نوع دوم را تعریف کاربردی آزمایشی می‌نامند.

### // مفاهیم اصلی

استدلال استقرایی	استدلال استقرایی
اصل موجز بودن	استقرای ناقص
تحقیق کاربردی	تحقیق بنیادی
تعریف کاربردی	تعریف ساختی
تعریف کاربردی اندازه‌گیری شده	تعریف کاربردی آزمایشی
رویکرد علمی	ثابت (عدد ثابت)
فرضیه	سازه‌ها
متغیر پیوسته	متغیر
متغیر مستقل	متغیر دو وجهی
متغیر مستقل فعال	متغیر مستقل خصیصه‌ای
متغیر مقوله‌ای	مدل
مفهوم	متغیر وابسته
	نظریه

### /// تمرینها

- ۱- در تمرینهای ذیل منابع علم - استدلال قیاسی، استدلال استقرایی، یا روش علمی - را که به طور چشمگیری مورد استفاده قرار گرفته‌اند، معین کنید:
 

الف - بعد از مشاهده وسیع واکنشها، لاوازیه نتیجه گرفت که سوختن فرایندی است که طی آن یک ماده مشتعل با اکسیژن ترکیب می‌شود. نتیجه کار او ضربت کشنده‌ای بر نظریه قدیمی قابلیت اشتعال<sup>(۱)</sup> در این زمینه وارد آورد.

ب - دالتون بعد از اندیشه بسیار، نتیجه گرفت که اجسام بایستی از ذرات کوچکی به نام اتم تشکیل شده باشند. فرضیات اولیه او پایه‌ای برای نظریه اتمی شد.

ج - پس از آنکه دانشمندان فرضیات دالتون را پذیرفتند، از آنها قیاس کردند و شروع به گردآوری اطلاعاتی کردند تا این فرضیات را تأیید کنند. مدارکی برای پشتیبانی از نظریه اتمی یافته شد.

د - انیشتین با این شناخت که مواد رادیو اکتیو همواره بدون آنکه ظاهراً از جرمشان کاسته شود ذرات انرژی را بیرون می‌دهند، فرمول  $E = mc^2$  را برای تبدیل اجسام به انرژی ابداع کرد.  
 ه - فرمی<sup>(۱)</sup> با پذیرش نظریه انیشتین، دست به تجربیاتی زد که در شکافتن اتم نتیجه بخش بود.

و - پس از بررسی نظریه تقویت، یک مدرس فرضیه می‌دهد که استفاده از برنامه کامپیوترهای شخصی باعث مهارت بیشتری در حساب می‌شود. او یک بررسی را ترتیب داد که طی آن دو گروه از دانش‌آموزان کلاس شش از کامپیوتر شخصی استفاده می‌کردند، در حالی که دو دسته دیگر از مواد درسی متداول استفاده می‌کردند.

۲- نقش نظریه در بررسی علمی چیست؟

۳- چه تفاوتی بین یک نظریه استقرایی و یک نظریه قیاسی وجود دارد؟

۴- بر اساس عنوان بررسی، تحقیقات ذیل را به بنیادی و کاربردی دسته‌بندی کنید:

الف - اثر تزریق آر ان ای (اسید ریبونوکلیک) در انتقال مهارتها در حیوانات تربیت شده نسبت به حیوانات تربیت نشده

ب - نتایج برنامه اصلاحی درس حساب

ج - شرطی کردن به عنوان یک عملکرد فاصله‌گذاری بین محرکات شرطی شده و اصلی

د - تدریس هندسه برای پرورش تفکر پویا: یک بررسی آزمایشی

ه - با استفاده از عناوین فعال، خصیصه‌ای، یا احتمالاً هر دو، متغیرهای ذیل را طبقه‌بندی

کنید:

الف - مقدار داروی تجویز شده

ب - اضطراب

ج - زمینه اقتصاد اجتماعی

د - روش تدریس

۶- در بیانیه‌های ذیل کدامیک از نگرشهای ویژه دانشمندان، نقض شده است:

الف - این بررسی عهده‌دار اثبات این است که استفاده از ماری‌جوآنا برای مهارتهای

دانشگاهی مضر است.

- ب - این موضوع به طور قطع ثابت شده است.
- ج - نتایج نشان می‌دهد که ماری‌جوآنا زیانبار است.
- ۷- ویژگیهای یک نظریه سودمند کدام است؟
- ۸- در یک بررسی که برای تعیین اثر انواع مختلف محرومیت از خواب در یادگیری هجاهای بی‌معنی طراحی شده است:
- الف - متغیر مستقل کدام است؟
- ب - متغیر وابسته کدام است؟
- ۹- متغیرهای ذیل را بر اساس نوع مقوله‌ای یا پیوسته طبقه‌بندی کنید:
- الف - پیشرفت تحصیلی
- ب - روش هجایی خواندن در مقابل روش بین - بگو در خواندن
- ج - مکالمه اسپانیایی، مکالمه انگلیسی، مکالمه فرانسوی
- د - مهارت عضلانی
- ه - فریحه موسیقی
- ۱۰- ویژگیهای تعاریف کاربردی کدام است؟
- ۱۱- کدام نوع از تعریف کاربردی (اندازه‌گیری شده یا آزمایشی) برای هر یک از متغیرهای ذیل مناسبتر است؟
- الف - تقویت
- ب - پیشرفت تحصیلی
- ج - نگرش
- د - روش تدریس
- ۱۲- شما آموزگاری هستید که به میزان رفتار پرخاشگرانه قابل مشاهده در بین کودکان علاقه‌مند بوده‌اید. با اولیای دانش‌آموزان درباره روش تربیتی (استفاده از تنبیه، تشویق و امثالهم) آنها در ارتباط با فرزندشان و مجموع زمانی که کودکان تلویزیون تماشا می‌کنند، مصاحبه می‌نمایید. اطلاعات زیادی گردآوری کرده، اما احساس می‌کنید که در این موارد یکسواختی وجود ندارد. طبق روش علمی، به احتمال قوی گام بعدی شما چه خواهد بود؟
- ۱۳- کدام روش تحقیقاتی (آزمایشی، علی - مقایسه‌ای، یا پیمایشی) به طور مؤثری به هر

یک از سوالات شما پاسخ می‌گوید؟

- الف - آیا کودکانی که صبحانه می‌خورند، نمرات بهتری در مدرسه به دست می‌آورند؟  
ب - آیا یک تغذیه مناسب، عادات صبحانه خوردن را در کودکان تغییر می‌دهد؟  
ج - چه تعدادی از کودکان در مدرسه گزارش می‌کنند که در خانه صبحانه نمی‌خورند؟  
د - آیا برقرار کردن یک برنامه آزاد برای صرف صبحانه در مدرسه تفاوتی در مهارتهای دانش‌آموزان ایجاد خواهد کرد؟

۱۴- بر اساس عناوین، هر یک از بررسیهای ذیل را بر اساس مناسبترین روش مورد استفاده در تحقیق طبقه‌بندی کنید:

- الف - تفاوت‌های مبتنی بر جنسیت در عملکرد درس ریاضی خود را نشان می‌دهند.  
ب - تأثیر بازخورد آزمون کامپیوتری در جدی بودن اشتباهات بعدی  
ج - دیدگاههای دانش‌آموزان کالج و ارزشیابی یک استاد ایده‌آل  
د. اثر فقدان زودرس پدر در استعداد‌های تحصیلی  
ه - یک دبیرستان منتخب: یک بررسی عمیق  
۱۵- مثالهایی از کاربرد منبع مطلع و تجربه به عنوان منابع دانش بیاورید.  
۱۶- یک مثال از اینکه چگونه تحقیق بنیادی در علوم زیست شناختی باعث رشد حرفه پزشکی شده است، بزنید.  
۱۷- یک مثال از اینکه چگونه تحقیق بنیادی در یادگیری باعث تکامل شیوه تدریس شده است، بزنید.  
۱۸- یک مثال از تحقیق کاربردی در زمینه مورد علاقه خود بزنید. زمینه‌های دیگری را نیز فهرست کنید که نیاز به انجام تحقیقات اضافی دارند.  
کدامیک از متغیرها، می‌باید در چنین تحقیقاتی مورد بررسی قرار گیرند.

/// پاسخها

۱- الف - استدلال استقرایی

ب - استدلال قیاسی

ج - روش علمی

د - روش علمی

- ۲- نظریه، یافته‌ها را تلفیق و اطلاعات را خلاصه می‌کند، و خط مشی‌هایی برای تحقیقات جدید فراهم می‌آورد، و ما را قادر می‌سازد تا پدیده‌ها را تبیین و پیش‌بینی کنیم.
- ۳- یک نظریه استقرایی به تبیین مشاهدات پیشین کمک می‌کند، در حالی که یک نظریه قیاسی، پیش از انجام مشاهدات وسیع، ساخته می‌شود.

۴- الف - بنیادی

ب - کاربردی

ج - بنیادی

د - کاربردی

۵- الف - فعال

ب - احتمالاً هر دو

ج - تخصیص‌ای

د - فعال

۶- الف - دانشمند، واقع‌گرا و بی‌طرف است.

ب - دانشمند شکاک است و یافته‌ها را به عنوان نتایجی آزمایشی تلقی می‌کند.

ج - دانشمند با واقعیات و نه ارزشها، سر و کار دارد.

۷- یک نظریه مفید، پدیده را به ساده‌ترین شکل ممکن بیان می‌کند، با مشاهدات مربوط و

مجموعه دانش پیشین سازگار است، وسایلی برای اثبات خود فراهم می‌سازد و بررسیهای جدیدی را دامن می‌زند.

۸- الف - میزان محرومیت از خواب

ب - تعداد هجاهای بی‌معنی آموخته شده

۹- الف - پیوسته

ب - مقوله‌ای

ج - مقوله‌ای

د - پیوسته

ه - پیوسته

۱۰- تعاریف قابل قبول، یک طرز عمل واضح برای تعیین جوهر پدیده و گستردگی آن ارائه می‌کنند.

۱۱- الف - آزمایشی

ب - اندازه گیری شده

ج - اندازه گیری شده

د - آزمایشی

۱۲- برای صورتبندی یک نظریه در مورد پرخاشگری در بین کودکان تلاش کنید.

۱۳- الف - علی - مقایسه‌ای

ب - آزمایشی

ج - پیمایشی

د - آزمایشی

۱۴- الف - علی - مقایسه‌ای

ب - آزمایشی

ج - توصیفی

د - علی - مقایسه‌ای

ه - کیفی

۱۵- پاسخها متفاوت خواهند بود.

۱۶- پاسخها متفاوت خواهند بود.

۱۷- پاسخها متفاوت خواهند بود.

۱۸- پاسخها متفاوت خواهند بود.



## بخش دوم

# زمینه تحقیق

## فصل دوم

### موضوع تحقیق

#### اهداف آموزشی

دانشجو پس از مطالعه این فصل، قادر خواهد بود:

- ۱- یک مسئله تحقیق را بیان کند.
- ۲- منابع بالقوه را در مورد مسائل مربوط به تحقیق در تعلیم و تربیت، تعیین کند.
- ۳- معیارهایی را برای ارزشیابی یک مسئله تحقیق بیان کند.
- ۴- با استفاده از معیارهای قابل قبول، مسئله مورد نظر را ارزشیابی کند.
- ۵- یک مسئله کلی را در زمینه مورد علاقه خود در تعلیم و تربیت انتخاب و آن را در شکل خاصی که برای بررسی تجربی آماده باشد، بیان کند.
- ۶- اصطلاحاتی مانند جامعه آماری و متغیرها را که در تحقیق مورد استفاده قرار می‌گیرند، تعریف کند.
- ۷- جامعه آماری و متغیرهای مورد استفاده در تحقیق مورد نظر را شناسایی کند.

تحقیق نظام‌مند با یک مسئله آغاز می‌شود. جان دیویی اولین قدم در روش علمی را تشخیص یک مشکل، یک مانع یا مسئله می‌داند که محققان را سردرگم می‌کند. انتخاب و صورتبندی یک مسئله یکی از مهمترین جنبه‌های تحقیق در هر زمینه است. محققان تازه‌کار اغلب از کشف این امر که این مرحله مقدماتی، بخش عمده‌ای از زمان کل پژوهش را در یک پروژه تحقیقاتی به خود اختصاص می‌دهد، شگفت زده می‌شوند. برای انجام تحقیق هیچ راهی بجز تشخیص یک مسئله، تفکر در مورد آن و صورتبندی آن به روشی سودمند وجود نخواهد داشت.

محقق در اولین وهله ضرورتاً می‌باید در مورد حیطه وسیعی که در آن به تحقیق خواهد پرداخت، تصمیم بگیرد. محقق باید در زمینه مورد بررسی، دانش و یا تجربه داشته باشد؛ به این ترتیب که سؤالاتی را طرح کند که با تحقیق بتوان به آنها پاسخ گفت. ما اغلب می‌شنویم که دانشجویان در درسهای مشکل اظهار می‌دارند که: «من به قدر کافی نمی‌دانم تا بتوانم سؤالی مطرح کنم.» همچنین، محقق بدون داشتن دانش یا تجربه مورد نظر در زمینه‌ای خاص نخواهد دانست که دانش اضافی مورد نیاز چیست یا چگونه می‌توان آن را در جریان انجام بررسیهای تجربی کسب کرد. بعلاوه، حیطه موضوع انتخابی برای بررسی باید از جمله موضوعات عمیقاً جالب باشد و یا در مورد آن کنجکاوی واقعی وجود داشته باشد. انتخاب موضوع ضرورتاً باید بسیار شخصی باشد. در غیر این صورت، شاید انگیزه محقق تا پایان کار تحقیق پایدار نماند. یک آموزگار دوره ابتدایی ممکن است به یافتن راه مؤثرتری در تدریس خواندن علاقه‌مند باشد. یک معلم انگلیسی دوره دبیرستان شاید بخواهد بداند که آیا استفاده از دستگاه واژه‌پرداز<sup>(۱)</sup> در نگارش دانشجویانش مؤثر خواهد بود یا خیر.

بعد از انتخاب مسئله، موضوع کلی به یک بیان مشخص از موضوع بررسی محدود می‌شود. بیشتر محققان تازه‌کار صورتبندی یک موضوع و یا مسئله قابل بررسی را مشکل می‌یابند. این امر به کمبود مسائل قابل بررسی در تعلیم و تربیت مربوط نمی‌شود. در واقع سؤالات نیازمند پاسخگویی، به قدری زیاد هستند که محقق اغلب در انتخاب یکی از میان آنها دچار زحمت می‌شود. یکی از مشکلات متداول این است که در اولین مرحله می‌باید مسئله تحقیق انتخاب و یک سؤال نیز صورتبندی شود، در حالی که درک محقق تازه‌کار از چگونگی انجام یک بررسی

بسیار محدود است. بعلاوه، عدم اطمینان از ماهیت مسئله‌های تحقیقاتی، محدود کردن مسئله، معیارهای پذیرش آن و چگونگی حل مسئله اغلب توانفرسا به نظر می‌رسد. حتی پژوهشگران با تجربه نیز اغلب این مسئله را ضروری می‌دانند که قبل از رسیدن به یک مسئله تحقیقاتی، برای دستیابی به معیارهای پذیرفته شده تلاشهای متعددی کنند. یک انتخاب یا یک صورتبندی اولیه را می‌توان با بررسی دقیق‌تر، نامعقول یا فاقد ارزش یافت. مهارت در انجام تحقیق تا حد زیادی بستگی به انتخاب عاقلانه زمینه مورد تحقیق دارد. مهارت در انجام تحقیق به زمان و تلاشهای مضاعفی نیاز دارد، ولی محققان مشتاق تازه کار می‌توانند به آن دست یابند.

بیان مسئله می‌باید بروشنی متضمن آن چیزی باشد که مورد بررسی قرار می‌گیرد. گزاره اصلی می‌تواند به صورت اخباری<sup>(۱)</sup> یا سؤالی مطرح شود. به سخن دیگر، بیان مسئله باید متضمن متغیرهای مورد نظر و رابطه ویژه بین متغیرهای مورد مطالعه باشد. برای مثال، مسئله تحقیق معلم دبیرستان یاد شده می‌تواند چنین بیان شود: «مسئله مورد بررسی در این پژوهش اثر بکارگیری واژه‌پرداز در نگارش انگلیسی دانش‌آموزان دوره دبیرستان است» یا «استفاده از واژه‌پرداز چه اثری در نگارش انگلیسی دانش‌آموزان دوره دبیرستان دارد؟» بسیاری از محققان شکل سؤالی را ترجیح می‌دهند، چون به نظر می‌رسد جهت‌گیری روشن‌تری در قبال یافتن پاسخ دارد و قابل قبول نیز هست.

هر چند بعید به نظر می‌رسد، ولی به مجرد آنکه حیطه موضوع برگزیده و یک سؤال یا بیانیه بروشنی صورتبندی شود، یکی از مشکل‌ترین مراحل فرایند تحقیق انجام یافته است.

### /// ماهیت مسائل

اگر چه در تعلیم و تربیت انواع مسائل قابل تحقیق وجود دارد، ولی همه آنها متضمن جستجو برای یافتن دانش مورد نیاز در این حیطه هستند. مسئله تحقیق روشن می‌کند که محقق چه می‌خواهد بداند و روش مورد استفاده برای کسب این آگاهی را نیز معین خواهد کرد. تحقیق آزمایشی به دنبال سؤالات مربوط به روابط موجود بین متغیرهایی است که محقق می‌تواند بر همه آنها کنترل داشته باشد. در تحقیق آزمایشی محقق ضمن دستکاری متغیر مستقل

علت) همه متغیرهای دیگر را کنترل می‌کند و هر یک از تغییرات بعدی را در متغیرهای وابسته (اثر) مورد مشاهده قرار می‌دهد. کنترل در یک آزمایش، حذف تبیین‌های جایگزین برای نتایج مشاهده شده را برای محقق امکان‌پذیر می‌سازد. اگر محقق بخواهد اثر روش آموزش را در یادگیری ترکیبات جمع، در بین کودکان دبستانی مورد بررسی قرار دهد، می‌تواند ضمن کنترل همه عوامل دیگری که می‌تواند در یادگیری ترکیبات جمع دخالت داشته باشد، روش آموزش را دستکاری کند و سپس اثر این روش را در یادگیری مورد مشاهده قرار دهد.

تحقیق علی - مقایسه‌ای همچنین در پی تعیین روابط موجود بین دو یا چند متغیر است، اما در اینگونه تحقیقات دستکاری متغیر مستقل ممکن نیست؛ چون متغیر مستقل قبلاً رخ داده است. بنابراین محقق قادر نیست در مورد چگونگی بکارگیری متغیر مستقل تصمیم‌گیری کند. اگر یک محقق بخواهد اثر طلاق اولیا را بر پیشرفت تحصیلی کودکان بررسی کند، مجبور خواهد بود که پیشرفت تحصیلی گروه‌های از پیش شکل یافته را با هم مقایسه کند: یک گروه از کودکانی که پدر و مادرشان از هم جدا شده‌اند، با گروه دیگری که پدر و مادرشان با هم زندگی می‌کنند. تحقیق توصیفی سؤالاتی را در مورد ماهیت، وقوع، پراکندگی متغیرهای آموزشی و یا روابط موجود بین این متغیرها مطرح می‌کند.

در این نوع تحقیق هیچ نوع دستکاری در متغیرها صورت نمی‌گیرد - تنها به توصیف متغیرها و روابط بین آنها که به صورت طبیعی اتفاق می‌افتد، پرداخته می‌شود. یک مثال از این نوع مسئله تحقیق این سؤال است که «دانش آموزان دوره اول دبیرستان در مورد فرایندی که طی آن احزاب سیاسی کاندیداهای خود را برای ریاست جمهوری انتخاب می‌کنند، چه می‌دانند؟» یا «چه رابطه‌ای بین تعداد سالهای فراگیری زبان فرانسه در دبیرستان و تعیین دوره درسی دانشجویان درس زبان فرانسه در دانشگاه وجود دارد؟»

### /// منابع مسائل و موضوعات

اولین سؤالی که اکثر دانشجویان مطرح می‌کنند، این است که: «چگونه یک مسئله تحقیق پیدا کنیم؟» اگر چه ضابطه مشخص برای تعیین یک مسئله وجود ندارد، ولی پیشنهادهای مشخصی می‌توان یافت که به این امر کمک می‌کند. مهمترین منابع در تعیین مسائل و موضوعات عبارتند از: تجربه و قیاس از نظریه و متون تحقیقاتی پیشین.

## // تجربه

یکی از سودمندترین منابع برای محقق تازه کار، تجربیات او در زمینه کار آموزشی است. او هر روز ملزم به تصمیم‌گیری در مورد اثرات احتمالی تجربیات آموزشی بر رفتار دانش‌آموزان است. اگر قرار است این تصمیمات درست باشد، کارشناسان تعلیم و تربیت می‌باید برای اطمینان از اعتبار مفروضات خود در مورد رابطه بین تجربیات یادگیری و تغییر در دانش‌آموزان، بررسی‌های انتقادی انجام دهند.

دریافته‌های روشها و تکنیکهای تدریس، تصمیمات مختلفی باید اتخاذ شود. مطمئناً روشهای تدریس، مستعد و نیازمند بررسیهای علمی است. رویکرد علمی در بررسیهای آموزشی متضمن آن است که تصمیم‌گیری در مورد نحوه انجام امور در تعلیم و تربیت مبتنی بر شواهد تجربی باشد، نه بر حدس و گمان، احساسات و جزم‌اندیشی. برای مثال، یک معلم دوره ابتدایی ممکن است در مورد اثر روش خود در تدریس قرائت سؤال کند. او ممکن است بخواهد روش معمول خود، یا هر روش شناخته شده دیگری را برای آگاهی از آنکه کدام یک مؤثرتر است، ارزشیابی کند، یا معلم دوره دبیرستان سؤال می‌کند که برای تدریس درس اقتصاد، روش مباحثه مؤثر است یا روش مقاله‌نویسی. استاد زیست‌شناسی ممکن است بخواهد در مورد اثر شبیه‌سازی از طریق کامپیوتر در مهارتهای حل مسئله از طرف دانشجویان تحقیق کند و یا این سؤال را مطرح کند که آیا شبیه‌سازی کامپیوتری از روشهای دیگر مؤثرتر است؟

مشاهده روابط مشخص برای مواردی که هنوز هیچ تبیین رضایت‌بخشی در مورد آنها وجود ندارد، یکی دیگر از منابع مسائل برای تحقیق است. معلم ممکن است دریابد که در زمانهای خاصی علائم آشکاری از اضطراب در دانش‌آموزان بروز می‌کند. این معلم، برای بررسی مسئله، می‌تواند تبیین‌های آزمایشی متعددی را در مورد منشأ این اضطراب طرح کند و سپس آنها را به طور تجربی مورد آزمون قرار دهد. این بررسی نه تنها مسئله آنی را حل می‌کند، بلکه سهم کوچکی نیز در درک عوامل مؤثر در اضطراب مربوط به کلاس درس ایفا می‌کند.

همچنین باید تصمیماتی درباره شیوه‌های معمول در مدارس گرفته شود. بسیاری از این شیوه‌ها اساساً بر سنتها یا نظریات گذشته‌گان بنا شده‌اند؛ در حالی که یا تایید علمی بسیار کمی پشتوانه آنها است و یا اصلاً چنین تاییدی وجود ندارد.

پس چرا نباید برخی از این شیوه‌ها را ارزشیابی کرد. برای مثال، چرا نباید برنامه آزمون

سالیانه را ارزشیابی کرد. آیا آزمونهای جایگزین<sup>(۱)</sup> معتبرتری از آنچه در حال حاضر مورد استفاده قرار می‌گیرد، برای این منظور خاص وجود دارند؟

بنابراین تجربیات روزمره متخصص تعلیم و تربیت می‌تواند مسائل ارزشمندی برای تحقیق فراهم آورد. در واقع، بیشترین ایده‌های تحقیقاتی، که به وسیله پژوهشگر تازه کار در تعلیم و تربیت ارائه می‌شود، ناشی از تجربیات شخصی او است. یک محقق ممکن است درباره روابط تازه و یا شیوه‌های مختلف دستیابی به هدفهای معین تصوّراتی داشته باشد، بنابراین، از طریق نوعی فرآیند شهودی، به ایده‌های خود برای تحقیق دست پیدا می‌کند؛ اگر چه حقیقت این است که این بررسیها عمدتاً نوعی راهبرد برای حل مسائل آتی هستند، با این حال این مسائل اغلب برای محققان تازه کار مفیدتر و پرمعناتر از مسائلی هستند که از طریق فرآیند منطقی قیاس از نظریه به دست می‌آیند. بعلاوه آنها می‌توانند بر اساس سهمشان در پیشرفت تجربیات تعلیم و تربیتی تایید و توجیه شوند.

### // قیاس از نظریه

قیاسها که می‌توانند بر مبنای نظریه‌های متعدد آموزشی و رفتاری که محقق با آنها آشناست، به عمل آیند، یک منبع عالی برای مسائل تحقیقاتی فراهم می‌آورند. نظریه‌ها با اصول کلی سروکار دارند که در مسائل خاص تعلیم و تربیت، مادامیکه در عمل مورد تایید قرار نگرفته‌اند، تنها به صورت فرضیه باقی می‌مانند. تنها در جریان تحقیق است که محقق تعیین می‌کند که آیا تعمیم‌های موجود در نظریه‌ها، قابل تبدیل به پیشنهادهای ویژه‌ای برای فعالیتهای آموزشی هستند یا خیر.

محقق، از یک نظریه می‌تواند فرضیاتی را استنتاج کند که یافته‌های مورد انتظار را در یک موقعیت کاری ویژه بیان می‌کنند.

بنابراین، محقق این پرسش را مطرح می‌کند که اگر نظریه بدرستی وضعیت موجود را خلاصه کرده باشد، چه روابطی بین متغیرها مشاهده خواهد شد، و سپس برای یقین از اینکه داده‌های تجربی به دست آمده، فرضیه و به دنبال آن نظریه را تایید می‌کند، یا خیر، دست به تحقیق نظام‌مند خواهد زد.

نظریه‌هایی در زمینه یادگیری، شخصیت، علوم اجتماعی، توسعه اجتماعی و بسیاری زمینه‌های دیگر وجود دارد که اعتبار، گستره و عمل‌پذیری آنها می‌تواند به طور سودمندی در موقعیتهای آموزشی، آزمون شود. نظریه تقویت بویژه می‌تواند به عنوان نقطه شروع در تحقیقات مربوط به شرایط کلاس درس مورد استفاده قرار گیرد. معانی ضمنی را که از یکی از مفروضه‌های نظریه تقویت می‌توان قیاس کرد، برای آزمون کلاسی در نظر بگیرید، با این معنی که، تقویت پاسخها موجب افزایش میزان و شدت پاسخها می‌شود. ما می‌دانیم که این نظریه محرک بسیاری از تحقیقات اخیر بوده است، ولی همچنان قیاسهای بسیاری وجود دارند که باید ساخته شوند و در شرایط کلاس مورد آزمون قرار گیرند. برای مثال، در مورد اثر فقدان تقویت آشکار و یا عدم تقویت پاسخهای صحیح دانش‌آموزان در شرایط کلاس درس، تحقیق کافی صورت نگرفته است. ما از طریق مطالعات آزمایشگاهی با حیوانات دریافته‌ایم که عدم تقویت، احتمال دادن پاسخ را کاهش می‌دهد و در نتیجه به خاموشی پاسخ منجر می‌شود. آیا این یافته می‌تواند به کلاس درس بسط داده شود؟ یعنی بر این اساس معلم می‌تواند فرض کند که پاسخهای درست دانش‌آموزان اگر به طور آشکار تقویت نشوند، خاموش خواهند شد؟ در حال حاضر، تحقیقات کافی در جهت آزمون این اصول در کلاس درس در دست نداریم.

در بین نظریه‌های دیگری که به نظر می‌رسد منابع سودمندی برای تحقیق باشند، می‌توان از نظریه اضطراب امتحان ساراسون<sup>(۱)</sup>، نظریه ناهماهنگی شناختی فستینگر<sup>(۲)</sup>، نظریه‌های متعدد خویشتن، نظریه انگیزه پیشرفت، نظریه قضاوت - اجتماعی، نظریه نقشها و نظریه پدیده‌شناختی یاد کرد.

برای مثال کیپنیز از نظریه فرایندهای مقایسه - اجتماعی فستینگر به عنوان مبنایی برای تحقیق در مورد رابطه بین ادراک دیگران و فرایند تغییر خود پنداره استفاده کرد.<sup>(۳)</sup> نظریه فستینگر این فرض را در بر دارد که خود سنجی‌ها از طریق انجام مقایسه بین خود و دیگران صورت‌بندی می‌شوند.<sup>(۴)</sup> تعدادی فرضیه از این نظریه قیاس شده‌اند که می‌گویند تغییرات مورد انتظار در خود پنداره در رابطه با ادراک و مقایسه با افراد مهم صورت می‌گیرد. آزمودنی‌های بررسی را ۸۷

1- Sarason, I. (1980). Introduction to the Study of Test Anxiety. In I. G. Sarason (ed). *Test Anxiety: Theory, Research, And Applications*. Hillsdale NJ: Erlbaum.

2- Festinger, L. A Theory of Cognitive Dissonance (New York, Har Per & Row)

3- Kipnis, D. M. (1961). Change in Self- Concepts in relation to Perceptions of Others. *Journal of Personality*, 29, 44g-465.

4- Festinger, L. (1954). A Theory of Social Comparison Process. *Human Relations*, 7, 117- 140.

دانشجویی تشکیل می‌دادند که با یکدیگر در خوابگاه دانشگاه زندگی می‌کردند. ارزیابی آنها از ویژگیهای شخصیتی خود در رابطه با ادراک آنها از بهترین دوستانشان مورد بررسی قرار گرفت. همه فرضیات مورد تایید داده‌ها قرار گرفت، و کسپنیز نتیجه گرفت که ادراک میان فردی و خودپنداره بشدت به یکدیگر نزدیکند.

این رویکرد نسبت به مسئله‌های تحقیقاتی به مطالعاتی منتج می‌شود که به دلیل مبتنی بودن بر یک نظریه مشترک، به آسانی با هم قابل تلفیق هستند. چنین تحقیقات مرتبط با هم بخصوص به عنوان وسیله‌ای برای گسترش دانش در حیطه‌ای مشخص ثمربخش است.

### // متون تحقیقاتی پیشین

یکی دیگر از منابع موضوعات پژوهشی، متون تحقیقاتی موجود در زمینه مورد علاقه محقق است. به هنگام مطالعه تحقیقات پیشین، در معرض نمونه‌هایی از مسئله‌های تحقیقاتی و شیوه‌های انجام آن تحقیقات قرار می‌گیریم. راههای متعددی وجود دارند که طی آن مرور تحقیقات پیشین می‌تواند در صورتبندی یک مسئله تحقیقاتی جدید مؤثر افتد.

۱- مطالعه تحقیقات پیشین ما را در صورتبندی سؤالات تحقیقاتی که نسبت به بررسیهای گذشته، گام منطقی بعدی تلقی می‌شوند، کمک می‌کند. بسیاری از اوقات نتایج بخشی از یک تحقیق، ما را به سمت سؤالات جدیدی هدایت می‌کند. در برخی از موارد، محققان در بخش نتیجه‌گیری گزارش تحقیقاتی خود سؤالات جدیدی را که پدید آمده‌اند، مطرح می‌کنند. یک راه ثمربخش برای وسعت بخشیدن به بررسیها، معرفی متغیرهای جدید برای کنترل بیشتر و شناسایی اثرات متقابل بین متغیرها است.

۲- مرور تحقیقات پیشین می‌تواند برای محقق انگیزه‌ای باشد، برای اینکه دریابد شیوه‌هایی که در آن تحقیقات مورد استفاده قرار گرفته‌اند، می‌توانند برای حل مسئله‌های تحقیقاتی دیگر مناسب باشند یا خیر، یا یک بررسی مشابه می‌تواند در زمینه یا حوزه متفاوت و یا با گروههای متفاوت آزمودنی انجام یابد یا خیر. برای مثال، محقق، پژوهشی را که اثر بخشی یک رویکرد چند رسانه‌ای را در آموزش شیمی مورد بررسی قرار داده است، مطالعه می‌کند. شاید یک بررسی مشابه بتواند در حوزه زیست‌شناسی یا حوزه دیگر انجام یابد، یا یک بررسی در مورد دانش‌آموزان دوره متوسطه ممکن است رهنمودی برای یک معلم دوره ابتدایی باشد که



می خواهد بداند آیا در سطح ابتدایی نیز همان روابط مابین متغیرها حاکم است یا خیر.

۳- مرور تحقیقات گذشته می تواند ضرورت تکرار آنها را برای تایید یافته‌های پیشین خاطر نشان کند. یکی از ویژگیهای اساسی یک تحقیق علمی، قابلیت تکرار آن است؛ به گونه‌ای که یافته‌های آن قابل اثبات باشد. تکرار یک بررسی با ایجاد تغییرات و یا بدون آن احتمالاً برای محقق تازه کار فعالیت ارزشمند و پر سودی است. تکرار یک بررسی گسترده، تعمیم یافته‌های تحقیق را وسعت می بخشد و همچنین شواهد اضافی درباره اعتبار آن یافته‌ها فراهم می آورد. در بسیاری از آزمایشهای آموزشی، امکان انتخاب آزمودنی‌ها به طور تصادفی وجود ندارد و ما می باید از گروههای کلاسی که از قبل سازماندهی شده‌اند، استفاده کنیم. این امر البته ظرفیت تعمیم یافته‌ها را محدود می سازد. به هر حال با تکرار تجربیات در زمانها و مکانهای مختلف و با تایید روابط مورد قبول در هر بررسی، اطمینان نسبت به اعتبار علمی یافته‌ها نیز افزایش می یابد. تکرار صرف بررسیهای دیگر، مهمترین خاستگاه تلاشهای تحقیقی نیست، ولی در زمینه مسائل تعلیم و تربیت نیاز مداوم به تایید و بسط یافته‌ها وجود دارد.

در بیشتر موارد، تکرار بررسیهای گذشته، دقیقاً شبیه همان تحقیق قبلی نیست. به هنگام تکرار تحقیقات پیشین، برای روشن ساختن برخی از وجوه یافته‌ها، برای آزمون اینکه یافته‌ها را تا چه حد می توان تعمیم داد، یا برای بررسی عواملی که در تحقیق اصلی گنجانده نشده بودند، تغییراتی اعمال می شود. برای مثال، بررسیهای پیاژه در مورد رشد داوریهای اخلاقی در کودکان، به دفعات در دیگر کشورها تکرار شده است.<sup>(۱)</sup> این بررسیها اگر چه از روش بنیادی پیاژه استفاده کرده‌اند، ولی رشد داوریهای اخلاقی را در کودکان طبقات اجتماعی متفاوت، در کودکان با مقطع سنی یکسان ولی با سطح هوش متفاوت، در کودکانی که در گروههای سنی خود میزان مشارکت متفاوتی دارند، در کودکانی که آداب تربیتی متفاوت دارند و در بین دختران و پسران، مورد بررسی قرار داده‌اند. اخیراً محققان دیگری نیز در تلاش برای اثبات یافته‌ها و نتایج پیاژه از روشهای متفاوتی استفاده کرده‌اند. در مجموع، قسمت عمده تحقیقات مشتق از پژوهشهای پیاژه، نتایج اصلی او را تایید کرده‌اند. به این ترتیب، یک بررسی منفرد، اگر با یک موضوع بااهمیت سروکار داشته باشد، و یا اگر یافته‌هایش هیجان‌انگیز باشد، می تواند برای بسیاری از بررسیهای دیگر الهام بخش باشد.

1- Piaget, J. (1932). *The Moral Judgment of The Child*. Glencoe, IL: Free Press

۴- تکرار تحقیقاتی که در گذشته انجام شده است، می‌تواند سؤالاتی در مورد قابلیت بکارگیری یافته‌های آن در فرهنگهای دیگر نیز برانگیزد. نتایج به دست آمده از یک تحقیق انجام یافته در فرهنگی خاص نمی‌تواند به طور خودکار در مورد فرهنگهای دیگر به کار گرفته شود و این یک دلیل برای آن است که در دهه‌های اخیر تأکید قابل توجهی بر بررسیهای میان فرهنگی شده است. در حیطه‌هایی مانند روان شناسی کودکان و بزرگسالان، یادگیری اجتماعی، رشد شناختی و زبانی، انگیزه پیشرفت، تکوین شخصیت، تجربیات آموزشی، مثالهای بی‌شماری از تحقیقات میان فرهنگی می‌توان یافت. تعداد رو به تزاید بررسیهای بین‌المللی در دانشگاههای آمریکا، هم علاقه به این نوع بررسی را افزایش داده و هم گردآوری داده‌های میان فرهنگی را تسهیل کرده است. تعدادی از رساله‌ها و پایان نامه‌های این دانشجویان در این زمهراند و بیتیانی برای بررسیهای میان فرهنگی بیشتر فراهم می‌سازند.

۵- مرور تحقیقات پیشین می‌تواند در شناسایی تضادها و تناقضات یا در ناخرسندی از مفهوم سازی، روش شناسی، وسایل اندازه‌گیری، و تجزیه و تحلیل‌های آماری مورد استفاده، ثمر بخش باشد. محققان اغلب می‌توانند در تحقیقات گذشته چیزی را برای اثبات بیابند. برای مثال، در سال ۱۹۲۸ هارت شورن و می این نظریه موجود را که صداقت یک ویژگی شخصیتی یکپارچه است، مورد چالش قرار دادند.<sup>(۱)</sup>

آنها رفتار چند هزار کودک دانش‌آموز را در موقعیتهای متعدد اغواکننده همبسته کردند و از همبستگی‌های پایین بین اندازه‌های اغواگرانه نتیجه گرفتند که صداقت، یک خصوصیت درونی مشترک نیست، بلکه خصوصیتی ویژه و تحت تأثیر موقعیتی است که افراد در آن قرار دارند. برتون در سال ۱۹۶۳ این بررسی را تکرار کرد<sup>(۲)</sup>؛ چراکه تأکید هارت شورن و می را درباره تأثیر عوامل محیطی بر رفتار صادقانه، مورد تردید قرار داده بود. برتون از روش تحلیل عوامل استفاده کرد و از تمام تکالیف محوله به کودکان، شواهدی در تایید ثبات درون فردی به دست آورد. او نتیجه گرفت که احتمالاً در فرد خصیصه‌ای بنیادین در مورد صداقت وجود دارد که آن را به هنگام مقاومت در برابر موقعیتهای اغواکننده متجلی می‌کند، ولی او نیز با هارت شورن و می در رد یک فرمول همه یا هیچ در مورد شخصیت افراد موافقت داشت. در سال ۱۹۶۵ هانت

1- Hart Shome, H. and M.A. May (1928). Studies In the Nature of Character. Studies in Deciet (vol. I). New York: Macmillan.

2- Burton, R. v (1963) - Generality of Honesty Reconsidered Psychological Review, 70, 481-499.

بررسی برتون را تکرار کرد، ولی به عنوان روش آماری، از تحلیل واریانس به جای تحلیل عوامل بهره گرفت.<sup>(۱)</sup> هانت اعتقاد داشت که تفاوت‌های شخصیتی، موقعیتها و تعامل بین فرد و موقعیت بایستی به عنوان منابع تغییر در رفتار صادقانه محسوب شوند و در تأیید فرضیه تعامل، شواهدی نیز فراهم کرد.

نلسن، گریندر، و موتر در سال ۱۹۶۹ همهٔ بررسیهای یاد شده را تکرار کردند و آنها را بسط دادند.<sup>(۲)</sup> آنها رویکردهای روش شناختی مختلف را با هم مقایسه کردند و نتیجه گرفتند که رفتار اغواگرانه، طی انجام مجموعه‌ای از تکالیف مختلف، تنها تا حدودی باثبات است. بنابراین، یافته‌های آنها با بررسیهای اولیه هارت شورن و می، به رغم بیش از چهل سال فاصله زمانی در جمع‌آوری اطلاعات مربوط به هر دو بررسی و به رغم جامعه‌های آماری متفاوت، تکالیف و شیوه‌های روش شناختی متفاوت، مطابقت داشت.

نمونه دیگری از بررسیهای برگرفته از بررسیهای پیشین، بررسیهایی است که اثرات گروه بندی بر اساس استعداد را در پیشرفت تحصیلی دانش‌آموزان مورد مطالعه قرار می‌دهد. در اولین سالهای قرن اخیر صدها بررسی در مورد این موضوع انجام یافت و در حال حاضر نیز موضوعی برای تحقیق به شمار می‌رود.<sup>(۳)(۴)(۵)</sup>

اغلب اوقات محقق از فواصل بارز در دانش نظام یافته در حیطه‌ای خاص آگاه است. بنابراین تحقیق باید طوری طراحی شود که به پر کردن این فواصل یاری رساند و به دانش معتبرتری منتج شود. برای دانشجویانی که در این مرحله از تحقیق قرار دارند، تشکیل جلساتی با استادان صاحب‌نظر، می‌تواند مثر ثمر باشد. بنابراین، دانشجو با تحلیل انتقادی تحقیقات منتشر شده در زمینه مورد نظر و با کمی خلاقیت، قادر خواهد بود که موضوعات متعددی را که بالقوه قابل تحقیق هستند، بیابد. درک جنبه‌های نظریه‌ای و کاربردی حیطه مورد نظر، محقق را قادر می‌سازد تا به شیوه‌ای انتقادی مطالعه کند و موضوعی درست را انتخاب کند.

- 1- Hunt, J. Mev. (1965). Traditional Personality Theory In The Light of Recent Evidence. *American Scientist*, 53, 80-96.
- 2- Nelsen. E.A., R.E. Grinder, and M.L. Mutterer (1969) Sources of Variance in behavioral measures of honesty In temptation Situations: Methodological analysis. *Developmental Psychology*, 1, 265-279.
- 3- Slavin, R. E. (1987). Ability Grouping and Student Achievement in Elementary Schools: A Best Evidence Synthesis. *Review of Educational Research*, 57, 293- 335.
- 4- Hiebert, E. (1987). The Context of Instruction and Student Learning: An Examination of Slavin's Assumption. *Review of Educational Research*, 57, 337- 340
- 5- Gamoran, A. (1987). Organization, instruction, and the Effects of Ability Grouping: Comment on Slavin's "best- evidence Synthesis" *Review of Educational Research*, 57, 341 - 345.

## // منابع غیر آموزشی

تجربیات مشاهدات در کل جهان، به همراه فعالیت‌های تخصصی ما، می‌تواند منابع پرثمری در انتخاب موضوع تحقیق باشد. نظریه‌ها یا شیوه‌هایی که محقق در سایر زمینه‌ها با آن برخورد می‌کند، می‌تواند برای استفاده در حیطه تعلیم و تربیت منطبق شوند. اغلب جنبش‌هایی که از خارج زمینه تخصصی ما سرچشمه می‌گیرند، ما را به سمت راه‌های جدیدی از پژوهش سوق می‌دهند. جنبش زنان، ما را به سمت بررسی قالب پردازی‌های جنسی در مواد آموزشی، اثر مدارس در یادگیری نقش‌های جنسی، تفاوت‌های جنسی در شخصیت و از این قبیل هدایت کرده است. جنبش حقوق مدنی بررسی‌های بیشماری را در مورد آموزش کودکان گروه‌های اقلیت به راه انداخت. شیوع بیماری ایدز در کشور موجب شده است که بررسی‌های زیادی برای تعیین بهترین شیوه‌ها و مواد آموزشی در زمینه آگاه ساختن دانش‌آموزان از خطرات بیماری و بهترین نحوه حفاظت از آنها در مقابل بیماری، صورت پذیرد.

بیشتر تحقیقات ارزشمند در تعلیم و تربیت از چنین منابع غیر آموزشی الهام می‌گیرند.

## /// ارزشیابی مسئله تحقیق

بعد از آنکه یک مسئله آزمایشی انتخاب شد، می‌باید مورد ارزشیابی قرار گیرد. محقق باید مطمئن باشد که زمینه مربوط به مسئله از آنچنان اهمیتی برخوردار است که تحقیق درباره آن را توجیه می‌کند؛ هر چند که تعیین این مسئله همیشه آسان نیست. قضاوت در مورد ارزش یک مسئله غالباً امری است که به ارزش‌های فردی و عقاید شخصی مربوط می‌شود. به هر حال، معیارهای خاصی وجود دارند که در فرآیند ارزشیابی اهمیت مسئله تحقیق باید مورد استفاده قرار گیرند.

۱- در بهترین حالت، مسئله باید به نوعی باشد که حل آن در پیکره دانش سازمان یافته تعلیم و تربیت سهم داشته باشد. محقق باید نشان دهد که بررسی او در پی پر کردن فواصل موجود در دانش کنونی و یا کمک به حل پاره‌ای تناقضات در تحقیقات گذشته است. شاید یک تحقیق بتواند مطالعات قبلی را به گونه‌ای غنی‌تر سازد که دانش معتبرتری حاصل شود. بیشتر دانشمندان بر این باورند که مسائلی که از نظریه سرچشمه گرفته‌اند، امکان بالقوه بیشتری برای حصول به این معیار دارند. محقق می‌تواند از خود سؤال کند که آیا بررسی او به شناخت روابط

جدید منجر خواهد شد یا به سادگی، یافته‌های قبلی را تکرار خواهد کرد.

مطالعات خاصی ممکن است هم در نظریه و هم در عمل سهم داشته باشند. به هر حال اگر مسئله‌ای فاقد مفاهیم آشکار نظری باشد، در نهایت باید دارای چند مفهوم عملی باشد. محقق باید با توجه به بررسی‌اش قادر به پاسخگویی به این سؤال باشد که «بسیار خوب، نتیجه چه خواهد بود؟» آیا راه حل مسئله تفاوتی در شیوه‌های تعلیم و تربیت ایجاد خواهد کرد؟ آیا دیگر متخصصان تعلیم و تربیت نیز به این یافته‌ها علاقه‌مند هستند؟ آیا یافته‌ها در موقعیتهای تصمیم‌گیری آموزشی سودمند خواهند بود؟

از آنجا که مسائل زیادی با مفاهیم عملی و نظریه‌ای وجود دارند که نیازمند راه حل هستند، صرف وقت و تلاش برای پاسخگویی به سؤالاتی که فاقد یکی از این دو جنبه مهم هستند، توجیه ناپذیر است.

بعضی اوقات، یک دانشجو در تلاش برای تعیین یک مسئله، سؤالی را انتخاب می‌کند که روابط ضعیفی را مطرح می‌کند. برای مثال، اینکه «چه رابطه‌ای بین محبوبیت فرد در میان همسالان و سرعت خواندن وجود دارد؟» یک مسئله جزئی محسوب خواهد شد. این امر برای تجارب آموزشی اهمیت بسیار کمی دارد و یا اصلاً اهمیت ندارد. همچنین سؤالی نظیر «چه رابطه‌ای بین هوش و پیشرفت خواندن وجود دارد؟» نیز سؤالی بی‌اهمیت قلمداد می‌شود؛ چرا که تاکنون اطلاعات معتبری در این زمینه به دست آمده است و تحقیقات بیشتر، غیر ضروری به نظر می‌رسد. «هیچ نیازی برای اختراع مجدد چرخ نیست.»<sup>(۱)</sup> بسیاری از مطالعات پیشنهادی باید به سهولت رد شوند؛ چرا که نه در نظریه و نه در کل، سودی برای تعلیم و تربیت ندارند. برای مثال، محقق تازه کاری که می‌خواهد مزیت‌های تطبیقی دو روش متفاوت را در شروع آموزش خواندن ارزشیابی کند، ممکن است بررسی نگرش آموزگاران کلاس اول را در مورد این مسئله پیشنهاد کند. چنین پیمایشی، پاسخ با معنایی در مزایای نسبی این دو سیستم به دست نخواهد داد؛ چرا که تنها از گروه کوچکی از معلمان، که احتمالاً به اندازه محقق در این زمینه جاهل هستند، به دست آمده است. رویکرد بهتر، یک آزمایش کنترل شده برای مقایسه پیشرفت دو گروه از دانش‌آموزان همسان تحت یک سیستم است که هر کدام با یکی از روشهای مذکور با خواندن آشنا شده‌اند. سهولت در پیمایش نظرات، بسیاری از افراد را در بکارگیری این روش،

حتی زمانی که پاسخ به پرسش مطرح شده نابجاست، اغوا می‌کند.

۲- یک مسئله تحقیق باید ما را به مسائل جدید و در نتیجه به تحقیقات بیشتر هدایت کند. این امر زمانی اتفاق می‌افتد که محقق بتواند مسئله تحقیق خود را با دانش سازمان یافته پیوند دهد و توجه خود را به نوع مطالعه‌ای که به طور منطقی می‌تواند پی‌آمد کار او باشد، معطوف کند. هنگامی که یک مطالعه مطلوب به پاسخ یک سؤال دست پیدا می‌کند، معمولاً سؤالات دیگری ایجاد می‌کند که نیاز به بررسی دارند. این مسئله در مورد بررسیهایی که با نظریه تقویت در کلاس درس سروکار دارند، واقعیت داشته است. در مقابل، بسیاری از تحقیقات توصیفی، به این معیار دوم دست نمی‌یابند. در این زمینه، مؤلفان کتاب پیشنهاد می‌کنند که دانشجویان مبتدی در امر تحقیق، در انتخاب موضوع به مسائلی توجه داشته باشند که امکان توسعه و پیگیری آتی آنها در بررسیهای دوره‌های عالی‌تر دانشگاهی وجود داشته باشد.

۳- مسئله باید قابل تحقیق باشد. اگرچه این معیار بدیهی به نظر می‌رسد، ولی در عمل بسیاری از مسائل در زمره سؤالاتی نیستند که بتوان آنها را مورد بررسی علمی قرار داد. یک مسئله، برای آن که قابل تحقیق باشد، باید این قابلیت را داشته باشد که در عمل مورد بررسی قرار گیرد. بسیاری از سؤالات جالب نمی‌توانند موضوع تحقیق تجربی واقع شوند، ولی باید طی بررسیهای فلسفی مورد پژوهش قرار گیرند. یک چنین سؤالاتی همانند «آیا آموزش مسائل جنسی در مدارس ابتدایی واجب است؟» یا «آیا ما باید کمونیزم را در دبیرستان تدریس کنیم؟» همگی جزو مسائل فلسفی هستند که با تحقیق علمی قابل پاسخگویی نیستند. هر چند با سؤالات مذکور نمی‌توان به طور تجربی کار تحقیق را آغاز کرد، اما می‌توان آنها را در قالب سؤالات قابل تحقیق مجدداً صورتبندی کرد. به عنوان مثال، باید سؤال اول را به این شیوه روشن‌تر بیان کنیم که آموزش مسائل جنسی در مدارس ابتدایی چه اثری بر روی نگرش دختران دوره اول دبیرستان در مورد تجربه جنسی قبل از ازدواج دارد. یک بررسی می‌تواند در جهت دستیابی به اطلاعات لازم برای یک چنین سؤالاتی طراحی شود. هر چند که سؤالات فلسفی از این نوع برای تحقیق علمی مناسب نیستند، با وجود این اطلاعاتی که تحقیق فراهم می‌آورد، می‌تواند در پلختن پاسخ سؤالات فلسفی و اخلاقی کاربرد داشته باشد. به این ترتیب اطلاعات به دست آمده در جریان تحقیق بر روی مسئله‌ای خاص می‌تواند در تصمیم‌گیری در مورد اخلاقیات و ارزشها، کارشناسان تعلیم و تربیت را یاری کند.

محقق همچنین باید به تعریف و اندازه‌گیری متغیرهای دخیل در سؤال مطرح شده توجه داشته باشد. مسئله‌ای نظیر «تغییر الویتهای ملی بر آینده تعلیم و تربیت در آمریکا چه تاثیری دارد»، برای تحقیق مناسب نخواهد بود. تعریف عباراتی مانند تغییر الویتهای ملی و آینده تعلیم و تربیت در آمریکا، به گونه‌ای که قابل اندازه‌گیری باشند، مشکل خواهد بود.

۴- مسئله می‌باید یا ویژگیهای محقق مناسبت داشته باشد. ممکن است مسئله از نظر معیارهای تبیین شده عالی باشد، ولی شاید برای محقق مهم جلوه نکند. تعدادی از جنبه‌های مشخصی که می‌باید مورد ملاحظه قرار گیرند، عبارتند از:

الف - مسئله باید از نوعی باشد که محقق به آن علاقه‌رستین دارد و مشتاقانه بتواند در مورد آن به تحقیق بپردازد. راه حل مسئله باید برای محقق مهم باشد. از آنجا که راه حل مسئله می‌تواند در افزایش دانش محقق در حیطه‌ای خاص، یا بهبود عملکرد او به عنوان یک متخصص تعلیم و تربیت مؤثر باشد، این راه حل باید برای محقق مهم باشد. اگر مسئله تحقیق پر معنا و جالب نباشد، بعید به نظر می‌رسد که به عنوان وظیفه‌ای تمام و کمال محقق را به صرف وقت و انرژی وادارد.

ب - مسئله می‌باید در زمینه‌ای باشد که محقق در مورد آن صاحب دانش و مقداری تجربه است. او نیازمند آشنایی با نظریه‌ها و مفاهیم موجود و پایه‌ریزی واقعیات به شیوه‌ای است که منجر به تعیین مسائلی پر ارزش شود. بعلاوه محقق باید روشن سازد که آیا از مهارتها و صلاحیتهای لازم برای به اتمام رساندن مطالعه برخوردار است؟ امکان دارد تهیه و اعتباریابی ابزار تحقیق و انجام تحلیل‌های آماری پیچیده ضروری باشد.

ج - مسئله مورد نظر تحقیق باید از محیطی که خود محقق نیز در آن حضور دارد، انتخاب شود. او باید مطمئن باشد که اطلاعات ضروری برای پاسخگویی به سؤال موردنظر قابل دستیابی است یا خیر. او باید بررسی کند تا اطمینان یابد که آزمودنیهای موردنیاز با پیشینه‌های تحصیلی مناسب، قابل دستیابی هستند. مسئولان مدارس اغلب اوقات با انجام تحقیق در مدرسه‌شان مخالف هستند. آنها به سهولت به دانش‌آموزان اجازه شرکت در بررسی را نمی‌دهند، بنابراین جز آنکه محقق در زمان بررسی در مدرسه‌ای مشغول به کار باشد، وسیله‌ای برای حل مشکل خود نخواهد داشت. یکی از نویسندگان کتاب حاضر قبل از آنکه بتواند اجازه انجام یک تجربه را در زمینه تعلیم و تربیت به دست آورد، مجبور شد که به چهار مدرسه مراجعه کند.

د - مسئله باید قابلیت بررسی و تکمیل شدن در زمانی معین را داشته باشد. مسئله‌ای را انتخاب نکنید که بسیار بزرگ و یا درگیر کننده باشد. از برنامه‌ریزی زمان مورد نیاز برای ساخت وسایل اندازه‌گیری و بکارگیری این وسایل، تجزیه و تحلیل اطلاعات و نوشتن گزارش مطمئن باشید.

### /// بیان مسئله

بعد از انتخاب و تعیین ویژگیهای مسئله، هنوز وظیفه صورت‌بندی یا بیان مسئله، به شکلی که پاسخگویی تحقیق باشد، وجود دارد. اگر بخواهیم یک مسئله خوب بیان شود، باید (۱) آن چیزهایی را که قرار است تعیین و یا حل شوند، دقیقاً روشن کنیم و (۲) دامنه بررسی را به سؤالی خاص محدود نماییم. ما نمی‌توانیم بر اهمیت بیان روشن و موجز سؤال تحقیق تاکید بیش از حد داشته باشیم. نیاز به گفتن نیست که فرد قبل از آنکه بتواند کار زیادی برای حل مسئله انجام دهد، باید آن را تعیین کند. محقق تازه‌کار دارای یک فکر کلی در مورد مسئله است، اما در صورت‌بندی آن به عنوان یک سؤال عملی تحقیق دچار مشکل می‌شود. او در می‌یابد که فکر کلی اولیه‌اش، اگر چه برای انتقال آن به دیگران و درک آن از طرف دیگران کافی به نظر می‌رسد، ولی به قدر کافی مشخص نیست تا به او اجازه دهد که به طور تجربی با مسئله مواجه شود و او تا زمانی که به یک بیان عینی از سؤال تحقیق دست نیابد، نمی‌تواند تحقیق خود را به پیش ببرد. برای روشن شدن مطلب، فرض می‌کنیم یک محقق تازه‌کار اظهار می‌کند که به مطالعه در مورد تأثیر برنامه‌دستی جدید علوم در مدارس دوره متوسطه، علاقه‌مند است. همانطور که گفته شد، می‌توان منظور محقق را به طور کل درک و با او به صورتی کلی ارتباط برقرار کرد، ولی اگر او خواهان یافتن روشی برای حل مسئله است، باید مسئله را به روشنی تمام مشخص کند.

گام اول با تعریف اصطلاحات تحقیق آغاز می‌شود. منظور از تأثیر، برنامه درسی علوم و مدارس دوره متوسطه چیست؟ تعاریف مورد نیاز تحقیق اکثراً از طریق فرهنگ لغات کامل نمی‌شود. برای مثال، تأثیر، خلق «نتیجه مورد نظر و یا مورد انتظار» تعریف می‌شود. این تعریف ساخت کلی تأثیر را توصیف می‌کند، ولی برای هدف تحقیق بسنده نیست. ما باید بتوانیم دقیقاً مشخص کنیم که محقق از چه چیزی به عنوان شاخص تأثیر استفاده خواهد کرد یا محقق برای سنجش وجود یا فقدان پدیده‌ای که از آن به عنوان مفهوم تأثیر نام می‌برد، چه کاری انجام خواهد داد. این امر در مورد سایر اطلاعات نیز صدق می‌کند. به عبارت دیگر، محقق باید متغیرهای



مسئله را به طور کاربردی تعریف کند. جهت تعریف کاربردی، محقق باید برای ارائه مفاهیم مورد نظر، برخی از رفتارها و یا وقایع آشکار را که مستقیماً به وسیله خود محقق و یا سایر افراد قابل مشاهده و اندازه‌گیری هستند، مشخص کند. همانطور که در فصل اول تشریح شد، تعریف کاربردی، تعریفی است که یک مفهوم را بر حسب عملیات و یا فرآیندهایی که برای اندازه‌گیری مفاهیم به کار می‌روند، تعریف کند.

در این بررسی محقق باید تاثیر را یا به عنوان پیشرفت در نمرات یک آزمون تفکر انتقادی و یا آزمون استاندارد شده علوم تعریف کند. اصطلاح برنامه درسی به مثابه دوره زیست‌شناسی برنامه‌ریزی شده که به دانش‌آموزان سال دوم دبیرستان ارائه می‌شود، تلقی می‌شود. مدارس دوره متوسطه می‌توانند به دبیرستانهایی اطلاق شود که از لحاظ تعداد دانش‌آموز، نحوه فعالیت و غیره، ویژگیهای معینی را دارا باشند. مسئله اولیه اکنون به این سؤال تبدیل می‌شود: «تاثیر دوره زیست‌شناسی برنامه‌ریزی شده در درک مفاهیم زیست‌شناختی دانش‌آموزان سال دوم دبیرستان، که درس زیست‌شناسی را تازه شروع کرده‌اند، چیست؟» تعاریف کاربردی، دامنه سؤال کلی را به متغیرهای قابل مشاهده ویژه محدود می‌کند.

اکنون کار تحقیق با وضوح و تمرکز بیشتر نشان داده شده است. محقق می‌تواند با طراحی یک مطالعه آزمایشی برای مقایسه نمرات پیش و پس آزمون مفاهیم زیست‌شناختی دانش‌آموزانی که در دوره زیست‌شناسی برنامه‌ریزی شده شرکت کرده‌اند و دانش‌آموزان مشابهی که برنامه زیست‌شناسی جایگزین آن را گذرانده‌اند، کار تحقیق را به پیش برد. اکنون محقق شواهدی عینی درباره یک برنامه درسی خاص و موقعیت خاص را گردآوری کرده است که می‌تواند پرتوی بر سؤال کلی اولیه بیفکند.

بعلاوه، در تبیین مسئله، محقق باید برای ایجاد تعادل بین کلیات و جزئیات کوشش کند. اگر مسئله بیان شده بسیار کلی باشد، محقق با پهنه‌ای مبهم و بدون هیچ نشانه آشکاری از مسیر تحقیق روبرو خواهد بود. برای مثال سؤالی نظیر «تاثیر آموزش برنامه‌ریزی شده بر پیشرفت علمی در دبیرستان چیست؟»، بسیار کلی است. خیلی بهتر خواهد بود که بپرسیم «اثر استفاده از دوره‌های جبر برنامه‌ریزی شده بر پیشرفت دانش‌آموزان تیزهوش دبیرستان چیست؟». این نحوه بیان فوراً تعداد نمونه مورد نیاز، متغیرهای مرتبط، و نوع اطلاعاتی را که گردآوری خواهند شد، به نمایش می‌گذارد.

از سوی دیگر، مسئله نباید آنقدر کوچک باشد که به نظر ناچیز و بی معنا جلوه کند. محقق طالب موضوعی است که به قدر کافی جامع باشد تا بر مبنای معیارهای بحث شده با اهمیت جلوه کند، در عین حال آن قدر مشخص باشد که برای دانشجو در موقعیت خاص او انجام پذیر شود. هنگامی که مسائل تحقیقاتی بر رابطه بین دو یا چند متغیر متمرکز باشند، این بدان معنا نیست که عبارت «رابطه بین — و — چیست؟» در بیان مسئله آمده باشد. بیان مسئله ممکن است در این قالب بیاید یا وجود رابطه را به طور ضمنی برساند و در غیر این صورت، بیان مسئله باید در قالب سؤال مطرح شود. دانشجویان اغلب در این مرحله گیج می شوند. برای مثال، (۱) بیان اخباری که «این بررسی اثر مربوط به روش خواندن الف و ب را در آموزش افراد کند آموز بررسی می کند» و (۲) بیان سؤالی که «اثر بخشی روش خواندن الف در مقایسه با روش ب در آموزش افراد کند آموز چیست؟»، رابطه بین متغیرها را پی جویی می کنند، بدون آنکه عبارت مذکور را به کار ببرند، و هر دو شیوه های قابل پذیرشی برای ارائه مسئله تحقیق هستند. همان گونه که قبلاً گفته شد، بعضی از محققان صرفاً شیوه سؤالی را ترجیح می دهند؛ چون این شیوه ساده و روشن است و از نظر روان شناختی نیز چنین به نظر می رسد که محقق را به سمت وظیفه اش، یعنی طراحی روشی برای یافتن پاسخ سؤالش، جهت می دهد. در بعضی از موارد، هم بیان اخباری و هم سؤال، توأمان مطرح می شوند: بیانهای اخباری عام تر ممکن است با یک یا چند سؤال بسیار ویژه دنبال شوند.

هنگامی که یک سؤال بالقوه صورت بندی می شود، باید روشن بودن و قابلیت انجام آن مورد ارزیابی قرار گیرد. بسط یک مسئله قابل انجام، یک فرایند تکاملی است که متضمن بسیاری از کوششها برای روشن سازی مفاهیم، تعریف شیوه عمل و بررسی راههای گردآوری داده ها است.

### /// تعیین جامعه مورد بررسی و متغیرها

یک راهبرد مناسب برای شکلدهی مسئله مورد نظر - با یک تصور مبهم از چیزی که فرد تمایل به بررسی آن دارد - در قالب یک مسئله قابل تحقیق، تفکر در مورد جامعه آماری تحقیق و متغیرهاست.

برای مثال، اجازه بدهید به مسئله مورد نظر خانم بورک<sup>(۱)</sup>، یکی از مسئولان دوره ابتدایی،

توجه کنیم که آیا تدریس خصوصی توسط دانش‌آموزان سالهای بالاتر به دانش‌آموزانی که متوسط نمره قرائت آنها پایین است، اثر مثبت خواهد داشت؟

اغلب اوقات تعریف جامعه آماری بررسی، آسان‌ترین کارهاست، و آن افرادی هستند که محقق مشتاق دانستن چیزهایی درباره آنهاست. جامعه مورد نظر دانش‌آموزانی هستند که خواندن آنها از حد متوسط پایین‌تر است. در این سؤال، توانایی خواندن یک متغیر نیست؛ چون همه کودکان مورد نظر به عنوان دانش‌آموزان پایین‌تر از متوسط تلقی شده‌اند.

با تعیین دانش‌آموزان پایین‌تر از متوسط به عنوان جامعه مورد مطالعه در بیان اولیه مسئله، خانم بورک اکنون باید از خود سؤال کند که آیا این همان جامعه آماری دلخواه اوست؟ او احتمالاً باید تصمیم بگیرد که «دانش‌آموزانی که خواندن آنها پایین‌تر از سطح متوسط است» یک گروه بسیار وسیع است و لذا باید خود را به سن خاصی، یعنی دانش‌آموزان پایین‌تر از متوسط دوره دوم، محدود سازد.

اکنون او آماده جستجوی متغیرهای موجود در بیان اولیه مسئله خود است. تدریس خصوصی، به دلیل تغییر دادن انواع تدریس‌های مورد استفاده یا تغییر دادن مقدار زمان تدریس و یا دریافت تدریس خصوصی از سوی بعضی از کودکان و عدم دریافت آن از طرف بقیه، می‌تواند یک متغیر باشد. خانم بورک تصمیم می‌گیرد که این مورد آن چیزی است که او واقعاً می‌خواهد بداند. بنابراین او بخش مربوط به این سؤال را به این صورت بازنویسی می‌کند که: آیا دریافت یک میزان مشخص تدریس خصوصی در مقابل هیچگونه تدریس خصوصی...؟ بنابراین، تدریس خصوصی متغیر غیر وابسته است؛ چون بر مهارت قرائت مقدم است، و نکته مهم پیش‌بینی این امر است که آیا تدریس خصوصی در مهارت خواندن که یک متغیر وابسته است، اثر خواهد داشت؟

اکنون آشکار شده است که کلمه تدریس خصوصی بسیار کلی است و بدون آنکه همه آزمودنیها یکسان و به یک میزان تدریس خصوصی دریافت‌کنند، نتایج بررسی بی‌معنی خواهد بود. خانم بورک تصمیم می‌گیرد که از تمرین یادآوری سریع کلمه، به عنوان نوع ویژه‌ای از تدریس خصوصی و ۱۵ دقیقه زمان در هر روز، به عنوان مقدار زمان استفاده کند.

عبارت یک تاثیر مثبت دارد، تا وقتی که خانم بورک به آن به مثابه متغیر غیر وابسته ننگرد، کاملاً مبهم است. تمرین یادآوری سریع کلمات بر چه ... تاثیر دارد؟ او می‌داند که بر یادآوری

سریع کلمه اثر دارد، ولی او می‌خواهد اثرات آنرا بر دیگر جنبه‌های رفتاری خواندن، که ممکن است مشاهده شود، بررسی کند: خواندن پر احساس با صدای بلند، بی‌صدا خوانی، احساسات مثبت نسبت به خواندن، تعداد کتابهای قرائت شده، ادراک و غیره. در میان دانش‌آموزان کلاس دوم که خواندن آنها از سطح حد متوسط پایین‌تر است، آیا تفاوتی در نمره قرائت CAT (آزمون پیشرفت کالیفرنیا)<sup>(۱)</sup> بین کسانی که روزانه ۱۵ دقیقه توسط دانش‌آموزان کلاسهای بالاتر به تمرین یادآوری کلمات می‌پردازند و کسانی که هیچ تمرینی در این زمینه نداشته‌اند، وجود دارد؟ او به یک متغیر وابسته فارغ از قضاوت معلم نیاز دارد و تصمیم می‌گیرد که از نمره قرائت آزمون پیشرفت کالیفرنیا به عنوان متغیر وابسته استفاده کند.

بیان مسئله تجدید نظر شده خانم بورک حالا اینگونه عنوان می‌شود: در میان دانش‌آموزان کلاس دوم که خواندن آنها از سطح حد متوسط پایین‌تر است، آیا در نمره قرائت CAT بین کسانی که روزانه ۱۵ دقیقه توسط دانش‌آموزان کلاسهای بالاتر به تمرین یادآوری سریع کلمات می‌پردازند و کسانی که هیچ تمرینی در این زمینه نداشته‌اند، تفاوتی وجود دارد؟ این سؤال می‌گوید که او در مورد چه کسانی بررسی می‌کند، برای بعضی، به نسبت بقیه، چه کار متفاوتی صورت خواهد گرفت و او انتظار چه اثری را در مقابل تیمار متفاوت خود دارد. همچنین توجه داشته باشید که قضاوت ارزشی تاثیر مثبت در این سؤال حذف شده است.

پیگیری این فرایند به شیوه متداول و مشابه با آنچه که در طراحی یک جمله به کار می‌رود، مفید است. فرد می‌تواند کار را با ترسیم یک خط عمودی و نوشتن جامعه آماری در سمت چپ و متغیرها در سمت راست آغاز کند. این عوامل موجود در بررسی سپس در پایین خط افقی فهرست می‌شوند. در مثال ما نمودار به صورت ذیل دنبال می‌شود.

جامعه آماری	متغیرها
دانش‌آموزان کلاس دوم که خواندن آنها از سطح متوسط پایین‌تر است.	- انجام تمرین هر روز به مدت ۱۵ دقیقه توسط دانش‌آموزان بالاتر از متوسط در مقابل انجام ندادن هیچ‌گونه تمرین
	- نمره خواندن در تست CAT

اجازه بدهید سؤال دیگری را مطرح کنیم. اثر تجربه کردن یک برنامه پیش دبستانی با نداشتن چنین تجربه‌ای در پیشرفت خواندن در کلاس اول دبستان چیست؟

متغیرها	جامعه آماری
- تجربه کردن یا نکردن یک برنامه پیش دبستانی (متغیر غیروابسته) - پیشرفت خواندن (متغیر وابسته)	کلاس اولی‌ها

این سؤال کامل است؛ زیرا یک جامعه آماری معین و دو متغیر دارد. از آنجا که برنامه پیش دبستانی بر پیشرفت خواندن کلاس اولی‌ها مقدم است، اولی می‌تواند به عنوان متغیر مستقل و دومی به عنوان متغیر وابسته تعیین شوند.

اجازه دهید مثال دیگری را بررسی کنیم: آیا آموزش رانندگی در دبیرستان فایده‌ای دارد؟ همانطور که اشاره شد، این سؤال نه جامعه آماری دارد و نه متغیر. یک محقق که با این سؤال شروع می‌کند، نخست باید تصمیم بگیرد رانندگان ۱۹ ساله‌ای را که آموزش رانندگی دوره دبیرستان را دیده‌اند، با آنهایی که این آموزش را ندیده‌اند، مقایسه کند. حالا ما یک جامعه آماری معین و یک متغیر مستقل داریم. اکنون می‌توانیم توجه خود را به انتخاب یک متغیر وابسته معطوف کنیم. آموزش رانندگی یا عدم آموزش آن در دوره دبیرستان ممکن است چه اثری بر رانندگان ۱۹ ساله داشته باشد؟ چنین تصمیم گرفته شده است که نرخ تصادف می‌تواند یک متغیر وابسته مناسب باشد. با گذاشتن این عوامل در یک نمودار، خواهیم داشت:

متغیرها	جامعه آماری
- گرفتن و یا نگرفتن تعلیم رانندگی در دوره دبیرستان (متغیر مستقل) - نرخ تصادف (متغیر وابسته)	رانندگان ۱۹ ساله

اکنون می‌توان یک سؤال کامل را مطرح کرد: آیا نرخ تصادف رانندگان ۱۹ ساله‌ای که در دوره دبیرستان، تعلیم رانندگی دیده‌اند، از رانندگان ۱۹ ساله‌ای که تعلیم رانندگی ندیده‌اند، کمتر است؟ این سؤال که چه رابطه‌ای بین جزم اندیشی و نگرشهای سیاسی در میان دانشجویان سال اول دانشکده وجود دارد، نکته دیگری را روشن می‌کند.

متغیرها	جامعه آماری
- جزم اندیشی - نگرشهای سیاسی	دانشجویان سال اول دانشکده

این سؤال با داشتن جامعه آماری و دو متغیر، کامل است. اما تا وقتی که نتوان تعیین کرد که کدامیک بر دیگری مقدم است، نخواهیم توانست متغیرها را به عنوان متغیر مستقل و وابسته برچسب بزنیم.

اگر بررسی انجام بگیرد که به جای رابطه بین متغیرها، وضعیت موجود را بسنجد، ممکن است تنها با یک متغیر کامل شود. برای مثال، محقق ممکن است عقیده دانشجویان دانشکده را در مورد قانونی شدن مصرف ماری جوآنا، مورد بررسی قرار دهد. در این مورد، جامعه آماری، دانشجویان ارشد دانشکده و تنها متغیر، عقاید آنها در مورد موضوع است.

روشهای مختلفی برای پاسخ دادن به انواع سوالات تحقیقاتی به کار گرفته می شود. هر گاه یک متغیر مستقل بتواند توسط محقق دستکاری شود، روشهای آزمایشی مناسب هستند (به فصل ۹ مراجعه کنید). اولین مثال در این بخش مربوط است به اثر تمرین یادآوری سریع کلمه در آزمون قرائت CAT که یک تحقیق آزمایشی است. بسیاری از متغیرها در امر آموزش نمی توانند مورد دستکاری قرار گیرند، بنابراین روش آزمایشی نمی تواند در بررسی این متغیرها مورد استفاده قرار گیرد. موضوع تحقیقی مربوط به مقایسه نرخ تصادف رانندگان ۱۹ ساله ای که تعلیم رانندگی گرفته اند و یا نگرفته اند، احتمالاً روش علی - مقایسه ای را می طلبد (به فصل ۱۰ مراجعه کنید). متغیر مستقل (گرفتن تعلیم رانندگی و یا نگرفتن تعلیم رانندگی) توسط محقق مورد دستکاری قرار نگرفت.

برای بعضی از موضوعات تحقیقی، بسته به آنکه چگونه طراحی شده باشند، هم روش آزمایشی و هم علی - مقایسه ای، مناسب است. در مثال تحقیقاتی یادشده در مورد اثر برنامه های پیش دبستانی در پیشرفت خواندن، محقق بسته به آنکه دستکاری متغیر مستقل، یعنی برنامه های پیش دبستانی را برگزیند یا آزمودنی هایی را انتخاب کند که قبلاً تربیت پیش دبستانی داشته و یا نداشته اند، می تواند از هر یک از این دو روش استفاده کند.

موضوعات تحقیقی دیگر با توصیف ماهیت وقوع یک یا چند متغیر آموزشی سروکار دارند.

روش توصیفی یکی از مناسبترین روشها برای این گونه مسائل است (به فصل ۱۱ مراجعه کنید). مثال مربوط به رابطه بین جزم اندیشی و نگرشهای سیاسی بین دانشجویان کالج، یک تحقیق توصیفی است. سنجش عقاید در مورد مسائلی از قبیل قانونی شدن مصرف ماری جوآن نیز یک تحقیق توصیفی است.

## // مقاله مجلات

تعیین جامعه آماری و متغیرهای مستقل و وابسته، چهارچوبی برای تجزیه و تحلیل گزارشهای تحقیقاتی چاپ شده به دست می‌دهد. مارش در مقاله تحقیقات آموزشی عواملی را به نمایش می‌گذارد که معمولاً در گزارش هر تحقیق آزمایشی مشاهده می‌شود.<sup>(۱)</sup> مقدمه این گزارش دلایل منطقی برای انجام بررسی و نیز تاریخچه کوتاهی از یافته‌ها و تحقیقات موجود را در مورد موضوع ارائه می‌کند. در بخش روشها، عوامل متعددی وجود دارد: (۱) نمونه (۱۰ کلاس از ۵ دانشگاه) که انتخاب شده است تا معرف جامعه آماری مورد نظر (دانشجویان دانشگاه) باشد؛ (۲) متغیر مستقل (امتحان در خانه، در مقابل امتحان کلاس) توصیف و به طور کاربردی تعریف شده است؛ (۳) متغیرهای وابسته (پاسخگویی دانشجویان به سؤال تحقیق در مورد اینکه آنها چقدر انگیزه دارند تا جهت آمادگی برای هر یک از انواع امتحانات و کسب نمرات خاصی از طریق یک آزمون درک و بکارگیری مطلب، که بدون اطلاع قبلی صورت می‌گیرد، به مطالعه و یادگیری بپردازند) توصیف و به طور کاربردی تعریف شده‌اند؛ (۴) شیوه‌ها با جزئیات کافی توصیف شده‌اند تا سایر محققان را قادر سازند این آزمایش را با گروه نمونه متفاوت تکرار کنند. بخش نتایج، داده‌های حاصله از آزمایش و نتایج آزمونهای آماری را به دست می‌دهد. بخش نتیجه‌گیری حاوی تفسیر محقق از این نتایج است.

### مقایسه‌ای بین امتحان در خانه، در مقابل امتحان در کلاس

این بررسی تلاشی بود برای تعیین اینکه آیا برای آنکه یادگیری صورت گیرد، امتحان در خانه به اندازه امتحان سنتی در کلاس خوب است، یا از آن بهتر است. نتایج این آزمایش مهم است؛ چون کارشناسان امر آموزش به طور فزاینده‌ای در حال حذف امتحان در سر کلاس هستند (گی و

گالاگر<sup>(۱)</sup>، ۱۹۷۶)، و به نظر می‌رسد هیچ پایه عملی برای این باور که امتحان انجام شده در منزل شیوه‌ای مؤثر است، وجود ندارد. بنابراین، روندهای جاری در جهت حذف امتحانهای کتبی رسمی ممکن است در واقع برخلاف یکی از اهداف اساسی آموزش و پرورش، یعنی یاد داری مفاهیم آموخته شده از طرف دانش‌آموزان، عمل کند.

احتمالاً نگرانی اصلی از امتحانهای کلاسی این است که این امتحانات موجب اضطرابهای تضعیف‌کننده می‌شوند؛ هر چند تحقیق در مورد این ادعا نتایج متناقضی به بار آورده است. برای مثال، ثابت شده است که در شرایط امتحانی، اضطراب یک اثر منفی دارد (ساراسون، دیویدسون، لایتال، ویت، و روبوش، ۱۹۶۰)<sup>(۲)</sup>، یک اثر مثبت دارد (کاستاندا، مک کندلس، و پالرمو، ۱۹۶۴)<sup>(۳)</sup> و بی تاثیر است (دنی، پاترسون، و فلدهاسن، ۱۹۶۴)<sup>(۴)</sup>. فرضیه بعضی از محققان این است که این میزان اضطراب است که در عملکرد امتحان تاثیر می‌گذارد. آناستازی<sup>(۵)</sup> (۱۹۷۴) معتقد است که امتحان، اگر به طور درست گرفته شود، یک اثر کلی مثبت دارد. او می‌گوید که رابطه بین اضطراب و عملکرد امتحان احتمالاً خطی نیست و بنابراین مقداری اضطراب مفید است؛ در حالی که میزان زیاد آن زیانبخش است. آزمایشهای زیادی برای از میان بردن اضطراب در شرایط امتحان کلاسی طراحی شده است، اما آناستازی اظهار می‌کند که «خود روش آزمونگران و نیز انجام یک امتحان خوب سازماندهی شده و از آسان به مشکل، به همین هدف دست خواهد یافت» (آناستازی، ۱۹۷۶).

تجربیات دانشجویان سطوح مختلف در مورد اضطراب آنها قبل از امتحانات، احتمالاً حذف امتحان را به دلیل ایجاد اضطراب تایید نمی‌کند و هدف ما نباید احتراز کامل از تنش باشد؛ چرا که این امر در دنیای ما امکان‌ناپذیر است، اما آموختن چگونگی تشخیص پاسخ نوعی به تنش و سپس تلاش برای سازگار کردن شیوه مواجهه ما با آن، امری شدنی است (سلی، ۱۹۷۸)<sup>(۶)</sup>.

پیشینه تحقیق در خصوص مقایسه امتحانهای انجام شده در خانه و کلاس بسیار قلیل است. شواهدی وجود دارد که امتحان شفاهی و کتبی هر دو به یک میزان تاثیر دارند و هر دوی این روشها بهتر از اصلاً امتحان نکردن است (کلهن، ۱۹۶۲)<sup>(۷)</sup>.

1- Gay and Gallagher

2- Sarason, Davidson, Lighthall, Walte, and Ruebush, 1960

3- Castaneda, McCandless, and Palermo, 1964

5- Anastasi

7- Calhoun, 1962

4- Denny, Paterson, and Feldhusen, 1964

6- Selye, 1978



همچنین، بررسیها نشان می‌دهد که پیشرفت تحصیلی دانشجویان دوره لیسانس بر اساس نظام رده‌بندی قبول / رد، کمتر از دانشجویان تحت نظام رده‌بندی استاندارد است (بین، هیلز، و راند، ۱۹۷۳؛ گلد، ریلی، سیلبمن، و لهر، ۱۹۷۱)<sup>(۱)</sup>.

در سال ۱۹۷۶ یک بررسی (گی و گالاگر، ۱۹۷۶) انجام شد که در آن یک کلاس دوره لیسانس به طور تصادفی به سه قسمت تقسیم شده بود. یک گروه تمرینهای دوره‌ای دریافت می‌کردند تا در خانه انجام دهند، گروه دوم آزمونهای دوره‌ای دریافت می‌کردند؛ و به سومین گروه حق انتخاب یکی از این دو روش داده شد. جالب است اشاره کنیم که هیچیک از افراد گروه سوم، آزمون دوره‌ای را انتخاب نکردند. بجز این نحوه برخورد متفاوت، همه دانشجویان در حد ممکن یکسان رفتار کردند. در پایان دوره، یک آزمون غیر منتظره یکسان به هر سه گروه داده شد. نتایج این آزمون نشان داد نمرات گروهی که قبلاً آزمون دوره‌ای دریافت می‌کردند، به طور قابل ملاحظه‌ای از دو گروه دیگر بیشتر بود.

این بررسی از جنبه‌های مختلف با سایر تحقیقات متفاوت است. نخست آنکه، این بررسی کلاسهای متنوعی را در رشته‌های ادبی، رفتاری و کاربردی در بر می‌گرفت؛ بنابراین نتایج احتمالاً قابل تعمیم به مخاطبان وسیعتری هستند. در این بررسی تلاش بر این بود تا سطوح ویژه یادگیری بر اساس سلسله مراتب بلوم در یادگیری شناختی اندازه‌گیری شود (بلوم، انگل‌هارد، فارست، هیل، و کرث‌هول، ۱۹۵۶).<sup>(۲)</sup> بالاخره تلاش شد که با توجه به نتایج آزمایش، علت مسئله شناسایی شود.

### روشها

دانشجویان ۱۰ کلاس از ۵ دانشگاه، به عنوان شرکت کنندگان این بررسی انتخاب شدند. مجموع نمونه را ۲۵۸ نفر دانشجو از کلاسهای روان شناسی، آمار، تعلیم و تربیت، تحقیق و تاریخ تشکیل می‌دادند. در شروع ترم تحصیلی استاد به همراه یک پاکت محتوی دستورالعملهای دقیق، توضیح کوتاهی در مورد شیوه‌های کار به هر دانشجو می‌داد. به دانشجویان گفته می‌شد که نیمی از آنان یک امتحان در خانه دریافت خواهند کرد، که یک

1- Bain, Hales, and Rand, 1973; Gold, Relly, Silbeman, and Lehr, 1971

2- Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956

آزمایش علمی است، هیچ کس از نتیجه‌ای که به دست می‌آورد، لطمه نخواهد خورد و هر کس که از نمره خود رضایت نداشته باشد، حق انتخاب یک آزمون دیگر را خواهد داشت. انتظار می‌رفت که این شیوه در عامل اضطراب اثر خنثی سازی داشته باشد.

حدود دو هفته قبل از شروع نیمه دوم دوره، کلاسها به طور تصادفی به دو گروه مساوی الف و ب تقسیم شد. در هر کلاس، به افرادی که به طور تصادفی در گروه الف انتخاب شده بودند، یک امتحان در منزل داده شد که می‌باید یک هفته بعد برمی‌گرداندند. به اعضای گروه ب یک آزمون یکسان در سر کلاس داده شد. یک هفته بعد در سر کلاس عادی به هر دو گروه یک آزمون غیر منتظره داده شد. در آزمون اخیر همان مواد درسی گنجانده شده بود که هفته قبل نیز مورد آزمون قرار گرفته بودند. ضمن آنکه اعضای هر دو گروه تحت شرایطی یکسان و زیر نظر یک استاد آموزش می‌دیدند، گروه آزمایشی و گروه گواه در حد امکان به طور یکسان در معرض آزمایش از نوع کلاسی قرار گرفتند. تنها روش (نوع آزمون) برای هر کلاس متفاوت بود.

آزمون غیر منتظره حاوی ۵۰ سؤال چند جوابی بود، که ۲۵ سؤال اطلاعات دانشجویان و ۲۵ سؤال دیگر ترکیب فهم مطلب و کاربرد محتوای دوره را می‌سنجید. آخرین دسته‌بندیها، سه سطح اولیه یادگیری شناختی در سلسله مراتب بلوم را (بلوم و دیگران، ۱۹۵۶) شامل می‌شد. برای کمک به تعیین میزان انگیزه هر دانشجو در خواندن و یاد گرفتن درس، جهت آماده شدن برای هریک از انواع امتحان، یک پیمایش صورت گرفت. ۹۶/۱ درصد دانشجویان به این پرسشنامه پاسخ دادند. در نهایت، یک تحلیل واریانس دو عاملی مورد استفاده قرار گرفت. از آنجا که دو گروه از ۱۰ کلاس مختلف انتخاب شده بودند، طرح بررسی را یک جدول  $2 \times 10$  تشکیل می‌داد (اری و جیکوبز، ۱۹۷۶)<sup>(۱)</sup>.

### نتایج

نمرات گروه ب (در آزمون کلاسی) به طور قابل ملاحظه‌ای از گروه الف (آزمون در منزل) در همه طبقه بندیهای آزمون غیر منتظره، بالاتر بود.

گروه آزمایشی به طور قابل توجهی یادگیری بیشتر، دانش بیشتر، و فهم بیشتری را از مطالب و کاربرد مواد درسی به نمایش گذاشت. به این ترتیب تجزیه و تحلیل‌های آماری در مورد هر طبقه

نشان داد که متغیر مستقل (نوع آزمون) و متغیر وابسته (یادگیری) به شدت به هم مربوطند. در پیمایشی که در مورد آزمون غیر منتظره انجام شد، یکی از سؤالات این بود: «اگر شما یک آموزگار بودید و می‌خواستید شاگردانتان مراد درسی را بواقع یاد بگیرند، کدامیک از انواع آزمون را به آنها می‌دادید؟» ۲۵/۸ درصد از پاسخگویان اظهار داشتند که آزمون قابل انجام در منزل را خواهند داد، ۵۹/۱ درصد آزمون کلاسی را انتخاب کردند، و ۱۵/۱ درصد قادر به تصمیم‌گیری نبودند. این طور به نظر می‌رسید که آنها به نسبت ۲ به ۱ این روش را ترجیح می‌داده‌اند.

دو سؤال برای تعیین نگرش آزمودنیها در مورد نوع خاصی از آزمون، طراحی شده بود. در اولین سؤال از دانشجویان خواسته می‌شد تا مجموعه تلاش خود را برای آموختن مطالب موجود در آزمون معلم ساخته آنها که هفته پیش به عنوان آزمون غیر منتظره دریافت کرده بودند، در یک پیوستار ده درجه‌ای، مشخص کنند. در سؤال دوم، که آن هم شامل یک پیوستار ده درجه‌ای بود، از آنها خواسته شد تا پاسخ دهند چه میزان تلاش برای انجام دیگر انواع آزمون (آزمون قابل انجام در منزل یا کلاسی) بایستی در نظر گرفته شود.

امتیاز گروه الف به طور قابل ملاحظه‌ای در سؤال اول پایین تر بود. بنابراین، دانشجویان اظهار داشتند که آنها در صورت داشتن آزمون کلاسی، تلاش بیشتری از خود نشان خواهند داد. بر عکس امتیاز گروه ب به طور قابل ملاحظه‌ای در سؤال دوم پایین تر بود. محقق می‌تواند این نتیجه را چنین تفسیر کند که دانشجویان در صورت داشتن آزمون قابل انجام در منزل کمتر درس خواهند خواند. نتیجه‌ای که تفاوت در یادگیری بین گروهها را توضیح می‌دهد، این است که آن دسته از دانشجویانی که آزمون کلاسی داشته‌اند، به اعتراف خود برای مطالعه تلاش بیشتری کرده‌اند.

### نتیجه گیری

این داده‌ها شاهی است بر مقبولیت آزمون کلاسی، در مقابل آزمون قابل انجام در منزل، در بین دانشجویان مختلف دانشگاه، در رشته‌های روان‌شناسی، آمار، تحقیق، تعلیم و تربیت و تاریخ. در این آزمایش دو نوع آزمون به کار گرفته شد. در مجموع، یادگیری دانشجویان در سه سطح اولیه سلسله مراتب بلوم در یادگیری شناختی، یعنی دانش، فهم مطالب و کاربرد، تاثیر داشته است. به طور وضوح آزمون کلاسی شیوه بهتری بود، و به نظر می‌رسید که دانشجویان با آن موافق هستند. اما، به دلایل بسیار، محقق نمی‌تواند اظهار کند که آزمونهای سنتی کلاسی

چون یادگیری مؤثر را افزایش می‌دهند، تنها روش معتبر هستند. با تأکید بر آرمانها و اهداف آزمونها به تحقیق بیشتری در مورد روشهای اندازه‌گیری و مقایسه سطوح یادگیری در دامنه‌ای وسیع نیاز است.

### /// خلاصه

اولین وظیفه‌ای که در پیش روی محققان است، انتخاب و صورت‌بندی یک مسئله است. یک مسئله تحقیقی، عبارتی سؤالی یا اخباری در مورد رابطه موجود بین متغیرهاست. در تلاش برای یافتن یک مسئله قابل تحقیق، پژوهشگر می‌تواند به تجربیات شخصی خود، نظریه‌هایی که می‌توان از آنها سؤالاتی استنتاج کرد یا به متون تحقیقی جاری در حوزه‌های مورد نظر مراجعه کند. آنها باید اهمیت مسئله پیشنهادی را برحسب معیارهای ویژه ارزیابی کنند، یعنی از خود بپرسند: آیا این مسئله در پیکره دانش امروزی مؤثر خواهد بود؟ آیا این مسئله امکان حرکت به سمت تحقیق آتی را فراهم می‌آورد؟ آیا قابل آزمون است، یعنی آیا متغیرها می‌توانند مشاهده و اندازه‌گیری شوند؟ مسئله تا چه حد با علایق، تجربیات و دانش من در این زمینه متناسب است؟ آیا به داده‌هایی که برای این مسئله مورد نیاز است، دسترسی دارم و آیا وسایل اندازه‌گیری موجود است، یا می‌توان آنها را برای اندازه‌گیری متغیرها ساخت؟ سؤالات نباید مستقیماً با مسائل فلسفی درگیر و همچنین نباید آنقدر کلی باشند که تحقیق آن غیرممکن به نظر رسد. بیان پرسش باید جامعه آماری مورد نظر و متغیرهای مورد بررسی را معین کند.

### // مفاهیم اصلی

بیان مسئله	جامعه آماری
متغیر مستقل	متغیر وابسته
مسئله تحقیق	معیارهایی برای مسئله تحقیق
منابع موجود در زمینه مسئله تحقیق	

### /// تمرینها

- ۱- یک گزارش تحقیقی چاپ شده در یک مجله پیدا کنید و بر اساس مطالب خوانده‌شده به پرسشهای زیر پاسخ دهید:

الف - چه مسئله‌ای در این بررسی مورد پژوهش قرار گرفته است؟

ب - فرضیه (ها) کدامند؟

ج - متغیرهای مستقل و وابسته کدامند؟

د - مسئله و فرضیه (های) بیان شده در گزارش را در کجا یافتید؟

ه - آیا مسئله و فرضیه (ها) آنقدر روشن بیان شده‌اند که شما بدانید چه چیزی دقیقاً باید

مورد پژوهش قرار گیرد؟

۲- در حیطه‌ای که احتمالاً به تحقیق در آن علاقه دارید، یک زمینه وسیع را انتخاب کنید. یک

جنبه خاص از این زمینه وسیع در نظر بگیرید و سپس یک مسئله تحقیق را که به آن علاقه‌مند

هستید، برگزینید. این مسئله را در یک قالب تحقیقی بیان کنید.

۳- نمونه‌های زیر بیانهای نامناسبی از مسائل تحقیقی هستند. هر یک را دوباره چنان بیان

کنید که به یک سؤال خاص مناسب برای تحقیق تبدیل شوند.

الف - اثرات شیوه‌های مختلف یادگیری مفاهیم علمی

ب - اضطراب و پیشرفت تحصیلی

ج - نگرشهای کودکانی که دارای محرومیت فرهنگی هستند.

د - مشاوره و دانش‌آموزان ضعیف

ه - اثر بخشی روش کوئیزنر<sup>(۱)</sup> در تدریس ریاضیات ابتدایی

۴- مسائل تحقیقی ذیل را ارزشیابی کنید:

الف - آیا فلسفه دکتر اسپاک در مورد تربیت کودکان سهل‌انگار، تأثیر نامطلوبی در آموزش و

پرورش آمریکا داشته است؟

ب - چه رابطه‌ای بین شیوه پا روی پا انداختن و هوش دانشجویان زن در کالج آمریکایی

وجود دارد؟

ج - با توجه به بررسیهای کاربردی اخیر در پیشرفت ریاضیات، آیا باید ریاضیات «جدید» در

مدارس ابتدایی منسوخ شود و روش سنتی ریاضیات جایگزین آن شود؟

د - دانش‌آموزان نقش مدیر مدرسه را در Central Middle School چگونه تعبیر می‌کنند؟

ه - محتمل‌ترین متغیرهای مستقل و وابسته را در بررسیهای زیر بیان کنید:

- الف - لین، اس، و ج، برگن<sup>(۱)</sup> (۱۹۸۸). اثرات متغیرهای آموزشی در توانایی یادگیری زبان در بین کودکان پیش دبستانی. *مجله تحقیقات آموزشی آمریکا*<sup>(۲)</sup>، ۲۸۴ - ۲۷۱ و ۲۵.
- ب - لین، آر و سی هاستنیزگر<sup>(۳)</sup> (۱۹۸۴). اعتبار پیش‌بینی‌کننده‌های موفقیت در دانشکده حقوق. *مجله اندازه‌گیری آموزشی*<sup>(۴)</sup>، ۲۵۸ - ۲۴۵ و ۲۱.
- ج - هاتیوا، ان<sup>(۵)</sup> (۱۹۸۸). مشق کامپیوتری و تمرین در حساب: وسیع تر شدن شکاف بین دانش‌آموزان قوی و ضعیف. *مجله تحقیقات آموزشی آمریکا*، ۳۹۷-۳۶۶ و ۲۵.
- د - چاپمن، ج.<sup>(۶)</sup> (۱۹۸۸). خود پندارهٔ کودکان ضعیف در یادگیری. *مجله بررسی تحقیقات آموزشی*<sup>(۷)</sup>، ۳۶۶ - ۳۴۷ و ۵۸.
- ه - ویر، ان و وی لی<sup>(۸)</sup> (۱۹۸۸). تفاوت‌های جنسی در انتخاب رشته‌های علوم در دانشکده. *مجله تحقیقات آموزشی آمریکا*، ۶۱۴ - ۵۳۹ و ۲۵.
- ۶ - منابع مسائل تحقیقاتی مورد استفاده محققان را فهرست کنید و برای هر کدام یک مثال بیاورید.
- ۷ - خواندن کتابهای خوب چه تأثیری در مهارت‌های خواندن کودکان دارد؟ آیا این سؤال قابل تحقیق است؟ اگر نیست، برای آنکه بتوان آنرا به طور کاربردی مورد پژوهش قرار داد، چه تغییراتی را پیشنهاد می‌کنید؟

### /// پاسخها

- ۱- پاسخها متنوع خواهند بود.
- ۲- پاسخها متنوع خواهند بود.
- ۳- الف - آیا بین نمرات پیشرفت درس علوم دانش‌آموزان کلاس ششم انجمن تحقیق علم (SRA)<sup>(۹)</sup> که یک سال برنامهٔ آموزش علوم ابتدایی (ESS) را گذرانده‌اند، با نمرات آنهايي که دورهٔ یکساله مبتنی بر کتاب درسی علوم را دیده‌اند، تفاوتی وجود دارد؟

1- Lane, S., and J. Bergan (1988)

3- Linn, R., and C. Hastings (1984).

5- Hativa, N. (1988).

7- View of Educational Research, 58, 347 -366.

9- Science Research Associates (SRA)

2- American Educational Research Journal 25, 271 -284.

4- Journal of Educational Measurement, 21, 245- 258.

6- Chapman, J. (1988).

8- Ware, N., and V. Lee (1988).

ب - آیا بین نمرات آزمون ترکیبی استانفورد و نمرات مقیاس اضطراب آشکار<sup>(۱)</sup> در بین دانش‌آموزان دوره دبیرستان تفاوتی وجود دارد؟

ج - محقق ابتدا باید مشخص کند که کدامیک از نگرشهای کودکان مورد بررسی قرار گرفته‌اند و سپس آنها را با نگرشهای یک گروه کنترل مناسب (شاید کودکانی که دچار محرومیت فرهنگی نیستند) مقایسه کند، یا در مورد رابطه بین انواع نگرشها و دیگر رفتارها فرضیه‌ای ارائه کند. برای مثال، آیا کودکان دارای محرومیت فرهنگی با نگرش مثبت نسبت به مدرسه، از کودکان با نگرش منفی نسبت به مدرسه، نمرات بهتری کسب می‌کنند؟

د - اثر یک برنامه مشاوره بر نگرشهای دانش‌آموزان ضعیف دوره راهنمایی در مورد مدرسه چیست؟

ه - آیا نمرات درس حساب دانش‌آموزان کلاس سوم، که یک سال از روش کوئیزر در تدریس ریاضیات ابتدایی بهره‌مند بوده‌اند، با نمرات آنهایی که روش تدریس دیگری داشته‌اند، متفاوت است؟

۴- الف - این سؤال متضمن یک قضاوت ارزشی است که بررسی کاربردی آن غیرممکن است.

ب - این سؤال پیش پا افتاده است، و پاسخ آن سهم ناچیزی در دانش موجود دارد.  
ج - تحقیق قادر به پاسخگویی به سؤالهای ارزشی نیست؛ تنها قادر به فراهم ساختن اطلاعاتی است که بر مبنای آن بتوان تصمیماتی اتخاذ کرد.  
د - با اینکه سؤال می‌تواند مورد بررسی قرار گیرد، اما از جنبه تعمیم دادن به سایر موقعیتها بسیار محدود است.

۵ - الف - متغیر مستقل: روشهای آموزشی

متغیر وابسته: برخی از اندازه‌های توانایی زبان

ب - متغیر مستقل: پیش‌بینی‌های متعدد (نمره‌های آزمونها و غیره)

متغیر وابسته: بعضی اندازه‌ها در عملکرد دانشکده حقوق (شاید معلمان)

ج - متغیر مستقل: استفاده از تمرین کامپیوتری در ریاضیات.

متغیر وابسته: پیشرفت در درس حساب برای دانش‌آموزان ضعیف و قوی.

د - متغیر مستقل: ضعیف بودن در یادگیری

متغیر وابسته: بعضی اندازه‌های خود پنداره

۵ - متغیر مستقل: جنسیت

متغیر وابسته: انتخاب رشته علمی در دانشکده

۶ - تجربیات روزمره کارشناسان آموزش و پرورش

قیاس از نظریه

تحقیقات پیشین

منابع غیر آموزشی

۷ - اصطلاح کتاب خوب و مهارت خواندن نیاز به تعریف دارد. بیان مجدد آن اینگونه است

که اثر خواندن چهار متن کلاسیک مشخص، در گنجینه لغات کودکان چیست.



# فصل سوم

## تحقیقات پیشین

### هدفهای آموزشی

دانشجو پس از مطالعه این فصل، قادر خواهد بود:

- ۱- هفت کارکرد اصلی در تحقیق را که از مرور تحقیقات پیشین حاصل می‌شود، فهرست کند.
- ۲- منابع مرجع اصلی در تعلیم و تربیت را توضیح دهد.
- ۳- نمایه‌های اریک را نام ببرد و تشابهات و تفاوت‌های آنها را تشریح کند.
- ۴- یک رویکرد کلی برای استفاده نمایه‌های نشریات علمی در تحقیق را خلاصه کند.
- ۵- حداقل سه نمایه غیر از اریک را نام ببرد و از اطلاعات زمینه‌ای در هر کدام نمونه‌ای ذکر کند.
- ۶- چگونگی استفاده از SSCI را در مروری از تحقیقات پیشین تشریح کند و عملکرد هر یک از چهار فهرست را توضیح دهد.
- ۷- کاربرد سالنامه اندازه‌گیری روانی<sup>(۱)</sup> را به عنوان یک منبع اطلاعاتی در آزمون‌ها و ابزار اندازه‌گیری، توصیف کند.
- ۸- شیوه‌های یک جستجوی کامپیوتری از تحقیقات پیشین را توضیح دهد.
- ۹- پیشرفت نظام‌مند مرحله به مرحله در سازماندهی تحقیقات پیشین را شرح و مقصود از هر مرحله را توضیح دهد.

1- Mental Measurement Yearbooks.

این فصل با همکاری سامرال، تی. هرآنگ، بنیانگذار کتابخانه دانشگاه کارولینای شمالی، تجدید نظر شده و اطلاعات آن به روز شده است.

وقتی که موضوع انتخاب شد، محقق طبیعتاً مشتاق عمل خواهد بود. به هر حال حمله شتابزده برای طراحی و انجام بررسی، قبل از یافتن زمینه کاملی از آنچه که پیش از این در حیطة مورد نظر انجام یافته، یک اشتباه است. محقق باید موضوع مورد نظر خود را به دانش موجود در زمینه مورد علاقه خود ارتباط دهد. قابلیت تعیین، سازماندهی و استفاده از منابع در زمینه مورد نظر، برای کارشناس تعلیم و تربیت و سایر افراد درگیر در تحقیق اهمیت یکسانی دارد.

این فصل مربوط است به (۱) نقش تحقیقات پیشین در یک پروژه تحقیقی، (۲) منابع مرجع در تعلیم و تربیت و (۳) سازماندهی تحقیقات پیشین برای ارائه در گزارش.

### /// نقش تحقیقات پیشین در یک پروژه تحقیقی

قبل از شروع بررسی باید جستجو برای تحقیقات پیشین به طور عملی تکمیل شده باشد. این مرحله به اهداف متعدد مهمی خدمت می‌کند.

۱- آگاهی از تحقیقات پیشین محقق را قادر می‌سازد که حدود موضوع تحقیق خود را تعیین کند. اگر از قیاس تمثیلی استفاده کنیم، یک کاشف ممکن است بگوید: «ما می‌دانیم در دو هزار مایلی غرب و در پشت آن گیاهان یک رشته کوه قرار گرفته است، ولی نمی‌دانیم که در پشت این رشته کوه چه موقعیت و وضعیتی وجود دارد. من پیشنهاد می‌کنم که از منطقه گیاهان گذر کنیم. به بالای کوهها برویم و از آنجا به طرف غرب حرکت کنیم.» به همین نحو، یک محقق می‌گوید: «محققان الف، ب و ج سؤال مرا تا این حد پاسخ داده‌اند، محقق د نیز با تحقیقات خود، این قدر به دانش ما افزوده است. من پیشنهاد می‌کنم برای اینکه از تحقیقات د فواید بیشتری بگیریم، به این شیوه عمل کنیم.»

۲- جستجوی کامل در نظریه‌ها و تحقیقات مربوطه، محقق را قادر می‌سازد که سؤال خود را در یک چشم انداز قرار دهد. محقق باید تعیین کند که آیا تلاشهایش به شیوه‌ای معنادار احتمالاً بر دانش موجود خواهد افزود؟ دانش موجود در هر یک از زمینه‌های مورد نظر مبتنی بر انباشت نتایج بررسیهای نسلهای پیشین محققان و نیز نظریه‌های طراحی شده برای تلفیق این دانش و توضیح پدیده مورد مشاهده است. یک محقق ممکن است جستجو در تحقیقات پیشین را با هدف یافتن پیوستگی بین بررسی خود و مجموعه دانش موجود در حوزه مورد نظر آغاز کند. بررسیهایی که فاقد هر گونه پیوستگی با دانش موجود هستند، به ندرت به زمینه خاص خود

کمک معنی داری می‌کنند. این‌گونه بررسیها به اطلاعاتی مهجور یا بهره‌وری محدود تبدیل می‌شوند.

۳- جستجو در تحقیقات پیشین به محقق کمک می‌کند تا سؤال خود را محدود کند و مفاهیم بررسی را تعریف و روشن سازد. یک سؤال تحقیقی ممکن است برای انجام بررسی، بسیار گسترده باشد و یا برای آنکه بتوان به طور عینی کار را با آن شروع کرد، بسیار مبهم باشد. جستجوی دقیق در تحقیقات پیشین می‌تواند به محققان در جرح و تعدیل سؤال اولیه به نحوی که قابل تحقیق باشد، کمک کند. این کار همچنین در روشن ساختن مفاهیم موجود در بررسی و تبدیل این مفاهیم به تعاریف کاربردی مفید است. بسیاری از سازه‌های آموزشی و رفتاری از قبیل تنش، خلاقیت، ناکامی، پرخاشگری، پیشرفت، انگیزش و سازگاری، نیاز به روشن شدن و تعریف شدن به شکل کاربردی دارند. این سازه‌ها به موازات بسیاری از مفاهیم آموزشی و رفتاری دیگر، تا زمانی که نتوان آنها را کمی کرد، برای تحقیقات علمی مناسب نیستند. محقق، در جستجوی منابع پیشین، با تلاشهای گذشته برای روشن ساختن این مفاهیم و تعریف کاربردی آنها، آشنا می‌شود. جستجوهای موفق اغلب در تشکیل فرضیات مربوط به روابط بین متغیرهای موجود در بررسی نتیجه‌بخش است. بررسیهایی که در آنها فرضیاتی مورد آزمون قرار می‌گیرند، اغلب از آنهایی که فاقد فرضیه هستند، مفیدترند.

۴- یک جستجوی موشکافانه در تحقیقات پیشین اغلب به شناخت دلایل مربوط به نتایج متناقض در یک حیطه خاص منجر می‌شود. نتایج متضاد برای عموم شناخته شده نیستند. دلایل این تناقضات را می‌توان در انواع رویکردهای انتخاب شده برای حل مسئله یا انواع وسایل به کار گرفته شده، روشهای مورد استفاده؛ و یا تجزیه و تحلیل‌های انجام شده، یافت. حل چنین تضادهایی به بحث و جدل می‌انجامد، اما آنها می‌توانند به دانش موجود در حوزه مورد نظر به طور چشمگیری کمک کنند.

۵- پژوهشگراز طریق جستجوی تحقیقات پیشین در می‌یابد که کارآیی کدام روشها به اثبات رسیده است و کدامیک اطمینان بخش به نظر نمی‌آیند. محقق با غور و تفحص در تحقیقات پیشین و تعالی بخشیدن به دانش خود ممکن است بزودی راهی را بیابد که از آن طریق بتواند مطالعات خود را به پیش برد. البته پس‌نگری<sup>(۱)</sup> همیشه بهتر از پیش‌نگری<sup>(۲)</sup> است؛ از این رو شاید این امر که مطالعات اولیه در حیطه‌ای خاص اغلب خام و نامؤثر به نظر می‌رسند،

اجتناب ناپذیر باشد. بسیاری از پروژه‌های تحقیقاتی به دلیل استفاده از شیوه‌ها، ابزار، طرح بررسی و یا تجزیه و تحلیل‌های آماری نامناسب، رد می‌شوند. واری‌های کامل روش‌های به کار گرفته شده در بررسی‌های پیشین اغلب هم در یافتن دلیل شکست آنها و هم در ایجاد بینش برای انتخاب روش مناسب برای بررسی مورد نظر محقق ثمر بخش است. هم موفقیت و هم عدم موفقیت کارهای گذشته متضمن بینشی برای محقق در طراحی بررسی‌اش خواهد بود. با بنیان نهادن بررسی‌های آتی بر تحقیقات پیشین، می‌توانیم پیشرفت فزاینده‌ای در دانش خود در زمینه تحقیقات آموزشی انتظار داشته باشیم.

۶- یک جستجوی کامل در تحقیقات مرتبط با موضوع، از تکرار سهوی بررسی‌های گذشته جلوگیری می‌کند. بارها اتفاق افتاده است که یک محقق ایده‌ی پژوهشی را ارائه و سپس کشف کرده است که این ایده در تحقیقی مشابه قبلاً به کار گرفته شده است. در چنین مواردی او باید تصمیم بگیرد که آیا باید بررسی قبلی را به طور عمد تکرار کند و یا برنامه‌اش را تغییر دهد و جنبه‌ی دیگری از مسئله را بررسی کند.

۷- مطالعه پیشینه تحقیق، محقق را برای تفسیر اهمیت نتایج خود در موقعیت بهتری قرار می‌دهد. آشنایی با نظریه مرتبط با موضوع و تحقیقات گذشته، محقق را برای هماهنگ‌سازی یافته‌های تحقیقاتش با پیکره دانش موجود در همان حیطه، آماده می‌سازد.

### /// منابع مرجع در تعلیم و تربیت

برای هر دانشمند یا محقق آگاهی از نحوه دستیابی به کارهای پیشین در زمینه مورد نظر، امری بسیار ضروری است. برای انجام این مهم، محقق باید با (۱) با منابع کارهای پیشین، (۲) سازمانی که این اطلاعات را جمع‌آوری و آنها را در بانکهای اطلاعاتی سازماندهی می‌کند، (۳) راه دستیابی به این بانکهای اطلاعاتی و (۴) راههای مؤثر برای یافتن اطلاعات مورد نیاز، آشنا باشد. برای استفاده از این منابع، فرد باید از تسهیلات و خدمات کتابخانه‌ای قابل دسترس مطلع باشد. بسیاری از کتابخانه‌ها کتابچه راهنمایی چاپ کرده‌اند که خدمات و مقررات آنها را شرح می‌دهند یا به منظور آشنا ساختن علاقه‌مندان، بازدیدهایی ترتیب می‌دهند.

این مسئله بویژه برای آشنایی با نحوه سازماندهی کارت کاتالوگها بسیار مهم است. بسیاری از کتابخانه‌ها از طریق سیستم کامپیوتر کتابخانه، از یک فهرست کامپیوتری استفاده می‌کنند. اگر

سیستم کتابخانه‌ای شما کامپیوتری شده باشد، شاید بخواهید با چگونگی استفاده مؤثر از این سیستم آشنا شوید. در این بین شما باید دریابید که آیا کتابخانه می‌تواند کتابها و سایر مدارک نظیر رساله‌ها و مقالات نشریات علمی را از طریق سرویس امانتدهی بین کتابخانه‌ای، از سایر موسسات به امانت بگیرد یا خیر.

### // راهنماهای اصلی

برای آغاز کردن بررسیهای تحقیقی، رجوع به راهنماهای اصلی تحقیقات پیشین مفید است. یکی از این راهنماهای مفید راهنمای کتابشناسی در تحقیقات آموزشی بری<sup>(۱)</sup> است. این راهنما به «عنوان یک راهنمای فشرده، برای کمک به دانشجویان در رشته‌های تعلیم و تربیت و برای استفاده مؤثر آنها از منابع کتابخانه دانشکده یا دانشگاهشان، در نظر گرفته شده است.» این یک فهرست مشروح بالغ بر ۷۰۰ منبع تحقیقی اصلی است که بر اساس انواع مواد، اعم از نشریات علمی، بررسیهای تحقیقی، نشریات دولتی، متون مرجع و آزمونها، مرتب شده است. متون مرجع شامل سالنامه، کتاب راهنما، منابع زندگینامه‌ای و کتابچه‌های راهنما در روش شناسی تحقیقات آموزشی، در سبک و سیاقی مناسب برای نوشتن مقالات تحقیقی است.

منابع کتابشناسی روز در تعلیم و تربیت<sup>(۲)</sup> که به سه زبان (انگلیسی، فرانسه و اسپانیایی) منتشر می‌شود، عمدتاً بر اساس مجموعه مراجع دفتر بین‌المللی اسناد آموزشی یونسکو در پاریس پایه‌ریزی شده است. این نشریه شامل کتابشناسی‌ها، فهرستهای نشریات علمی و نامنامه تحقیقات به ترتیب کشور مربوطه، یا فهرستهای مجزا برای سازمانهای نشر بین‌المللی و منطقه‌ای است.

راهنمای منابع اطلاعات آموزش وودبری<sup>(۳)</sup> راهنمای جامع برای یافتن اطلاعات در زمینه تعلیم و تربیت است. این راهنما، منابع چاپی، غیر چاپی و سازمانی را انتخاب و توصیف می‌کند و در بسیاری از موارد آنها را برای تحقیق در تعلیم و تربیت مورد ارزشیابی قرار می‌دهد. ترتیب آن اساساً بر مبنای نشریات، با فصولی در مورد تامین مالی و دولت، تعلیم و تربیت استثنایی، مواد آموزشی، آزمونها و وسایل سنجش است. این فهرست همچنین شامل توصیفی از مراحل یک تحقیق مؤثر و یک راهنما برای پژوهشگران است.

1- Berry, D. M. (1980). A Bibliographic Guide to Educational Research (2d ed.). Metuchen, NJ: Scarecrow Press

2- Unesco (1984). Current Bibliographical Sources in Education (2d ed.). Paris: Unesco, International Bureau of Education.

3- Woodbury, M (1982). A Guide to Sources of Educational Information (2d ed.). Arlington, VA: Information Resources

راهنمای کتابهای مرجع شی‌هی<sup>(۱)</sup> چند هزار منبع مرجع را در موضوعات مختلف، از جمله تعلیم و تربیت، به طور مختصر توصیف و ارزشیابی می‌کند. اطلاعات آن از طریق ضمایم دو سالانه به روز در می‌آیند. کتابهای مرجع سالانه آمریکا<sup>(۲)</sup> کتابهای مرجع منتشره سال قبل (شامل کتابهای تجدید چاپ شده) را در کلیه موضوعات در بر می‌گیرد و حواشی توصیفی و ارزیابانه‌ای را فراهم می‌آورد.

### // نشریات علمی در زمینه تعلیم و تربیت

نشریاتی که بررسیهای تحقیقاتی را در موضوعات خاص ادغام و خلاصه می‌کنند، محقق را قادر می‌سازند تا بررسی خود را با مرور تحقیقات پیشین، آغاز کند.

دائرةالمعارف تحقیقات آموزشی میتزل<sup>(۳)</sup> برای ارائه یک ترکیب و تحلیل انتقادی از گزارشهای تحقیقی در تعلیم و تربیت طراحی شده است و شامل مقالات مؤثر با بحثهای مستند به مدارک کتابخانه‌ای از جریانه‌ها و تحولات اخیر، به همراه موضوعات سنتی است. این دائرةالمعارف چهار جلدی شامل حدود ۱۶۴ موضوع است. این دائرةالمعارف منبع اصلی مناسب برای تحقیق در مورد موضوعات بنیادی در زمینه تعلیم و تربیت می‌باشد. دائرةالمعارف ارزشیابی آموزشی<sup>(۴)</sup> سعی بر ارائه مفاهیم و تکنیکهای اصلی برای ارزشیابی برنامه‌های تربیت معلم دارد و مقالات آن به ترتیب الفبا تنظیم شده است. دائرةالمعارف کارشناسان تعلیم و تربیت آمریکا<sup>(۵)</sup> شامل حدود ۲۰۰۰ مدخل کوتاه برای نامها و اصطلاحاتی است که اغلب در متون تخصصی تعلیم و تربیت به آنها برمی‌خوریم. دائرةالمعارف بین‌المللی تعلیم و تربیت: «پژوهشها و بررسیها»<sup>(۶)</sup>، یک بررسی کلی از جدیدترین دستاوردهای دانش‌پژوهی در حل مسائل تعلیم و تربیتی، تجربیات، و سازمانهای تحقیقی در سراسر جهان است. این دائرةالمعارف کارهای تحقیقی و کارشناسی در تعلیم و تربیت را به مفهوم وسیع دربرمی‌گیرد و آخرین پیشرفتها در شاخه‌های مختلف تعلیم و تربیت، قابلیت دستیابی به اطلاعات علمی

press

ادامه پاورنی صفحه قبل

- 1- Sheehy, E. P. (1986). Guide to Reference Books (10 Th ed.). Chicago: American Library Association.
- 2- Wynar, B. S. (ed.) (1970-). American Reference Books Annual. Littleton, co: Libraries Unlimited.
- 3-Mitzel, H. E. (ed.) (1982). Encyclopedia of Educational Research (5th ed., vols. 1-4). New York: Free Press.
- 4- Anderson, S. B., et al (1975). Encyclopedia of Educational. San Francisco: Jossey - Bass.
- 5- Dejnozka, E.L., and D.E. Kapel (1982). American Educators' Encyclopedia. Westport, CT: Greenwood Press.
- 6- The International Encyclopedia of Education: Research and Studies (1985). (10vols) Oxford: pergamon.

مطمئن و معتبر در این حیطة و انواع تحقیقات تکمیلی مورد نیاز را مورد بررسی قرار می‌دهد. دائرةالمعارف بین‌المللی آموزش عالی<sup>(۱)</sup> شامل ۲۸۲ مقاله در مورد موضوعات جدید در آموزش عالی است که توسط پیشتازان تعلیم و تربیت در سراسر دنیا نوشته شده است. یکی دیگر از منابع مفید، کتابچه راهنمای تحقیق در تدریس است.<sup>(۲)</sup> از این کتابچه راهنما در فاصله ۱۰ سال سه چاپ متفاوت منتشر شده است. این کتابچه تحقیقات موجود در عرصه تدریس را فهرست، خلاصه و تحلیل انتقادی می‌کند. هر چاپ شامل مقالات معتبر کارشناسان در موضوعات منتخب در این زمینه است. کتابشناسی‌های جامع تحت موضوعات منتخب در آن گنجانده می‌شود. در بین موضوعات چاپ سوم می‌توان از «اندازه‌گیری تدریس»، «روشهای کمی تحقیق در زمینه تدریس»، «روشهای کیفی تحقیق در زمینه تدریس»، «مشاهده به عنوان کندوکاو و روش»، «ترکیبات تحقیق در تدریس»، «تدریس راهبردهای یادگیری»، «رفتار معلمان و پیشرفت دانش آموزان»، «تحقیق در مورد تدریس هنرها و زیبایی‌شناسی» و بررسیهای تحقیقی بشمار دیگر در زمینه تدریس نام برد. دائرةالمعارف تعلیم و تربیت استثنایی<sup>(۳)</sup> اطلاعات پایه‌ای را برای بسیاری از شیوه‌ها و تخصصهای مربوط به تعلیم و تربیت کودکان استثنایی و ویژگیهای مخصوص به آنها، نیازها و مسائل آنها، فراهم می‌سازد. این دائرةالمعارف سه جلدی شامل متجاوز از ۲۰۰۰ موضوع است. هر چند ترتیب این کتاب الفبایی است، اما بسیاری از موضوعات را می‌توان از نظر محتوایی در هفت زمینه اصلی دسته‌بندی کرد: زندگینامه‌ها، آزمونهای آموزشی و روان‌شناختی، مداخله و ارائه خدمات، شرایط معلولیت، خدمات مربوطه، مسائل حقوقی و موضوعات متفرقه.

چهار چاپ از نشریه بررسی تعلیم و تربیت استثنایی<sup>(۴)</sup> به همراه نشریه پیشرفتهای تعلیم و تربیت استثنایی<sup>(۵)</sup> بررسیهای جامعی در مورد حیطةهای تعلیم و تربیت استثنایی ارائه می‌کند و به ارزشیابی‌های محدودتر حیطةهای خاص آن می‌پردازد. بررسی تحقیقات آموزشی<sup>(۶)</sup> یک

- 1- Knowles, A. S. (ed). The International Encyclopedia of Higher Education (1978). (10vols). San Francisco: Jossey, Bass.
- 2- Gage, N. L. (ed). (1983). Handbooks of Research on Teaching: Chicago: Rand McNally.
- 3- Travers, R. (ed) (1973). Second Handbook of Research on Teaching (2d ed.) Chicago: Rand McNally.
- 4- Reynolds, C.R. and L. Mann (eds) (1987). Encyclopedia of Special Education: A Reference for the Education of the Handicapped and Other Exceptional Children and Adults (3vols.). New York: John Wiley.
- 5- Mann, L. and D. Sabatino (eds.) (1973-1980). Review of Special Education (1st- 4th eds.) Philadelphia: JSE Press.
- 6- Keogh, B. K (1981). Advances in Special Education: A Research Annual, Greenwich, CT: JAI Press.
- 7- Review of Educational Research (1931-). Washington: American Educational Research Association.

فصلنامه از انجمن تحقیقات آموزش آمریکا (AERA) است که بررسیهای ترکیبی و تفسیر متون تحقیقات آموزشی را هم در مسائل ماهوی و هم در مسائل روش شناختی، منتشر می‌کند. از سال ۱۹۳۱ تا ۱۹۷۰ در هر شماره نشریه، به ترتیب، یکی از ۱۵ موضوع گسترده در زمینه تعلیم و تربیت مورد بررسی قرار گرفت. در ژوئن سال ۱۹۷۰ نشریه این سیاست را ترک و شروع به چاپ مقاله‌های ارائه شده کرد، از آن پس هر شماره آمیزه‌ای از موضوعات گوناگون است. در تلاشی برای تهیه چکیده‌هایی از بررسیها در زمینه موضوعات گسترده یک حیطه تخصصی، AERA مجموعه بازنگری تحقیق در تعلیم و تربیت<sup>(۱)</sup> را که از سال ۱۹۷۳ هر ساله چاپ می‌شود، تدارک دید. تلاش این مجموعه بر آن است که مشخص کند در حیطه تعلیم و تربیت چه تحقیقاتی انجام شده است، چه تحقیقاتی در دست اجرا است و چه تحقیقاتی لازم است که صورت گیرد. برای مثال، جلد سیزدهم بازنگری تحقیق در تعلیم و تربیت<sup>(۲)</sup> شامل مرور پیشرفتهای انجام شده در تحقیق و در عمل، در چهار زمینه اصلی: آموزش ابتدایی قرائت، پیشرفت مهارتهای نوشتن، نیروی تدریس و فرا تحلیل انجام آزمایشهای گوناگون در تعلیم و تربیت است.

بررسیهای روان شناختی نقش مهمی در تحقیقات آموزشی ایفا می‌کند. سالنامه بررسیهای روان شناختی، که از سال ۱۹۵۰ هر ساله چاپ می‌شود، متخصصان حیطه‌های مختلف روان شناسی را به کار می‌گیرد تا در مورد متون تحقیقاتی گزارش تهیه و آن تحقیقات را ارزیابی کنند، و روندها و پیشرفتهای جدید را در همه جنبه‌های علم روان‌شناسی شناسایی و حوزه‌های فراموش شده را مشخص کنند. کتاب راهنمای روان‌شناسی عمومی<sup>(۳)</sup> یک منبع مطالعاتی مفید است که به حوزه‌های وسیعی از تاریخ، نظریه، روش‌شناسی، ارگانیزم انسان، ادراک، یادگیری، زبان، تفکر و هوش، انگیزش و هیجان، شخصیت و سایر حوزه‌های خاص می‌پردازد. این نشریه شامل منابع کتابشناختی از بررسیهای تحقیقی موجود است.

// نمایه نشریات دوره‌ای، مجلات خلاصه تحقیقات و فهرست نقل نامها  
 امروزه با پی‌ریزی شالوده‌گسترده‌ای از تحقیق، نظریه و عقاید عمومی در یک موضوع خاص،

1- Review of Research In Education (1973 -). Washington: American Educational Research Association.

2- Annual Review of Psychology (1950-). Palo Alto: Annual Reviews.

3- Wolman, B. B. (1973). Handbook of General Psychology, Engle Wood- Cliffs, NJ: Prentice - Hall.



محقق می‌تواند مدارک و مطالب اضافی را که در منابع اصلی بررسی نقل نشده‌اند، پیدا کند. نمایه فصلنامه‌ها، مجلات خلاصه تحقیقات و فهرست نقل نامها فرد را قادر می‌سازد تا این اطلاعات را بیابد. این نشریات، که در شماره‌های متوالی و در مقاطع زمانی منظم چاپ می‌شوند، در یافتن اطلاعاتی که به طور وسیع در مجلات و دیگر منابع پراکنده هستند، به عنوان راهنما عمل می‌کنند. ناشران این نمایه‌ها از کارشناسانی که منابع چاپ شده و چاپ نشده را بررسی و دسته‌بندی می‌کنند، استفاده به عمل می‌آورند. این مقالات سپس بر حسب موضوع، و گاهی با حاشیه‌نویسی، گروه بندی می‌شوند؛ به گونه‌ای که محققان از طریق فهرست نسبتاً جامعی از آثار موجود در حوزه تخصصی‌شان، که به‌طور منطقی به روز درآمده‌اند، مطلع می‌شوند.

#### / نمایه‌های اریک

مرکز اطلاعاتی منابع آموزشی (ERIC) به منظور گردآوری، نگهداری و انتشار اطلاعات در زمینهٔ تعلیم و تربیت که توسط دفتر تعلیم و تربیت آمریکا (USOE) تأسیس شده است، دو نمایه مهم تهیه می‌کند.

پیش از تأسیس سیستم اریک، گزارشهایی که توسط طرفهای قرار داد و اهداکنندگان مقالات به USOE ارائه می‌شد، به صورت ابتدایی و پراکنده توزیع و همانند گزارشهای منابع دیگر، ناپدید می‌شد. هدف اریک تصحیح این شرایط آشفته، به منظور گردآوری و حفظ اسناد و مدارک غیر چاپی «ناپایدار» و مورد نظر دست اندرکاران تعلیم و تربیت و قرار دادن این بانک اطلاعاتی در اختیار مردم بود.

اکنون سیستم اریک، که بودجهٔ آن را انجمن ملی تعلیم و تربیت (NIE) تأمین می‌کند، اطلاعات را گردآوری، ارزشیابی و فهرست می‌کند تا از طریق شبکه‌ای مشتمل بر ۱۶ دفتر مرکزی واقع در دانشگاههای متعدد و سازمانهای حرفه‌ای در سراسر کشور و با بهره‌جویی از پردازش اطلاعات مرکزی، وارد نمایه‌های اریک شوند. هریک از دفاتر مرکزی، مسئول یک زمینه خاص در تعلیم و تربیت است. یک کتابچهٔ راهنما، با عنوان *راهنمای اطلاعاتی خدمات اریک*<sup>(۱)</sup> این دفاتر مرکزی و سایر سازمانهایی را که راه دستیابی به بانک اطلاعاتی اریک و منابع

1- Brandhorst, T., and J. Eustace (eds.) (1986). *Directory of ERIC Information Service Providers*. Washington: U.S.

مربوط به آن را برای مراجعان و مشتریان خود فراهم می‌سازند، فهرست می‌کند. سه نشریه علمی از مهمترین نشریاتی که توسط آریک تهیه می‌شود، نمایه روز مجلات تعلیم و تربیت<sup>(۱)</sup> (CIE) و منابع تعلیم و تربیت<sup>(۲)</sup> (RIE)، و منابع آموزشی کودکان استثنایی<sup>(۳)</sup> (ECER)، هستند. برای دستیابی به اسناد و مدارک آریک از یک سیستم فهرستبندی پیشرفته و دقیق استفاده می‌شود. این سیستم، در گنجینه توصیفی آریک<sup>(۴)</sup> که در بیشتر کتابخانه‌ها موجود است، تعریف شده است.

### / ماهنامه نمایه روز مجلات تعلیم و تربیت

ماهنامه نمایه روز مجلات تعلیم و تربیت (CIE) از محصول کار کارشناسان مراکز اطلاعاتی آریک ترکیب یافته است. مقالات موجود در بالغ بر ۷۰۰ مجله طبق سیستم گنجینه آریک دسته‌بندی و فهرست می‌شوند. متجاوز از ۳۵۰۰۰۰ مقاله در CIE فهرست و هر ساله در حدود ۱۸۰۰۰ مقاله به آنها اضافه می‌شود.

CIE به چهار بخش تقسیم می‌شود: (۱) فهرست موضوعی، (۲) فهرست مؤلف، (۳) بخش مدخل اصلی، و (۴) فهرست محتوایی مجلات. فرد می‌تواند عنوان و شماره‌های مقالات مورد علاقه خود را با اولین نگاه به فهرست موضوعی بیابد و سپس از این شماره‌ها برای پیدا کردن مدخلها در بخش مدخل اصلی استفاده کند. اگر نتوانید موضوع مورد علاقه خود را در فهرست موضوعی بیابید، در فهرست موضوعات مشابه به جستجو پردازید. گنجینه لغات آریک برای تدارک یک سیستم فهرست بندی نظام‌مندتر تعمداً محدود شده است. فهرست نویسندگان متفرقه برای آن دسته از افراد مفید است که مشتاق یافتن کارهای محققان خاصی هستند.

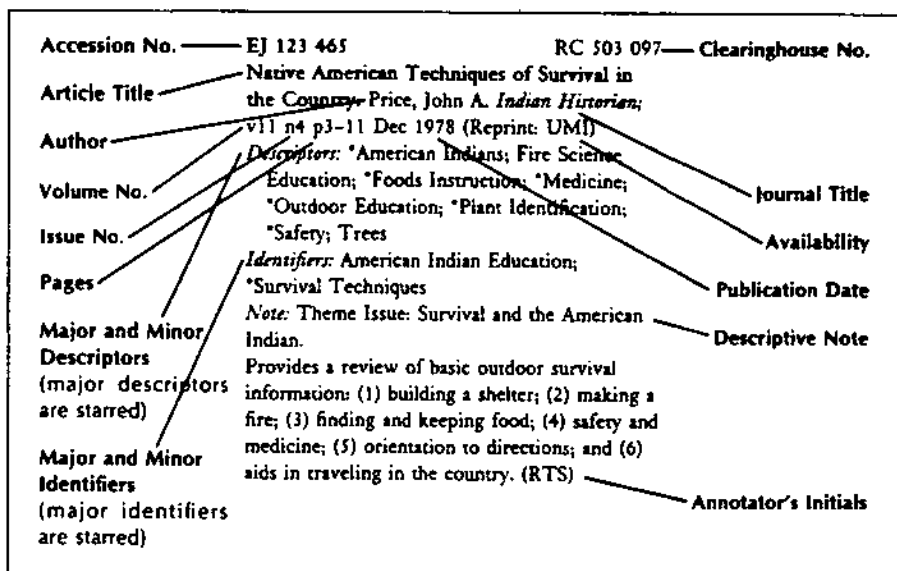
فهرست محتوایی مجلات شامل محتوای کلی مجلات و محتوای هر مقاله است.

شکل ۱ - ۳ نمونه‌ای از نحوه معرفی یک مقاله در بخش مدخل اصلی، به همراه تشریح قسمتهای مختلف آن به همان نحوی است که به وسیله CIE برای روشن شدن نوع اطلاعات در فهرست محتوایی مجلات شامل محتوای کلی مجلات و محتوای هر مقاله است.

Government Printing Office.

- 1- Current Index to Journals in Education (1989). Phoenix: Oryx Press (1979).
- 2- Resources in Education (1988). Washington: U.S. Government Printing Office.
- 3- Exceptional Child Education Researches (1969). Reston, VA: Council For Exceptional Children.
- 4- Houston, J.E. (ed.) (1987). Thesaurus of ERIC Descriptors (11th ed.). Phoenix: Oryx Press

شکل ۱-۳ نمونه‌ای از یک مدخل *CIEE*



شکل ۱-۳ نمونه‌ای از نحوه معرفی یک مقاله در بخش مدخل اصلی، به همراه تشریح قسمتهای مختلف آن به همان نحوی است که به وسیله *CIEE* برای روشن شدن نوع اطلاعات در هر دفتر، تدارک دیده شده است.

حاشیه نویسی بویژه بسیار مفید است؛ چون توضیح کوتاهی در مورد هر مقاله ارائه می‌کند. توجه کنید که در نمونه بالا عنوان آن تنها ایده مختصری از محتوای هر مقاله به دست می‌دهد. قبل از ژانویه ۱۹۷۰، زمانی که *CIEE* حاشیه نویسی را ضمیمه کار خود کرد، محققان در مراجعه به مقالاتی که عناوین گمراه کننده و نارسا داشتند و در بررسی بیشتر ثابت شد که با موضوع بی ارتباط هستند، با ناکامی بزرگی مواجه شدند. اگر عنوان یک مقاله بتواند گرایش اصلی آن را به طور مشخص منتقل کند، خلاصه کنندگان، حاشیه‌ای به آن ضمیمه نمی‌کنند. انتشار *CIEE* از سال ۱۹۶۹ آغاز شد. این مجله به طور ششماهه و سالانه منتشر می‌شود.

شیوه درست در استفاده از *CIEE* شامل شش مرحله است:

۱- کلمات کلیدی را که ممکن است مقالات مربوط به بررسی شما در زیر آنها فهرست شده

باشند، معین کنید. این کلمات کلیدی نوعاً شامل جامعه مورد بررسی و متغیرهایی است که شما باید در تبیین مسئله تحقیقی خود آنها را تعریف کنید.

۲- برای دریافتن آنکه کدامیک از کلمات کلیدی شما به عنوان توصیفگر مورد استفاده قرار خواهند گرفت، گنجینه توصیفی اریک را بازبینی کنید. شما احتمالاً برای کلمات کلیدی که فهرست کرده‌اید، نیاز به یافتن معادلها دارید.

۳- بررسی خود را از آخرین شماره‌های *CIE* شروع کنید و سپس به کارهای گذشته در نشریات پیشین پردازید.

۴- کلیه منابع داده شده برای هر عنوان را که احتمالاً می‌تواند مفید باشد، از *CIE* رونوشت بردارید. این روش کار، یافتن مقالات اصلی را آسان می‌کند.

۵- مقالات را در مجلات مربوط به آنها جستجو کنید. اگر کتابخانه شما فاقد آن مجله بود، ممکن است بتوانید نسخه زیراکس شده آن را از کتابخانه دیگری به عاریت بگیرید.

۶- اگر یک چکیده به همراه خود مقاله ضمیمه شده بود، اول آن را بخوانید تا دریابید که آیا خواندن کامل مقاله ضروری است یا خیر. در صورت فقدان چکیده از بخش خلاصه و نتیجه‌گیری مقاله آغاز کنید.

### / منابع آموزشی کودکان استثنایی

این فصلنامه (سابقاً کتابچه تحقیقاتی مربوط به تعلیم و تربیت کودکان استثنایی) شامل چکیده تحقیقاتی ذخیره شده در بایگانی کامپیوتری مرکز اطلاعات شورای کودکان استثنایی به عنوان بخشی از برنامه‌ریزی است. چکیده‌های مربوط به تعلیم و تربیت معلولین و کودکان جوانان سرآمد نیز فهرست شده‌اند. ترتیب آنها شبیه *CIE* و *RIE* است. تا سال ۱۹۷۵، هر مجلد شامل فهرست تراکمی سالهای قبل بوده و از ۱۹۷۶، فهرستهای تراکمی مؤلف، عنوان و موضوع در نسخه نهایی سال گنجانده شده است. کتابها با مقالات موجود و مجلات، اسناد دولتی و غیره از موارد مشمول *ECER* هستند.

### / منابع تعلیم و تربیت

چکیده‌گزارشهای تحقیقی از منابعی به غیر از مجله‌ها نیز فهرست و به صورت ماهنامه به

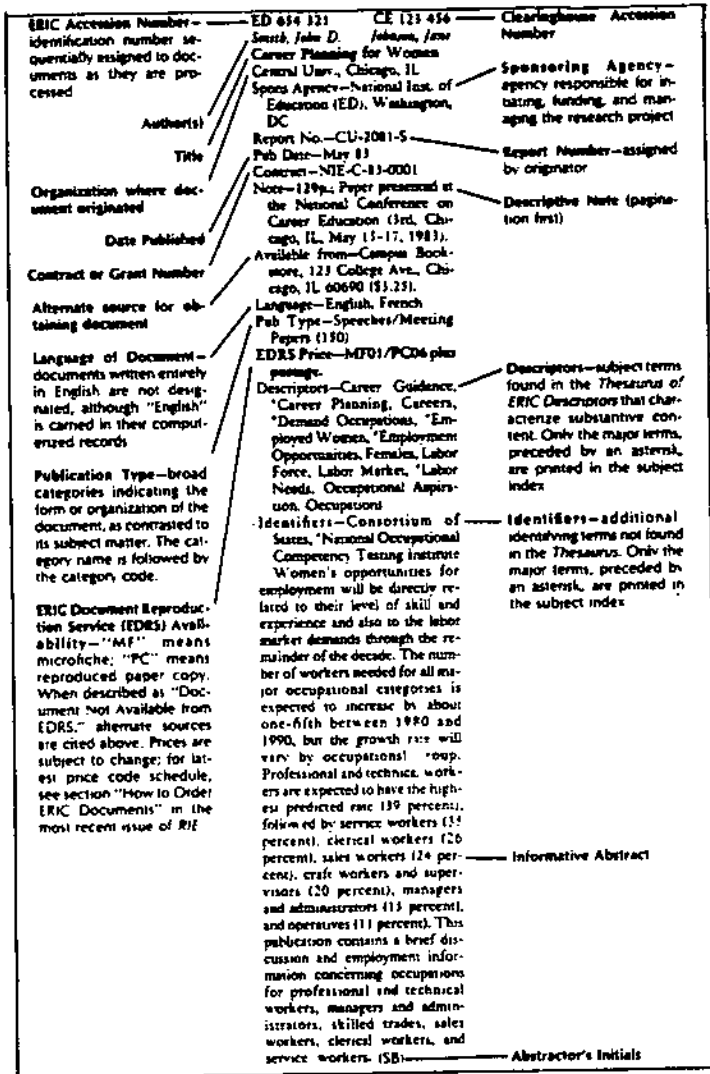
وسیله اریک در نشریه منابع تعلیم و تربیت (RIE) منتشر می‌شود. در این زمینه یک مجموعه شش ماهه از بخش فهرستها نیز به صورت تراکمی منتشر می‌شود. هر مرکز اطلاعاتی اسناد و مدارک مربوط به حیطه تخصصی خود را جمع‌آوری و سپس آنها را وارد فهرست می‌کند و فهرستها و چکیده‌ها را تهیه می‌کند. اسناد و مدارک مادر (اصلی)، به همراه چکیده‌های تهیه شده توسط مراکز اطلاعاتی، بر طبق یک روال منظم به بخش پردازش اطلاعات، جایی که به یک بخش از بانک اطلاعات مرکزی تبدیل می‌شوند، ارسال می‌شوند، به طریقی که مراجعه کنندگان سیستم اریک قادر به استفاده از آن باشند. بخش پردازش مسئول جمع‌آوری داده‌های همه مراکز اطلاعاتی و چاپ مجموعه به صورت ماهانه در RIE است. منابعی که مراکز اطلاعاتی تخصصی از آنها فهرست تهیه می‌کنند، شامل گزارشهای تحقیقات آموزشی با هزینه دولت فدرال، چکیده‌ها، جزوه‌ها، کتابچه‌های راهنما، مقالات برگزیده از پیشرفتهای انجمنها و سازمانهای علمی، کتابشناسی‌ها، نمونه برگزیده متون درسی دانشگاهی، راهنمای معلمان و طرح کلی برنامه درسی، به همراه تعداد بسیار زیادی پیشنهاد تحقیق و گزارش پروژه‌ها است. با چنین پوشش گسترده‌ای، روشن است که اسناد و مدارک مجموعه اریک مورد توجه شدید دست‌اندرکاران و همچنین محققان در زمینه تعلیم و تربیت باشد. در حال حاضر متجاوز از ۳۵۰۰۰۰ هزار سند و مدرک در این مرکز وجود دارد، و هر سال ۱۸۰۰۰ سند و مدرک به این مجموعه در حال گسترش اضافه می‌شود.

به منظور تعیین و سفارش نسخه‌ای از اسناد و مدارک مادر، هر چکیده با شماره دسترسی خود مشخص می‌شود. در حال حاضر سه فهرست در این زمینه منتشر می‌شود: فهرست نویسنده، مؤسسه و موضوع. محقق با استفاده از فهرست مناسب، می‌تواند عناوین اسناد و مدارک و شماره‌های دسترسی به آنها را پیدا و سپس از آنها برای یافتن چکیده‌های خاص استفاده کند. روال استفاده از RIE شبیه به آن چیزی است که در CIJE به کار می‌رود. محقق نوعاً بررسی خود را با جدیدترین نسخه شروع می‌کند و برای رجوع به کارهای گذشته به جستجوی موضوع یا موضوعات مورد علاقه می‌پردازد.

### / مجموعه اسناد اریک

نسخه میکروفیش اکثر اسناد ما در (اصلی) به محض درخواست در اختیار متقاضی قرار

شکل ۲-۳ نمونه‌ای از یک مدخل RIE



می‌گیرد، یا می‌توان یک قرارداد دائمی با ERIC منعقد کرد تا تمام اسناد به مجرد آماده شدن در اختیار طرف قرارداد قرار گیرد.

بسیاری از کتابخانه‌ها چنین قرار دادی با اریک دارند، بنابراین کلیه مجموعه اسناد اریک با شرایطی سهل در اختیار مراجعان آنها قرار می‌گیرند.

بیشتر اسناد به دو شکل در دسترس متقاضی قرار می‌گیرند: نسخه چاپی و میکروفیش. در بسیاری از کتابخانه‌ها سیستم میکروفیش به دلیل ارزان‌تر بودن و اشغال جای کمتر ترجیح داده می‌شوند. مجموعه کامل میکروفیش اریک، حجمی برابر یک قفسه بایگانی را اشغال می‌کند. دستگاهی که میکروفیش خوان نامیده می‌شود، برای بزرگ کردن صفحه میکروفیش بر روی پرده ثابت است و همچنین برای متقاضی امکان حرکت از تصویر یک صفحه به صفحه‌ای دیگر را فراهم می‌سازد. بیشتر کتابخانه‌ها و دانشگاه‌ها از چنین تسهیلاتی برخوردارند. قیمت هر یک از اسناد اریک در هر یک از فرمهای چاپی و میکروفیش در *RIE* فهرست می‌شود.

تمام پروژه‌هایی که به هزینه USOE انجام می‌شوند، به طور خودکار به سیستم اریک منضم می‌شوند و نسخه‌های گزارشهای کامل شده آنها از طریق اریک در دسترس مراجعان قرار می‌گیرند. اگر متن منتشر شده برای بار اول و با امتیاز حق مؤلف چاپ شده باشد، به عنوان مرجع از آن نام برده می‌شود؛ هر چند نسخه‌های آن از طریق اریک در دسترس همگان قرار نمی‌گیرد. روشهای مشابهی در مورد کتابها و سایر مواد چاپی که برای فروش آماده می‌شوند، نیز به کار می‌رود. اریک در عرصه تعلیم و تربیت ارزشمند است؛ چون مواد مورد نظر برای محققان در زمینه تعلیم و تربیت گستره وسیعی از منابع را گردآوری و خلاصه می‌کند. دستیابی به بسیاری از این منابع در گذشته تقریباً غیرممکن بوده است.

مراکز اطلاعاتی اریک در زمینه موضوعات منتخب، کتابشناسی‌هایی با حواشی فراهم می‌آورد. قبل از آنکه برای بررسی خود کتابشناسی تهیه کنید، بهتر است از طریق مکاتبه با مرکز اطلاعاتی اریک که مربوط به زمینه تحقیق شماست، تماس برقرار کنید تا چنانچه کارکنان آنجا کتابشناسی مفیدی در دست دارند، آن را در اختیار شما بگذارند.

توضیح کاملتر در مورد سیستم اریک و چگونگی استفاده از آن در کتابچه اریک تحت عنوان «این مرکز چه می‌تواند برای شما انجام دهد، چگونه از آن استفاده کنیم»<sup>(۱)</sup>، موجود است که در بسیاری از کتابخانه‌ها در دسترس متقاضیان قرار می‌گیرد.

1- Brown, J. W. M. K. Setts, and J. Yarborough (1977). ERIC: What It can do for you, How to Use It. Stanford, CA: ERIC Clearing House on Information Sources.

مراکز اطلاعاتی شبکه اریک امیدوار هستند که این مجموعه را تا حد امکان جامع تر کنند و از متخصصان تعلیم و تربیت درخواست می کنند تا مواد مربوط را به این مراکز ارائه کنند. اگر گزارشی از یک سخنرانی و یا مقاله ای تهیه کرده اید و علاقه دارید از طریق اریک در کشور منتشر شود، نسخه ای از آن را به بخش پردازش و منابع اریک به آدرس 4833 Rugby Avenue, Bethesda, MD 20814 یا مرکز اطلاعاتی مسئول برای گردآوری اسناد در حوزه موضوع مربوطه، ارسال دارید.

### دستیابی به سیستم اریک از طریق کامپیوتر

یک نسل قبل، جستجوی کامل ۵۰۰ مجله و ۱۰۰۰۰۰ سند و مدرک دیگر برای یافتن مطالب مربوط به مسئله مورد نظر، یک کار عظیم بود. امروزه دسترسی کامپیوتری به سیستم اریک امکان انجام چنین جستجویی را با استفاده از کامپیوتر شخصی یا ارتباط تلفنی، فراهم آورده است.

محتوای *CIEE* و *RIE* بر روی نوارهای کامپیوتری، موجود و در دسترس است و با استفاده از این نوارها می توان به سیستم اریک دست یافت. برای مثال، برای شناسایی مقالات و اسناد *RIE* که برای تدریس زبان فرانسه با استفاده از مطالب برنامه ریزی شده به کار می روند، کامپیوتر به نحوی برنامه ریزی شده است که با کاربرد توصیفگر آموزش، و نیز توصیفگر زبان فرانسه، تمام اسناد و مقالات مربوط به این موضوع را شناسایی می کند. برخی از برنامه های کامپیوتری تنها شماره های دسترسی به اسناد اریک *ED* و مجلات اریک *EL* را چاپ کامپیوتری می کنند، اما بیشتر برنامه ها مدخل کامل *CIEE* و مدخل کامل *RIE* هر مقاله و سند را که دارای هر دو توصیفگر هستند، چاپ می کنند.

متجاوز از ۹۰۰ مؤسسه نوارهای اریک را در اختیار دارند و می توانند به جستجوهای کامپیوتری بپردازند. بیشتر ادارات آموزش و پرورش ایالتی این نوارها را در اختیار دارند و برای دست اندرکاران تعلیم و تربیت در ایالت خود به جستجوهای کامپیوتری می پردازند. بیشتر این ادارات همچنین از متخصصان بازیابی اطلاعات استفاده می کنند که اطلاعات بازیابی کامپیوتری مورد نیاز مشتریان را از طریق برگه تقاضا برای گرفتن کمک از سیستم اریک، در اختیار آنها قرار می دهند. بسیاری از دانشگاهها خدمات مشابهی برای دانشجویان خود و دیگر متقاضیان فراهم ساخته اند.



برخی از کتابخانه‌ها نیز سیستم اریک و دیگر بانکهای اطلاعاتی را بر روی CD-ROM<sup>(۱)</sup> یعنی دیسک فشرده با حافظه غیر قابل دستکاری منتقل کرده‌اند و در اختیار دارند، که یک روش جدید برای دسترسی به اریک و سایر بانکهای اطلاعاتی است. مؤسسه سیلور پلاتر<sup>(۲)</sup> از سال ۱۹۶۶ تاکنون بانک اطلاعات اریک را به طور کامل بر روی سه CD-ROM منتشر کرده است. هر ثبت اطلاعات شامل شرح کتابشناختی، چکیده‌ها، و توصیفگران می‌شود. بر خلاف فهرستهای چاپی اریک، این دیسکهای CD-ROM را می‌توان در یک کیف دستی کوچک حمل کرد. برای استفاده از CD-ROM به یک کامپیوتر شخصی با حافظه ۵۱۲ کیلوبایت و یک دیسک‌گردان نیاز است. به غیر از بانک اطلاعات کامپیوتری اریک، حدود ۵۰۰ بانک اطلاعاتی مربوط به بسیاری از موضوعات، شامل چکیده‌های روان‌شناختی، چکیده‌های رساله‌های بین‌المللی، چکیده‌های جامعه‌شناختی، و فهرست نقل نامهای علوم اجتماعی نیز وجود دارد. بانکهای اطلاعاتی جدید تقریباً هر روز از طریق نرم افزار Vendors جهت جستجوهای کامپیوتری در حال تاسیس هستند. نمایه‌تعلیم و تربیت<sup>(۳)</sup>، نشریه یکی از بانکهای اطلاعاتی بیشمار است که به سیستم کامپیوتری ویلسون لاین وابسته است. این خدمات نیز در بسیاری از کتابخانه‌ها قابل دسترس است.

جستجو در سیستم اریک یک مرحله مهم در بررسی پیشینه تحقیقاتی است، ولی نمی‌توان فرض کرد که وقتی این مرحله خاتمه یافت، جستجوی پیشینه تحقیقاتی نیز به پایان رسیده است. مطالب مرتبط با سؤال محقق ممکن است وارد سیستم‌اریک نشده باشند.

### / نمایه سایر نشریات علمی

بسیاری از نشریات دیگر نیز برای دست یافتن به اطلاعات روز در زمینه تحقیقات نظریه‌ها و عقاید در تعلیم و تربیت سودمند هستند. یکی از مراجع استاندارد برای کارهای تحقیقاتی در زمینه تعلیم و تربیت، نمایه‌تعلیم و تربیت است که به طور منظم از سال ۱۹۲۹ توسط مؤسسه اچ.و. ویلسون در نیویورک منتشر می‌شود. این نمایه مقالات حدود ۳۵۰ نشریه علمی، سالنامه بولتن، صورتجلسه و تک‌نگاری را فهرست می‌کند. نمایه‌تعلیم و تربیت بهترین منبع برای

1- Compact Disc - Read - Only Memory.

2- Silver Platter Information, Inc., 37 Walnut st., Wellesley Hills, MA 02181.

3- Education Index (1929-). New York: H. W. Wilson

دستیابی به مقالات مندرج در مجلات، قبل از تأسیس نمایه روز مجلات تعلیم و تربیت در سال ۱۹۶۹، و نیز بهترین منبع برای دستیابی به مقالات جدید است؛ زیرا این نشریه مقاله‌ها را شش ماه قبل از *CIE* فهرست می‌کند.

علاوه بر این نمایه‌های عمومی تعدادی نمایه تخصصی مفید نیز در دسترس محققان قرار می‌گیرد. چکیده‌های روان‌شناختی،<sup>(۱)</sup> متون روان‌شناسی و مطالب مربوط به آن را در سراسر جهان فهرست می‌کند. این فهرست حاوی کتابها، رساله‌های دکترا و مقالات علمی به همراه خلاصه‌ای از هر یک است که خواننده را قادر می‌سازد تا ارتباط مطالب را با یکدیگر دریابد. چکیده‌ها و کتابشناسی رشد کودک<sup>(۲)</sup> مطالب موجود در حیطه‌های نوبیاوگی، طب بالینی، بهداشت عمومی، رشد و روان‌شناسی تطبیقی، روان‌شناسی آزمایشی (شامل پدیده‌های یادگیری، شخصیت و روان‌شناسی تعلیم و تربیت و مشاوره) را برحسب نام مؤلف و موضوع فهرست می‌کند. چکیده‌های جامعه‌شناسی تعلیم و تربیت<sup>(۳)</sup>، که یک نشریه بین‌المللی است، مقالات مجلات علمی، کتابها، و رساله‌ها را در حوزه خود، بر اساس نام مؤلف و موضوع، به همراه خلاصه‌ای از هر کدام فهرست می‌کند. چکیده‌های مدیریت آموزش<sup>(۴)</sup> با فهرست‌بندی مقالات نشریه‌های تخصصی در زمینه مدیریت آموزشی، راه دستیابی به مطالب آنها را فراهم می‌آورد. چکیده‌های آموزش عالی<sup>(۵)</sup> (سابقاً چکیده‌های دانشجویان - کارکنان کالج) ترکیبی از چکیده‌های مجلات، خلاصه مذكرات کنفرانسها و گزارشهای تحقیقی مربوط به دانشجویان و خدمات دانشجویی است. موضوعات شامل مشاوره، مسکن، کمکهای مالی، و آزمون و اندازه‌گیری است. نمایه علوم اجتماعی<sup>(۶)</sup> مقالات مجلات در زمینه علوم اجتماعی شامل مجلات آموزشی و نیز مجلاتی در زمینه روان‌شناسی، جامعه‌شناسی، علوم سیاسی، مردم‌شناسی، حقوق و اقتصاد را برحسب نام مؤلف و موضوع فهرست می‌کند.

هر رشته، مانند آموزش بازرگانی، مهارتهای صنعتی، پزشکی، و غیره، نمایه تخصصی خود را دارد. مراجعه به راهنماهای اصلی متون تحقیقاتی پیشین نام نمایه‌های تخصصی در زمینه‌های دیگر را نیز به خوانندگان ارائه می‌کند.

1- Psychological Abstracts (1927-). Lancaster, PA: American Psychological Association.

2- Child Development Abstracts and Bibliography (1927-). Lafayette, IN: Purdue University, Society For Research in Child Development.

3- Sociology of Education Abstracts (1965-). Abbingdon, England: CarFax Publishing Co.

4- Educational Administration Abstracts (1966-). Beverly Hills: Sage Publications.

5- Higher Education Abstracts.

6- Social Sciences Index (1974-). New York: H. W. Wilson.

## / فهرست رساله‌ها و پایان نامه‌ها

رساله‌های فوق لیسانس و دوره دکترا از منابع اطلاعاتی مفید برای محققان است. علاوه بر نمایه‌های مجلات علمی و مجلات خلاصه تحقیق، نظیر RIE اریک و چکیده تحقیقات روان شناختی، که برخی از رساله‌ها و پایان نامه‌ها را فهرست می‌کنند، راهنماها و نمایه‌های تخصصی متعدد نیز بسیار سودمند هستند.

نمایه جامع رساله‌ها<sup>(۱)</sup>

این نمایه تلاش دارد که کلیه رساله‌های پذیرفته شده در دانشگاه‌های آمریکا را طی سالهای ۱۹۶۱-۷۲ تحت پوشش قرار دهد. این نمایه شامل ۳۷ جلد در ۱۷ طبقه بندی اصلی موضوعی به همراه شرح کامل کتابشناختی برحسب مؤلف و موضوع است. یک ضمیمه سالانه برای روزآمد کردن این نمایه چاپ می‌شود. چکیده رساله‌های بین‌المللی (DAI)<sup>(۲)</sup> که هر ماه منتشر می‌شود، شامل چکیده رساله‌های دوره دکتراست که با همکاری دانشگاه‌ها به مرکز میکرو فیلم‌های بین‌المللی دانشگاهی تحویل می‌شود (حدود ۳۷۰ عدد در سال ۱۹۸۴). DAI به سه بخش تقسیم می‌شود:

علوم انسانی و علوم اجتماعی، علوم پایه و مهندسی و دانشگاه‌های اروپا. جستجوهای کامپیوتری در DAI از طریق بسیاری از بانکهای اطلاعاتی اصلی و ندرز از جمله، دیالوگ<sup>(۳)</sup>، BRS<sup>(۴)</sup> و کامپوسرو<sup>(۵)</sup> در دسترس است. بیشتر کتابخانه‌ها برای همکاری با محققان، جستجوهای کامپیوتری در بانکهای اطلاعاتی را تدارک می‌بینند.

یک نشریه دیگر مشابه با DAI، رساله‌های دوره دکترای آمریکاست<sup>(۶)</sup>. این نشریه رساله‌های پذیرفته شده دوره دکترا در دانشگاه‌های ایالات متحده و کانادا طی هر سال تحصیلی را به همراه رساله‌هایی که به شکل میکروفیلم در میکروفیلم‌های دانشگاهی موجود است، در یک فهرست مشترک گردآوری می‌کند. این نشریه حاوی تعدادی از رساله‌هاست که در DAI وجود

1- Comprehensive Dissertation Index (1851- 1972). (37vols.) Ann Arbor: University MicroFilms.

2- Dissertation Abstracts International (1938-). Ann Arbor: University MicroFilms.

3- Dialog Information Service, Inc., 3460 Hillview Avenue, Palo Alto, CA 84304.

4- BRS Information Technologies, 1200 Route 7, Latham, NY 12110.

5- Compu Serve Information Service, Inc, 5000 Arlington Center Boulevard, Columbus, OH43220.

6- American Doctoral Dissertations (1957-) Ann Arbor: University MicroFilms.

ندارد. رساله‌های دوره دکتری آمریکا برحسب طبقه‌بندی موضوعی و مؤسسه مرتب شده و شامل یک فهرست است، ولی هیچ چکیده‌ای تهیه نمی‌کند.

راهنمای فهرست پایان نامه‌های دوره فوق لیسانس مؤسسه بِلک<sup>(۱)</sup> به دو بخش اصلی تقسیم می‌شود: فهرست پایان نامه‌ها در زمینه‌های خاص، با حاشیه‌نویسی‌ها، و فهرست پایان نامه‌های مؤسسات خاص. چکیده پایان نامه‌های فوق لیسانس: چکیده پایان نامه‌های فوق لیسانس منتخب بر روی میکروفیلم<sup>(۲)</sup> شامل چکیده یک فهرست منتخب از پایان‌نامه‌های دوره فوق لیسانس از دانشگاهها و دانشکده‌های ایالات متحده آمریکا است که به صورت میکروفیلم در دسترس است.

پایان نامه‌های فوق لیسانس در تعلیم و تربیت مؤسسه سیلوی<sup>(۳)</sup>، یک فهرست سالانه از پایان‌نامه‌ها در زمینه تعلیم و تربیت از دانشکده‌ها و دانشگاههای ایالات متحده آمریکاست. برای هر پایان نامه فهرست شده، اطلاعات کتابشناسی داده می‌شود، اما چکیده‌ای ضمیمه آنها نیست.

#### فهرست نقل نامهای علوم اجتماعی (SSCI)

مقاله‌ای که با مراجعه به یکی از این نمایه‌ها یافته شده است، خود، مراجعی در زمینه کارهای قبلی در آن حیطه، ارائه می‌کند. زمانی که فرد مقالات مهم را در زمینه‌ای خاص می‌یابد، یک نمایه نقل نامها می‌تواند در سرعت بخشیدن بر کار او مؤثر باشد. برای مثال اگر شما یک مقاله فوق العاده مفید را که از سال ۱۹۷۶ منتشر شده است، خوانده باشید، از طریق نمایه‌های بعدی می‌توانید مقالاتی را بیابید که نام مقاله مورد نظر شما در کتابشناسی آن وجود دارد. بسیاری از این مقاله‌ها، تداوم، گسترش یا تغییراتی را که در بررسی مورد نظر اولیه رخ داده است، توصیف خواهند کرد.

فهرست نقل نامهای علوم اجتماعی (SSCI)<sup>(۴)</sup>، که هر ساله در سه جلد توسط مؤسسه اطلاعات علمی منتشر می‌شود، می‌تواند به شما بگوید که طی سال از کدام نویسندگان در همه

1- Black, D. (ed.) (1965). Guide to Lists of Master's Theses. Chicago: American Library Association.

2- Masters Abstracts: Abstracts of Selected Masters theses on Microfilm 1962- 1995; Masters Abstracts International (1986). Ann Arbor: University MicroFilms.

3- Slvey, H. M. (1951-). Master's Theses in Education. Cedar Falls, IA: Research Publications.

4- Social Sciences Citation Index (1973-) Philadelphia: Institute For Scientific Information.

زمینه‌های علوم اجتماعی، از جمله تعلیم و تربیت، چه چیزی ذکر شده و در زمینه‌های مختلف چه نوشته شده است، همچنین کتابشناسی اطلاعات مورد لزوم درباره مؤلفانی که از آنها از آنها ذکر شده یا از کسی ذکر کرده‌اند نیز ضمیمه است. این اطلاعات از طریق چهار نمایه قابل دستیابی است:

۱- نمایه مرجع یک فهرست الفبایی از نام همه مؤلفانی را که طی سال کتابی از آنها منتشر شده است، ارائه می‌کند، که حاوی اطلاعات کتابشناسی کاملی از جمله فهرست الفبایی منابع و مأخذی است که در کتابشناسی مولف مرجع آمده است.

۲- فهرست نقل نامها، نام نویسندگان نقل شده در نمایه مرجع را اخذ می‌کند و آنها را به ترتیب الفبایی به شیوه‌ای ارائه می‌کند که محقق دریابد یک مقاله خاص در کجا ذکر شده است. مراجع کتابشناختی برای هر مورد نقل شده ارائه می‌شود. یک نمایه از ناشر مؤلف در زمینه نقل نامها و یک نمایه از مقاله‌های بی‌نام مؤلف نیز تهیه می‌شود. این نمایه فرد را قادر می‌سازد تا ایده‌هایش را با سرعت دنبال کند.

۳- فهرست موضوعی اصطلاحات همه واژه‌های مهم را با واژه‌های مهم دیگر زیر هر عنوان زوج می‌کند. اریک عنوان هر واژه را در حکم واژه اولیه فهرست می‌کند و آن را با سایر واژه‌های معادل تلفیق می‌کند. یک فهرست الفبایی از نامهای مؤلفان، که عناوین کتابهای آنان حاوی آن واژه‌ها است، برای هر زوج واژه و معادل آن ارائه می‌کند. بنابراین اطلاعات کتابشناسی را می‌توان برای هر مولف در نمایه مرجع یافت.

۴- نمایه نشانی ناشر یک فهرست الفبایی از سازمانهای انتشاراتی است که مؤلفان طی آن سال به آن وابسته‌اند. زیر هر مدخل، شرکت نشر، فهرست مؤلفان با اطلاعات کتابشناسی کامل قرار دارد.

## // انتشارات دولتی

دولت فدرال یک منبع اصلی اطلاعات تعلیم و تربیتی است که بیشتر از هر یک از سازمانهای ایالات متحده، متعهد به انجام تحقیقات، پیمایشها، و گردآوری آمار است. اداره آموزش و پرورش به موازات مرکز ملی آمار در زمینه آموزش و پرورش، تعداد بیشمار نشریه، شامل گزارشهای تحقیقی، پیمایشها، فعالیتهای اجرایی و برنامه‌های توصیفی به چاپ می‌رساند.

برای یافتن نشریات ویژه دولت آمریکا، کاتالوگ ماهانه نشریات دولت آمریکا<sup>(۱)</sup> فهرستی اصلی برای مراجعه است. این نشریه شامل یک بخش اصلی است که مؤسسه انتشاراتی، اسناد منتشره و عنوان نویسنده، موضوع، و فهرست کلمات کلیدی عناوین را فهرست می‌کند. تا سال ۱۹۷۵ یک مجموعه سالانه برحسب عنوان و موضوع در نسخه چاپ دسامبر گنجانده شده بود، از سال ۱۹۷۶ این مجموعه‌ها به صورت شش ماهه منتشر می‌شود. همچنین یک نمایه تراکمی ۵ سالانه برای جستجوی سریعتر در تحقیقات سالهای گذشته وجود دارد.

نشریات وابسته به ادارات ایالتی آموزش و پرورش و دیگر مؤسسات ایالتی را می‌توان از طریق نمایه ماهانه نشریات ایالتی<sup>(۲)</sup> پیدا کرد.

### // منابع آزمونها

در انجام هر تحقیقی یک آزمون یا وسایل اندازه‌گیری مورد نیاز است. سالنامه اندازه‌گیری روانی بوروس<sup>(۳)</sup>، عمده‌ترین مأخذی است که آزمونها را فهرست می‌کند و به طور انتقادی به بررسی آنها می‌پردازد. این کتابها بخصوص برای کمک به متقاضیان در عرصه تعلیم و تربیت، روان‌شناسی و ترغیب کردن آنها به استفاده هر چه هوشمندانه‌تر از آزمونهای استاندارد شده، طراحی شده است. هر سالنامه، با این هدف با آرایش یکسان طراحی شده است که مجلات قبلی را تکمیل کند، تا اینکه جایگزین آنها شود. آزمونها بر اساس موضوع گروه‌بندی می‌شوند و توضیح در مورد هر آزمون با جستجوها و بررسیهای مهم و منابع بررسیهایی که آزمون در آن به کار رفته است، دنبال می‌شود. هر مجلد مروری بر بررسیها، قطعات منتخب و مراجع کتابشناسی مجلدات قبلی دارد. مجلدات شامل آزمونهای استعداد و پیشرفت تحصیلی در حیطه‌های موضوعی متعدد، آزمونهای شخصیت و حرفه‌ای و آزمونهای هوش و اطلاعات کامل در مورد قیمت و نحوه استفاده از هر کدام از آزمونهای هوش است. اطلاعات کامل در مورد قیمت و نحوه استفاده از هر کدام از آزمونها نیز موجود است. آزمونهای چاپ شده شماره<sup>(۴)</sup> به مثابه نمایه و

1- U.S. Superintendent of Documents (1895). Monthly cata of U.S. Government publications. Washington: U.S. Government Printing Office.

2- U.S. Library of Congress, Exchange and Gift Division (1910). Monthly Checklist of state publication. Washington: U.S. Government Printing Office.

3- Buros, O.K. (ed.) (1938-) Mental Measurement Yearbooks. Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

4- Mitchell, J.V., Jr. (ed.) (1983) Tests in Print III: An Index to Tests, Test Reviews, and the Literature on Specific Tests. Lincoln: Buros.

ضمیمه‌ای برای مجلدات گذشته سالنامه اندازه‌گیریهای روانی به کار می‌رود. مؤسسه بوروس همچنین مطالب موجود در سالنامه‌های اندازه‌گیری روانی را برای تکن‌نگاریهای ویژه در آزمونهای شخصیت، خواندن، هوش، مهارت شغلی و حرفه‌ای، زبانهای انگلیسی و خارجی، ریاضیات، علوم و بررسیهای اجتماعی سازماندهی می‌کند.

همچنین موسسه بوروس یک مجموعه از داده‌های کامپیوتری را از طریق مرکز تکنولوژیهای اطلاعاتی BRS ارائه می‌کند که اطلاعات آن به طور ماهانه به روز می‌شود.

نمای تمام و کمال هر سند به متقاضی امکان می‌دهد که متن کامل هر سند را پیدا کند و نسخه‌ای از آن را دریافت و یا آن را به طور کامل بررسی کند و بخشهای منتخب آن را از چاپگر کامپیوتر به دست آورد.

این سیستم در بین سایر سالنامه‌ها یک سیستم بازیابی اطلاعات عالی برای به روز کردن اطلاعات است.

یک مرجع استاندارد برای آزمونها، مرجع جامع ارزشیابی در روان‌شناسی تعلیم و تربیت و تجارت<sup>(۱)</sup> است که حاوی متجاوز از ۳۰۰۰ آزمون موجود به زبان انگلیسی است. هر آزمون یک دسته‌بندی مقدماتی به دست می‌دهد و در یکی از بخشها به طور مشروح تعریف و احتمالاً یک منبع رجوع در یک دسته‌بندی ثانوی محسوب می‌شود.

از سال ۱۹۸۴، نقد آزمونها<sup>(۲)</sup> تعریفها و نقدهای کلی از وسایل اندازه‌گیری را فراهم ساخته است. یک فهرست مفید از مجموعه موضوعات آزمونها، بر اساس نوع متغیرهای آنها نیز ضمیمه است. راهنماهای ویژه آزمونها شامل اندازه‌گیری نگرشهای روان‌شناسی اجتماعی<sup>(۳)</sup>، آزمونهای خواندن و مرور تحقیقات پیشین<sup>(۴)</sup>، راهنمایی بر ۶۵ آزمون تشخیص در تعلیم و تربیت استثنایی<sup>(۵)</sup>، راهنمای بین‌المللی برای آزمونهای فضایی<sup>(۶)</sup>، سنجش فراگیران ناتوان:

1- Sweetland, R.C., and D.J. Keyser (eds.) (1983). *Tests: A Comprehensive Reference for Assessments in Psychology, Education and Business*. Kansas City, Mo: Test Corporation of American.

2- Keyser, D.J., and R.C. Sweetland (eds.) (1984). *Test Critiques* vol. 1-. Kansas City, Mo: Test Corporation of America.

3- Robinson, J.P., and P.R. Shaver (1973). *Measures of Social Psychological Attitudes*. Ann Arbor: Survey Research Center, Institute for Social Research.

4- Buros, O.K. (ed.) (1988- 1975). *Reading Tests and Reviews* (2vols.). Lincoln: Buros Institute of Mental Measurement, University of Nebraska.

5- Compton, C. (1990). *A Guide to 65 Diagnostic Tests for Special Education*. Belmont, CA: Fearon Education.

6- Elliot, J., and I. MacFarlane (1983). *An International Directory of Spatial Tests*. New York: Humanities Press.

وسایل برگزیده<sup>(۱)</sup>، آزمونهای به صورت میکروفیش در آمده: فهرستهای حاشیه‌نویسی<sup>(۲)</sup> و کاتالوگ مجموعه آزمونهای مرکز خدمات آموزشی<sup>(۳)</sup> است.

اطلاعات در مورد آزمونها را می‌توان در نمایه‌های نشریات علمی، از جمله نمایه روز مجلات تعلیم و تربیت، نمایه تعلیم و تربیت، منابع تعلیم و تربیت کودکان استثنایی، و چکیده‌های روان‌شناختی، که تحت نام خاص خود در فهرست موضوعی ثبت شده‌اند، یافت. شما همچنین می‌توانید زیر عناوین موضوعی مانند «آزمونها» و «مروری بر آزمونها» به جستجو بپردازید.

### // جستجوی کامپیوتری

اکثر کتابخانه‌های دانشگاهی تسهیلات جستجوی کامپیوتری را تحت عناوین جستجوی کامپیوتری، جستجوی پایگاه اطلاعاتی، سرویس رجوع کامپیوتری، و از این قبیل تدارک دیده‌اند. در بیشتر کتابخانه‌ها این خدمات از طریق پست و در برخی اوقات و به طور محدود با مراجعه بدون وقت قبلی صورت می‌گیرد. بعضی از کتابخانه‌ها ممکن است به دلیل تقاضا برای جستجوی کامپیوتری اطلاعات ویژه (تخصصی) از فرد بخواهند که در آنجا حضور یابد. این امر نه تنها به فرد فرصت می‌دهد تا نمونه‌ای از نتایج اطلاعات بازیابی شده را ملاحظه کند، بلکه این بخت را به او می‌دهد که حافظه کامپیوتر را بیشتر بکاود و راهبردهای تحقیقی خود را بهبود بخشد.

بسیاری از منابع نقل قول شده در این فصل از جمله نمایه‌های تعلیم و تربیت، نمایه‌های اریک، منابع تعلیم و تربیت کودکان استثنایی، چکیده‌های روان‌شناختی، نمایه نقل نامهای علوم اجتماعی، چکیده‌های رساله‌های بین‌المللی، و سالنامه‌های اندازه‌گیری‌های روانی، در پایگاههای اطلاعاتی کامپیوتری گنجانده شده‌اند. بعلاوه، مجموعه داده‌های دیگری مانند سندرم نارسایی اکتسابی در سیستم ایمنی بدن (ایدز) تحقیقات علوم پزشکی، منابع در تعلیم و تربیت حرفه‌ای (RIVE)، مجموع آزمون مرکز خدمات آموزشی (ETSF) و چکیده‌های جامعه‌شناختی (SOCA)، نیز وجود دارند. برای تهیه فهرستهای مشروح از بیشتر پایگاههای

1- Mauser, A.J. (1977) *Assessing the Learning Disabled: Selected Instruments* (2d ed.). Navato, CA: Academic Therapy Publications.

2- *Tests in Microfiche: Annotated Index* (1975- ). Princeton: Educational Testing Service.

3- *The Educational Testing Service Test Collection Catalog* (1986- ). Phoenix: Gryx Press.



اطلاعاتی، موجود از دو کاتالوگ پایگاه اطلاعاتی منتشره توسط سرویس خدماتی دیالوگ و تکنولوژیهای اطلاعاتی BRS مشترکاً استفاده می‌شود.<sup>(۱)</sup> بسیاری از مؤسسات، کتابداران متخصص را که در جستجوی کامپیوتری مهارت دارند و می‌توانند در خدمات مختلف به شما کمک کنند، در استخدام دارند.

از فواید جستار کامپیوتری می‌توان به این موارد اشاره کرد: (۱) صرفه جویی در وقت: به دلیل وجود بسیاری از اصطلاحات مترادف و همانند که برای مفهوم یا مفاهیم مورد نظر مفید است، کامپیوتر در چند دقیقه می‌تواند اطلاعات را بازیابی و چاپ و از هدر رفتن ساعات و یا روزهای محقق در استخراج دستی جلوگیری کند؛ (۲) در جریان روز بودن: پایگاههای اطلاعاتی کامپیوتری مجموعاً بیشتر از فهرستهای چاپی در جریان اطلاعات روز هستند. داده‌های بیشتر پایگاههای اطلاعاتی هر هفته و یا دو هفته یک بار به روز در می‌آیند؛ (۳) پیوند: چون هر موضوع تحقیقی به ترکیبی از دو و یا چند موضوع مربوط است، کامپیوتر می‌تواند همزمان بیش از یک عنوان را جستجو کند؛ (۴) دقت: زمانی که یک موضوع را نمی‌توان به سهولت از طریق فهرستهای چاپی شناسایی کرد، کامپیوتر می‌تواند از طریق فراهم ساختن امکان دستیابی به مواد مختلف از قبیل: عناوین، کلمات کلیدی، رمزهای موضوع، یا سر تیرهای موضوعی ضمیمه که فهرستهای متداول قادر به ارائه آنها نیستند، این مهم را به انجام رساند. جستجوی کامپیوتری همچنین می‌تواند نتایج شما را بر اساس گزینه‌های معین، از جمله تاریخ و نوع نشریه، نام نویسنده و یا زبان محدود کند (۵) منابع منحصر به فرد: برخی از پایگاههای اطلاعاتی داده‌هایی را فراهم می‌کنند که در منابع چاپی موجود نیست و (۶) دسته بندی: جستجوهای کامپیوتری اطلاعات بازیابی شده را بر طبق تاریخ انتشار (درخواست حضوری یا غیرحضوری) یا نام نویسنده و عنوان با ترتیب حروف الفبایی دسته‌بندی می‌کند.

برای بیشتر محققان جستجوی کامپیوتری تنها روش برای انتخاب است. در یک جستجوی «دستی» فرد باید برای یک عنوان ویژه، تعداد بیشماری از فهرستها را بررسی کند، در بین فهرستها به دنبال آنها بگردد و تعداد ناچیزی داده‌های مربوط بیابد که اصطلاحات منتخب را با دیگر زمینه‌های مورد علاقه ترکیب کرده است. هر چند با استفاده از مفاهیم منطقی به نام عمل کننده‌های بول<sup>(۲)</sup> (از سیستم منطقی‌ای که توسط جرج بول، ریاضیدان انگلیسی، در قرن نوزدهم

1- 1988 Dialog Database Catalog (1988). Palo Alto: Dialog Information Services, Inc.; 1988 BRS Database Catalog (1988).

Latham, NY: BRS Information Technologies.

2- Boole.

تحول یافت) می‌توان همزمان برای بسیاری از موضوعات به جستجوی کامپیوتری پرداخت و آنها را با یکدیگر ترکیب کرد.

شکل ۳-۳ راههایی که مفاهیم و یا اصطلاحات مختلف را می‌توان با استفاده از عمل‌کننده‌های منطقی بول به یکدیگر پیوند داد، نشان می‌دهد.

طبق این شکل،  $A \text{ AND } B$  برای مواردی به جستجو می‌پردازند که هر دو اصطلاح کلیدی را دارا هستند. برای مثال، در مارس ۱۹۸۹ سیستم اریک ۱۰۴۷ مورد با اصطلاح کلیدی دانش‌آموزان بزرگسال و ۲۵۰۲ مورد با اصطلاح کلیدی آموزش ریاضیات در اختیار داشت. با استفاده از برنامه سیلور پلاتر چهار مورد با دو اصطلاح کلیدی معین شد.

در این شکل،  $A \text{ OR } B$  اصطلاحاتی را جستجو می‌کند که یک یا دو (و نه منحصرأً یکی) اصطلاح را دارا باشند. در مثال مشابهی درخواست برای آموزش ریاضیات یا پیشرفت ریاضیات به مناسبتی دست خواهد یافت که یک یا هر دو اصطلاح کلیدی را دارا باشد.

یک جستجوی معمولی از هر دوی این راهبردها استفاده می‌کند. یک مربی بالغ که به آموزش ریاضیات علاقه‌مند است، از عبارت  $\text{OR}$ ، شامل اصطلاحات کلیدی متعددی که احتمالاً دامنه آموزش بزرگسالان را پوشش می‌دهد و یا عبارت  $\text{OR}$ ، شامل عبارتهای متعددی که دامنه آموزش ریاضیات را در می‌گیرد، به همراه یک عبارت از نوع  $\text{AND}$  برای پیوند هر دو زمینه (عوامل) استفاده می‌کند.

عوامل آموزش ریاضیات

پیشرفت ریاضیات

OR

آموزش ریاضیات

OR

مهارتهای ریاضیات

عوامل آموزش بزرگسالان

آموزش ابتدایی بزرگسالان

OR

آموزش بزرگسالان

یادگیری بزرگسالان

OR

برنامه‌های بزرگسالان

OR

دانش آموزان بزرگسال

OR

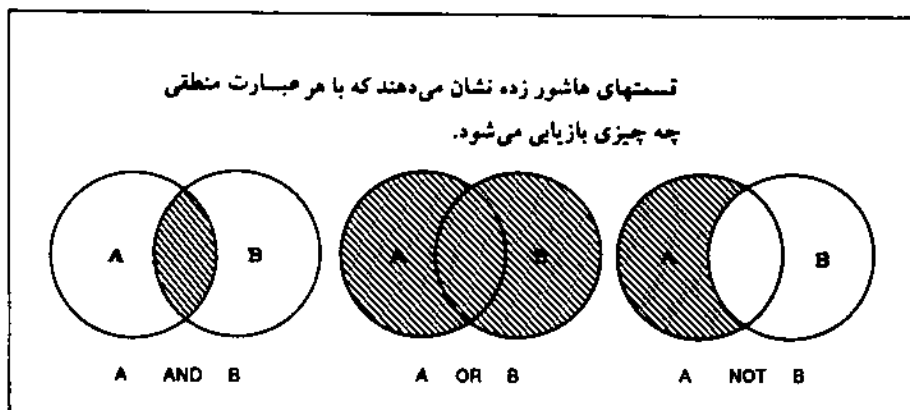
ادامه تحصیل

OR

دانش آموزان غیر سنتی

با استفاده از این عوامل، برنامه سیلور پلاتر، نشان داد که ۱۵۳۱۶ مورد از عامل اول و ۳۰۰۹ مورد از عامل دوم یک یا چند اصطلاح کلیدی داشتند. فقط ۹۴ مورد از هر دو عامل در نهایت

### شکل ۳-۳ منطق بول



یک اصطلاح کلیدی داشتند که از آن میان ۷۳ مورد از اسناد اریک و ۲۱ مورد مجلات تحت پوشش CIE بودند. (همه این اطلاعات از طریق کامپیوتر و در کمتر از چهار دقیقه به دست آمد!) در این مرحله، فرد می‌تواند تعداد موارد را با گزینش راهبردی که عوامل دیگری را ارائه می‌کند، محدود کند و موارد را به آنهایی منحصر کند که انگلیسی نیستند یا جستجو را به یک تاریخ خاص محدود سازد. همچنین فرد می‌تواند اجزای متعددی از اطلاعات را درباره هر مورد که بر صفحه کامپیوتر ظاهر می‌شود، بر اساس این تصمیم که چه چیزی به کارش مربوط می‌شود، انتخاب کند. در پایان فرد می‌تواند برای موارد باقیمانده، از این اطلاعات نسخه چاپی کامپیوتری تهیه کند. شکل ۳-۴ یک نسخه چاپ شده کامپیوتری از ۲۱ مقاله مجلات است که از طریق جستجوی کامپیوتری تعیین شده، و شامل شماره سند، نویسنده، عنوان، مقاله (حاوی عنوان، مجله، شماره بخشها، شماره، و تاریخ)، توصیفگرها و چکیده است.

مثال پیشین رایجترین موارد استفاده از منطق بول را نشان می‌دهد. عمل‌کننده‌های دیگر بعضی اوقات به کار می‌روند. برای نمونه، NOT می‌تواند برای موارد «آموزش بزرگسالان که دارای توصیفگر «آموزش ابتدایی بزرگسال» نیز نیستند، به جستجو پردازند.

## // واژه‌نامه‌ها

تقریباً می‌توان گفت که هر رشته علمی و ازگان تخصصی خود را داراست. عملکرد یک واژه‌نامه موضوعی، تعریف مختصر کلمات، چه اصطلاحات و چه اسامی است که زبان فنی یک رشته علمی خاص را به وجود می‌آورند. واژه‌نامه‌های متعددی در زمینه تعلیم و تربیت وجود دارند. واژه‌نامه تعلیم و تربیت گود<sup>(۱)</sup>، یک واژه‌نامه دانشگاهی حاوی کلمات و اصطلاحاتی است که در زمینه تعلیم و تربیت دارای معانی ویژه هستند. نام حدوداً ۲۰۰ هماهنگ کننده، دستیار و مرورگر در پایان چاپ سوم این واژه‌نامه، که حاوی بیش از ۳۳۰۰۰ مدخل است، فهرست شده است. اصطلاحات تعلیم و تربیتی مورد استفاده در کانادا، انگلستان و ولز در بخشهای جداگانه در انتهای این واژه‌نامه، ارائه شده است. واژه‌نامه بین‌المللی تعلیم و تربیت<sup>(۲)</sup> حاوی بالغ بر ۱۰۰۰۰ مدخل، شامل تعاریف و اصطلاحات، سازمانهای بین‌المللی، مؤسسات و سازمانهای اصلی ملی، کارشناسان تعلیم و تربیت، و از این قبیل است.

واژه‌نامه تعلیم و تربیت<sup>(۳)</sup>، که توسط هیل ویراستاری شده است، مقالات کوتاهی در ۱۵ زمینه تعلیم و تربیت ارائه می‌کند که براساس یک سیستم الفبایی از اصطلاحات (عمدتاً اصطلاحات انگلیسی) فهرست می‌شوند. جلد اول دائرةالمعارف بین‌المللی آموزش عالی<sup>(۴)</sup> نیز تعاریف کوتاهی از کلمات اختصاری و یک واژه‌نامه از اصطلاحات مهیا می‌سازد. در زمینه تعلیم و تربیت کودکان استثنایی، واژه‌نامه ناتوانیها در خواندن و یادگیری بوش<sup>(۵)</sup> و منابع راهنمای کارشناسان تعلیم و تربیت در آموزش کودکان استثنایی: اصطلاحات - قوانین - آزمونها و سازمانها تألیف دیویس<sup>(۶)</sup>، و واژه‌نامه اصطلاحات تعلیم و تربیت کودکان استثنایی<sup>(۷)</sup> سه واژه‌نامه عمده و کاربردی هستند.

1- Good, c.v. (1973). Dictionary of Education (3d ed.). New York: McGraw - Hill.

2- Page, G.P., J.B. Thomas, and A.R. Marshall (1977). International Dictionary of Education. New York: Nicholas Publishing.

3- Hill, P.J. (ed.) (1982). A Dictionary of Education. Boston: Kegan Paul.

4- Knowles, A.S. (ED.) (1977). The International Encyclopedia of Higher Education (10 vols.). San Francisco: Jossey- Bass.

5- Bush, G.L., and R.C. Andrews (1978). Dictionary of Reading and Learning Disabilities. Los Angeles: Western Psychological Services.

6- Davis, W.E. (1980). Educator's Resource Guide to Special Education: Terms - Laws- Tests- Organizations. Boston: Allyn & Bacon.

7- Moore, B.C., W. Abraham, and C.R. Laing (1980). A Dictionary of Special Education Terms. Springfield, IL: Thomas.

شکل ۴-۳ نمونه‌ای از یک برگه جستجوی کامپیوتری ERIC

AN: EJ293655  
AU: Dolan,-Lawrence  
TI: Affective Characteristics of the Adult Learner: A Study of Mastery-Based Instruction.  
JN: Community/Junior-College-Quarterly-of-Research-and-Practice;  
v7 n4 p367-78 Jul-Sep 1983  
DE: Affective-Behavior; Community-Colleges; Remedial-Mathematics; Two-Year-Colleges  
DE: \*Adult-Learning; \*Mastery-Learning; \*Mathematics-Achievement; \*Mathematics-Instruction; \*Reentry-Students; \*Student Characteristics  
AB: Presents an investigation of the benefits of mastery learning strategies for adults with mathematics skill deficiencies. Examines which student types benefit most from mastery learning, centering on affective characteristics. Demonstrates that students with low self-concept, control, and instructional mastery skills had higher achievement in mastery than nonmastery classes. (DMM)

// قرضه بین کتابخانه‌ای

خدمات قرضه بین کتابخانه‌ای نیز می‌تواند فوق‌العاده مؤثر باشد. اگر شما منابعی برای بررسی تحقیقی در مورد نشریه‌های علمی، کتابها، یا رساله‌هایی بیابید که در کتابخانه همیشه شما موجود نباشد، از طریق خدمات قرضه بین کتابخانه‌ای امکان به امانت گرفتن آن از کتابخانه مورد نظر وجود دارد.

/// سازماندهی تحقیقات پیشین

هنگامی که محقق متقاعد شد که یک بررسی منطقاً جامع از تحقیقات پیشین در زمینه مورد

نظرش انجام یافته است، می‌تواند به سازماندهی آن بپردازد. یک روش سودمند، منظم کردن بررسیها بر طبق موضوع و تعیین این است که هر یک از این موضوعات چگونه به بررسی مورد نظر مربوط هستند.

تحقیقات پیشین باید به نحوی ارائه شوند که با نشان دادن آنکه چه نتایجی به دست آمده و چه مواردی برای تحقیق در زمینه مورد نظر باقی مانده است، انجام آن را از طرف محقق توجیه کنند. فرضیات، چهارچوبی برای یک چنین سازماندهی فراهم می‌کنند. همانند مکتشفی که قصد سفر دارد، محقق نیز از قلمروی شناخته شده نقشه برداری می‌کند و راه رسیدن به قلمروی ناشناخته را نشان می‌دهد. اگر بررسی ابعاد متعددی داشته باشد یا در مورد بیش از یک فرضیه پژوهش کند، فرایند سازماندهی برای هر کدام مجزا خواهد بود.

محقق باید از وسوسه ارائه تحقیقات پیشین به عنوان مجموعه‌ای از چکیده‌ها پرهیز کند. به معنای دقیق‌تر، تحقیقات پیشین باید به گونه‌ای ارائه شوند که شالوده‌ای نظام‌مند برای بررسی فراهم سازند.

این مسئله نیز تقریباً اجتناب‌ناپذیر است که تعدادی از گزارشهایی که محقق به دقت مطالعه کرده و از آنها یادداشت نیز برداشته است، در رجوع بعدی با موضوع مورد مطالعه کم ارتباط به نظر آیند. آوردن کلیه بررسیهایی که محقق در جستجوی تحقیقات پیشین به آن برخورد می‌کند در طرح اولیه، نه ضروری و نه مطلوب است. خوانندگان به صرف کمیت تحقیقات پیشین تحت تأثیر قرار نخواهند گرفت. ربط و سازماندهی مواد از اولویت اول برخوردار است. چنانچه محقق از ابتدا روشی نظام‌مند برای گردآوری منابع و مآخذ موجود اختیار نکند، دچار بی‌نظمی خواهد شد. پیشنهادهای زیر می‌تواند در این زمینه چاره ساز باشد.

۱- از آخرین بررسیها در زمینه مورد نظر خود آغاز کنید و سپس به مطالعات قبلی باز گردید. یکی از نقاط قوت آشکار این رویکرد این است که شما با بررسیهایی آغاز می‌کنید که از تفکرات و یافته‌های گذشته ترکیب یافته‌اند. سوء تفاهمات اولیه تصحیح و از روشهای نامناسب دوری شده است. مزیت دیگر، این است که این بررسیها منابع مرجع کارهای اولیه را در بر دارند، بنابراین ما را به منابعی هدایت می‌کنند که در غیر این صورت، با آنها مواجه نخواهیم شد. آشکار است که در فرآیند گردآوری تحقیقات پیشین، باید محدوده‌ای تعیین شود. از یک سو، فراهم آوردن شالوده‌ای پر معنی برای تحقیق متضمن برخورداری از تمام کارهای مهم گذشته در

همان زمینه است و از سوی دیگر، صرف وقت بیش از حد در این کوشش می تواند خوانندگان گزارش شما را به دلیل ذکر جزئیات زاید خسته کند. مطمئن شوید که ماخذ و منابع مربوطه، به بررسی شما خدمت می کنند و نه حکومت.

۲- ابتدا چکیده یا خلاصه یک گزارش را مطالعه کنید تا بتوانید معین کنید که این گزارش با سؤال شما رابطه دارد یا خیر. این کار از اتلاف وقت زیاد در جریان خواندن مقالات غیر سودمند جلوگیری می کند.

۳- قبل از یادداشت برداری، گزارش را به طور سریع بخوانید تا بخشهایی را که با سؤال شما مرتبط است، پیدا کنید. این، راه دیگری برای جلوگیری از اتلاف وقت است.

۴- یادداشتهای خود را بر روی فیش کارت بنویسید، چون مرتب کردن و سازماندهی آنها نسبت به کاغذهای متفرقه راحت تر است. بسیاری، فیش کارتهای ۶-۴ اینچی را ترجیح می دهند، چرا که هم فضای قابل قبولی برای یادداشت برداری دارد و هم آنقدر کوچک است که می توان آنرا در پاکت و یا کیف جای داد.

۵- در جریان بررسی هر کاری کتاب شناسی کامل آنرا یادداشت کنید. اگر در آغاز کار بدانید که چه شیوه ای در پایان گزارش به کار خواهید برد، با یادداشت برداری بر اساس آن شیوه در زمان صرفه جویی خواهید کرد. شماره تلفن کتابخانه را نیز یادداشت کنید. این امر ممکن است در بازبازی مجدد بررسی ضروری باشد.

۶- برای تسهیل در مرتب کردن و سازماندهی کارتها، هرگز بیش از یک منبع را بر روی آنها یادداشت نکنید. مرتب کردن منابع شما به صورت الفبایی و یا با هر روش دیگری بدون در نظر گرفتن این مسئله ممکن نخواهد بود.

۷- مشخص کنید که چه بخشی از یادداشتها سوالات مستقیم مؤلف و کدامیک تفسیرهای شماست. عدم موفقیت در انجام این تفکیک، می تواند شما را سهواً به سوی دزدی تالیفات و افکار دیگران سوق دهد. همچنین عاقلانه خواهد بود که ارزیابی نویسنده از کار خودش را با ارزیابی خود از آن تفکیک کنید.

۸- اگر جستجوی کامپیوتری انجام شده باشد، شیوه جستجوی خود را در فایل کامپیوتر حفظ کنید. این اطلاعات به شما در بازبازی اطلاعات دقیق و کاهش هزینه روزآمد کردن اطلاعات مورد نیاز، کمک می کند.

## جدول ۱-۳ منابع تحقیقات مربوط به زمینه تعلیم و تربیت

منبع	محتوا
چکیده‌ها و کتابشناسی رشد کودک	چکیده‌های مقالات مجلات در زمینه رشد کودک
نمایه روز مجلات تعلیم و تربیت	عناوین، مؤلفان و مشخصات نشریات از زمینه مقالات مربوط به تعلیم و تربیت؛ با حاشیه نویسی‌های مورد نیاز. طبقه‌بندی برحسب موضوع، مؤلف، و مجله بدون چکیده.
چکیده رساله‌های بین‌المللی نمایه تعلیم و تربیت	چکیده رساله‌های دوره دکترا در ایالات متحده، کانادا و اروپا. عناوین، مؤلفان و مشخصات نشریات در زمینه مقالات مربوط به تعلیم و تربیت. فهرست شده بر اساس موضوع و عنوان. بدون چکیده.
دائرة المعارف تحقیق در تعلیم و تربیت مجموعه میکروفیشهای اریک	خلاصه‌ها و ارزشیابیهای تحقیقات منتشر شده در زمینه آموزش و پرورش در پایان هر دهه. اسناد کامل از تحقیقاتی که چکیده‌های آنها در منابع تعلیم و تربیت موجود است
منابع تعلیم و تربیت کودکان استثنایی	چکیده مقالات مجلات و منابع دیگر در تعلیم و تربیت
سالنامه اندازه‌گیری‌های روانی	اطلاعات و ارزشیابیهای آزمونهایی که به صورت تجاری در دسترس است.
کاتالوگ ماهانه نشریات دولتی	فهرست موضوعی نشریات دولت فدرال ایالات متحده
چکیده‌های روان‌شناختی	چکیده مقالات مجلات در زمینه روان‌شناسی
راهنمای خوانندگان در پیشینه مجلات علمی	عناوین، مؤلفان و مشخصات نشریات در زمینه مقالات دیگر نشریات مشهور؛ فهرست شده بر اساس موضوع و عنوان بدون چکیده.
منابع تعلیم و تربیت	چکیده گزارشهای تحقیقی و سایر اسناد گردآوری شده توسط دفتر مرکزی اریک
بررسی تحقیقات آموزشی	بررسی موضوعات متنوع در فصلنامه‌ها
نمایه نقل نامهای علوم اجتماعی نقد آزمون	اطلاعات کتاب‌شناختی از نویسندگان و موضوعات نقل شده بررسی وسایل اندازه‌گیری، شامل توصیفهای کلی؛ کاربردهای مفید / موارد استفاده؛ قدها، و جنبه‌های تکنیکی
آزمونهای چاپ شده شماره سه	فهرست و متمم سالنامه اندازه‌گیریهای روانی از شماره اول تا شماره هشت
گنجینه توصیفگرهای اریک	سیستم برای طبقه‌بندی و فهرست کردن اسناد اریک



### /// خلاصه

اگر محقق همهٔ راههای منتهی به اطلاعات مورد نظر را همانطور که پیشنهاد شد، بیمایند، تصویر کامل و قابل قبولی از جایگاه بررسی خود در آن حیطه تحقیقاتی خاص خواهد داشت. در جدول ۱-۳ مهمترین منابع و مضامین پیشینه تحقیقات در تعلیم و تربیت خلاصه شده است.

### // مفاهیم اصلی

جستجوی کامپیوتری	بانک اطلاعات
راهنماهای کتابشناختی	دائرةالمعارفهای تخصصی
سیستم اریک	راهنماهای اسناد دولتی
مرورهای دوره‌ای	فهرستهای رساله‌ها و پایان‌نامه‌ها
منطق بول	منابع آزمونها

### /// تمرینها

- ۱- سه نقش مهم تحقیقات پیشین در پروژه تحقیقی را بیان کنید.
- ۲- شایستگی CIJE و فهرست تعلیم و تربیت را با هم مقایسه کنید.
- ۳- کدامیک از راهبردهای ذیل برای سازماندهی جستجو در تحقیقات پیشین قابل قبول است.

- الف - سازماندهی بررسیها بر طبق موضوع
- ب - با مقالات اولیه آغاز کنید و به موقع به کارهای تازه بپردازید.
- ج - ابتدا بخشهای چکیده یا خلاصهٔ یک گزارش را مطالعه کنید.
- د - گزارش را به طور اجمال بخوانید تا بخشهایی را که به سؤالات شما مربوط است، بیابید.
- ه - برای هر کار یک منبع کتابشناسی کامل بنویسید.
- ۴- در چه نقطه‌ای از روند انجام پروژه تحقیقی، محقق باید تحقیقات پیشین و بررسیهای مربوط به سؤالات انتخابی و مورد نظر را مطالعه کند؟
- ۵ - حاشیه نویسی چیست؟

- ۶- اهمیت کتاب سالانه اندازه‌گیری ذهنی بوروس در ایجاد منابع آزمونها را توضیح دهید.
- ۷- چرا اریک چنین منبع مفیدی برای مشاوره در جستجوی پیشینه تحقیقات است؟
- ۸- موارد استفاده از فهرست نقل نامهای علوم اجتماعی در جستجوی پیشینه تحقیقات را توضیح دهید.
- ۹- مراحل معین در استفاده از سیستم اریک برای مروری بر تحقیقات پیشین کدام است؟
- ۱۰- مزایای استفاده از جستجوی کامپیوتری در پیشینه تحقیقات تخصصی بانکهای اطلاعاتی چیست؟
- ۱۱- چگونه از منطق بول در جستجوی بانکهای اطلاعاتی استفاده می‌شود؟
- ۱۲- در شکل ۱-۳ کدامیک از منابع مربوط به پیشینه تحقیقات در کتابخانه‌های قابل دسترس شما موجود است؟

### /// پاسخها

- ۱- شناخت تحقیقات پیشین محقق را قادر می‌سازد تا پیشتازان حیطه مورد نظر را تعیین کند، سؤال تحقیق را در یک دورنما قرار دهد، و از تکرار سهوی بررسیهای قبل اجتناب کند.
- ۲- فهرست تعلیم و تربیت راه دستیابی به مقالات منتشره را از ۱۹۲۹ تا به امروز فراهم می‌کند. CIEE بیشتر مجلات را پوشش می‌دهد و حاشیه‌نویسی را تهیه می‌کند، در حالی که فهرست تعلیم و تربیت مقالات را سریعتر از CIEE فهرست می‌کند.
- ۳- الف، ج، د و ه.
- ۴- محقق به مجرد آنکه موضوع تحقیق انتخاب و طرح شد، پیشینه تحقیقات را مورد بررسی قرار می‌دهد.
- ۵- یک حاشیه نویسی، توضیح کوتاهی از یک مقاله است و می‌تواند به محقق در انتخاب مقالات مربوط به موضوع کمک کند.
- ۶- آنها پر معناترین فهرست و تبیین از آزمونهای استاندارد شده موجود را ارائه می‌کنند. مرورهای دوره‌ای از آزمونها به محقق در انتخاب یک آزمون مناسب کمک می‌کند.
- ۷- اریک در پوشش دادن بسیاری از موضوعات در زمینه تعلیم و تربیت، یک سیستم جامع محسوب می‌شود. هر یک از بخشهای تخصصی به یک زمینه موضوعی ویژه اختصاص دارند.

که اطلاعات و مواد را جمع‌آوری و وارد سیستم می‌کنند. سیستم اریک همچنین محدوده وسیعی از مواد شامل: بررسیها و پروژه‌های تحقیقی، راهنماهای درسی، مقالات معتبر از انجمنها و مؤسسات شناخته شده، کتابشناسی‌ها، و موارد درسی را تحت پوشش دارد. بسیاری از این مواد چاپ نشده‌اند و امکان دستیابی به آنها در جای دیگر میسر نیست. سیستم اریک نه تنها مواد را فهرست می‌کند، بلکه از بیشتر آنها میکروفیش تهیه می‌کند.

۸- از طریق فهرست نقل نامها فرد می‌تواند با تعیین مقالاتی که از کارهای ویژه‌ای نام برده‌اند، یک خط تحقیقی رو به جلو را تعقیب کند. فهرست موضوعی اصطلاحات فرد را قادر می‌سازد تا مقالات را توسط کلمات معادل با عناوین آنها بیابد.

۹- کلمات کلیدی در بررسی خود را تعیین کنید.

الف - کلمات کلیدی یا معادلهای آنها را در دائرةالمعارف توصیفگران اریک پیدا کنید.

ب - از یک برنامه جستجوی کامپیوتری برای یافتن مقالات مجلات و آن دست از اسناد اریک، که حاوی اصطلاحات و یا توصیفگران کلیدی مورد نظر شماست، استفاده کنید، یا از طریق CIE و RIE برای یافتن شماره‌های دسترسی مقالات و اسناد با اصطلاحات معادل خود، به یک جستجوی غیر کامپیوتری (دست) پردازید.

ج - چکیده‌ها یا حاشیه‌نویسی‌ها را برای حذف مقالات و اسناد بی‌ربط مطالعه کنید.

د - مقالات و اسناد باقیمانده اریک را شناسائی کنید.

ه - ابتدا چکیده، خلاصه و نتیجه‌گیری را بخوانید.

۱۰- جستجوگر کامپیوتری می‌تواند در سیستم بانکهای اطلاعاتی سریعاً به جستجو پردازد. آنها بویژه برای شناسایی مقالات و اسناد دیگر، با کمک ترکیب خاص از کلمات کلیدی، مفید هستند. آنها می‌توانند به طرق مختلف، برای مثال: عنوان، تاریخ، نویسندگان، کلمات کلیدی، یا رمزهای موضوع، به جستجو پردازند. اطلاعات بانکهای اطلاعاتی کامپیوتری از فهرستهای چاپی تازه‌تر هستند و اطلاعاتی را شامل می‌شوند که درجای دیگری قابل دستیابی نیستند.

۱۱- جستجو به طریق A و B تنها اسناد دارای هر دو توصیفگر را شناسایی می‌کند.

جستجو از طریق A یا B همه اسناد با یک توصیفگر را شناسایی می‌کند.

جستجو از طریق A و نه B بدون شناسایی اسناد با توصیفگر B همه اسناد با توصیفگر

A را شناسایی می‌کند.

۱۲- پاسخها متفاوت خواهند بود.



## فصل چهارم

### فرضیه

#### اهداف آموزشی

دانشجو پس از مطالعه این فصل، قادر خواهد بود:

- ۱- فرضیه را تعریف کند.
- ۲- هدف از فرضیه (ها) را در یک بررسی تحقیقی بیان کند.
- ۳- بین فرضیه استقرایی و قیاسی تمایز قائل شود.
- ۴- معیارهای مورد استفاده در ارزشیابی فرضیات تحقیق را بیان کند.
- ۵- یک تعریف کاربردی را بیان و یک مثال ارائه کند.
- ۶- از مثالهای داده شده یک فرضیه قابل آزمون اقامه کند.
- ۷- یک فرضیه صفر را تعریف و هدف از آن را در یک بررسی تحقیقی تشریح کند.
- ۸- یک فرضیه تحقیقاتی و یک فرضیه صفر را برای یک بررسی تحقیقی طرح کند.
- ۹- فرضیات جهت‌دار و غیر جهت‌دار تمایز قائل شود.
- ۱۰- مراحل آزمون یک فرضیه را تعریف کند.
- ۱۱- هدف (هدفها) از انجام یک بررسی مقدماتی را بیان کند.
- ۱۲- عوامل مورد لزوم در یک طرح تحقیقی را فهرست کند.

فرضیه، ابزاری قدرتمند در تحقیق علمی است. این ابزار ما را قادر می‌سازد تا نظریه رابه‌مشاهده و مشاهده رابه نظریه مربوط سازیم. استفاده از فرضیه‌ها، انسان عصر حاضر را قادر می‌سازد که ایده‌های فیلسوفان استقرایی را با توجه به تاکید آنها بر مشاهده، و منطق فیلسوفان قیاسی را با توجه به تاکید آنها بر تعقل، با هم به کار گیرند. استفاده از فرضیه‌ها، تجربه و تعقل را برای خلق ابزاری قدرتمند جهت جستجوی حقیقت پیوند می‌دهد.

محقق، بعد از تعیین و بیان مسئله و بررسی متون تحقیقاتی پیشین، برای ساختن فرضیه خود آماده است. یک فرضیه می‌تواند به صورت یک گزاره<sup>(۱)</sup> آزمایشی پیشنهادی، به عنوان راه حل یک مسئله، یا به عنوان توصیف چند پدیده، به طور دقیق تعریف شود. فرضیه، بیان ساده انتظارات محقق نسبت به رابطه بین متغیرهای موجود در مسئله است و سپس در جریان یک بررسی تحقیقاتی مورد آزمون قرار می‌گیرد. اولین روست که فرضیه تنها به عنوان راه حل پیشنهادی برای یک مسئله و با درک این موضوع که نتیجه تحقیق می‌تواند باعث ابقا و یا رد آن شود، ارائه می‌شود.

برای مثال، محقق ممکن است با این سؤال آغاز کند: ادراک کودکان از خود، چه نقشی در فرآیند یادگیری خواندن آنها ایفا می‌کند؟ او سپس، باید فرض کند: بین ادراک کودکان کلاس اول از خود و پیشرفت در امر خواندن رابطه مستقیم وجود دارد. یا محقق ممکن است با سؤالی از این دست شروع کند: آموزش پیش دبستانی چه اثری بر پیشرفت تحصیلی کودکان کلاس اول که محرومیت فرهنگی دارند، می‌گذارد؟ فرضیه را می‌توان این گونه عنوان کرد: در کلاس اول، پیشرفت تحصیلی کودکان دارای محرومیت فرهنگی که آموزش پیش دبستانی را گذرانده‌اند، نسبت به کودکانی که محرومیت فرهنگی دارند، اما آموزش پیش دبستانی را ندیده‌اند، بیشتر است. در هر دو مثال، می‌توان مشاهده کرد که فرضیه، گزاره‌ای است که دو متغیر را به هم ربط می‌دهد. در مثال اول، متغیرها، ادراک خویشتن و پیشرفت در امر خواندن هستند، و در مثال بعدی متغیرها، آموزش پیش دبستانی و پیشرفت تحصیلی در کلاس اول هستند.

اگرچه صاحب‌نظران استفاده از فرضیه را توصیه می‌کنند، و همان طور که خواهیم دید، فرضیه در خدمت مقاصد متعدد و مهمی است، اما وجود آن در همه مطالعات تحقیقی کاملاً ضروری نیست. فرضیات در فرآیند تحقیق یک ابزار هستند، اما به خودی خود هدف نیستند. بررسیها

اغلب در حیطه‌ای صورت می‌گیرند که مجموعه‌ی کوچکی از اطلاعات پس زمینه‌ای در آن مورد وجود داشته باشد. اگر محقق پیشی نسبت به حوزه مسئله مورد نظر، یا متغیرهای اصلی که در یک پدیده مؤثر هستند، یا محیطی که متغیرها در آن بروز می‌کنند، نداشته باشد، آن وقت است که تبیین یک فرضیه با معنا، برای او بسیار دشوار خواهد بود. برای مثال، پیمایشهایی که سعی در توضیح ویژگیهای پدیده‌ای خاص دارند، یا در پی تعیین نگرشها و عقاید گروهها هستند، اغلب کار تحقیق را بدون فرضیه دنبال می‌کنند. در این مرحله، داده‌های تجربی که بر طبق آن فرضیاتی را بتوان بیان کرد، هنوز در دسترس محقق نیست. در چنین بررسیهایی بیان مسئله تحقیق کافی است. در تحقیقاتی که رابطه علی بین متغیرها مورد بررسی قرار می‌گیرد، فرضیه‌ها اهمیت بسیار پیدامی‌کنند.

فرضیه باید قبل از مرحله گردآوری داده‌ها ساخته شود. این امر به دو دلیل ضروری است: (۱) فرضیه با شالوده درست نشان می‌دهد که محقق در حیطه تحقیق خود دارای دانش معتبری است. (۲) فرضیه به گردآوری و تفسیر داده‌ها جهت می‌دهد، فرضیه به ما می‌گوید چه شیوه‌ای را در پیش بگیریم و چه داده‌هایی را گردآوری کنیم و بنابراین می‌تواند به مقدار زیادی از اتلاف وقت و تلاش محقق جلوگیری کند. باید تاکید شود که این امر، نه تنها برای تحقیق آزمایشی، بلکه برای همه انواع تحقیق درست است.

هدفهای کاربرد فرضیه را می‌توان اینگونه خلاصه کرد:

۱- فرضیه‌ها توصیفهایی آزمایشی از پدیده‌ها فراهم می‌آورند و میزان گسترش دانش ما را در یک حیطه خاص افزایش می‌دهند. انسان، برای رسیدن به دانش معتبر در زمینه مسائل تعلیم و تربیت باید از گردآوری صرف واقعیت‌های مجزا از هم فراتر رود تا به تعمیم‌ها و روابط متقابل موجود در این واقعیتها دست یابد. این روابط متقابل و تعمیم‌ها طرحی را فراهم می‌آورند که برای درک مسئله مهم است. چنین طرحی، احتمالاً تا زمانی که گردآوری داده‌ها جهت مشخصی نداشته باشد، آشکار نخواهد شد. فرضیاتی که خوب طراحی شده‌اند، جهت را مشخص و توصیفها را پیشنهاد می‌کنند. از آنجا که فرضیه‌ها می‌توانند با بررسی علمی، مورد آزمون قرار گیرند و اعتبار یابی شوند، امکان بسط دانش ما را فراهم می‌آورند.

۲- فرضیه، محقق را با یک بیان رابطه‌ای مجهز می‌کند که به طور مستقیم در جریان مطالعه علمی قابل آزمون است. سوالات نمی‌توانند به طور مستقیم مورد آزمون قرار گیرند. تحقیق با

یک سؤال آغاز می‌شود، ولی تنها روابط پیشنهادی بین متغیرها می‌توانند مورد آزمون قرار گیرند. برای مثال، محقق نمی‌تواند این سؤال را مورد آزمون قرار دهد: آیا اظهار نظرهای معلم در مورد نوشته‌های دانش‌آموزان باعث پیشرفت قابل توجهی در کار آنها می‌شود؟ در عوض، محقق فرضیه‌ای را که این سؤال بر آن دلالت دارد، آزمون می‌کند: اظهار نظر معلم در مورد نوشته‌های دانش‌آموزان موجب پیشرفت قابل توجهی در عملکرد آنها می‌شود. یا به طور مشخص‌تر، نمره عملکرد دانش‌آموزانی که از اظهار نظر معلم خود درباره نوشته‌های قبلی خویش برخوردار شده‌اند، از نمره عملکرد دانش‌آموزانی که چنین امکانی نداشته‌اند، بیشتر خواهد بود. سپس محقق به سوی بررسی رابطه بین این دو متغیر، یعنی نظرات معلم و عملکرد دانش‌آموز حرکت خواهد کرد.

۳- فرضیه‌ها تحقیق را هدایت می‌کنند. فرضیه هدف مشخصی را معرفی می‌کند و به این ترتیب ماهیت داده‌های ضروری را برای آزمون رابطه پیشنهادی تعیین می‌کند. به زبان ساده‌تر، فرضیه به محقق می‌گوید که چه کند. واقعیتها باید انتخاب شوند و مشاهدات در ارتباط با سؤال مشخص صورت گیرند و این فرضیه است که ارتباط بین این واقعیتها را تعیین می‌کند. فرضیه‌ها بنیانی برای انتخاب نمونه و روشهای مورد استفاده در تحقیق فراهم می‌آورند و تجزیه و تحلیل آماری مورد نیاز و رابطه‌هایی که باید مورد آزمون قرار گیرند، نیز از طریق فرضیه بیان می‌شوند. بعلاوه، فرضیه بررسی را در چهارچوب لازم نگه می‌دارد و از گستردگی بیش از حد آن جلوگیری می‌کند.

برای مثال، فرضیه مربوط به آموزش پیش دبستانی کودکان دارای محرومیت فرهنگی و پیشرفت تحصیلی آنها را در کلاس اول در نظر بگیرید. این فرضیه، روش تحقیق ضروری و نمونه مورد استفاده را نشان می‌دهد و حتی محقق را در بکارگیری آزمونهای آماری مورد لزوم در تجزیه و تحلیل داده‌ها، راهنمایی می‌کند. از نحوه بیان فرضیه روشن است که محقق یک بررسی علی - مقایسه‌ای انجام خواهد داد، یعنی کودکان کلاس اول را که دارای محرومیت فرهنگی هستند و از آموزش پیش دبستانی برخوردار بوده‌اند با گروه مشابه، که آموزش پیش دبستانی نداشته‌اند، مقایسه می‌کند. هر تفاوتی در میانگین پیشرفت تحصیلی دو گروه را می‌توان با روشهای معنی‌دار آماری نظیر آزمون ۱ و تحلیل واریانس مورد تجزیه و تحلیل قرار داد (این روشها در فصل ۶ مورد بحث قرار خواهند گرفت).

۴- فرضیه‌ها چهارچوبی برای گزارش نتایج بررسی فراهم می‌آورند. چنانچه محقق، هر فرضیه را جداگانه در نظر گیرد و نتیجه‌گیریهای مربوط به هر یک را تبیین کند، کار خود را بسیار



آسان خواهد یافت. به این ترتیب، او می‌تواند این بخش از گزارش‌نویسی را بر حول تداوک پاسخهایی برای فرضیات اصلی سازمان دهد. این امر موجب می‌شود که گزارش پرمعنا تر و خواندنی‌تر شود.

### /// پیشنهادهایی برای استنتاج فرضیه‌ها

محقق چگونه یک فرضیه را استنتاج می‌کند؟ همچنان که در فصل دوم توضیح داده شد، یک بررسی می‌تواند از مسائل عملی، از موقعیتهای رفتاری مشاهده شده که نیازمند توصیف است، یا حتی بهتر از آن، از نظریه‌های آموزشی، روان‌شناختی یا جامعه‌شناختی نشئت گرفته باشد. بنابراین، فرضیه‌ها به طور استقرایی از مشاهده رفتار و یا به طور قیاسی از نظریه یا یافته‌های تحقیقات پیشین استنتاج می‌شوند.

### // فرضیه‌های استقرایی

روش استقراء عبارت از صورت‌بندی یک فرضیه، از طریق تعمیم رابطه‌های مشاهده شده است؛ به این معنا که، محقق مشاهداتی در مورد رفتار صورت می‌دهد، به روندها یا رابطه‌های احتمالی توجه می‌کند و سپس برای توصیف این رفتار مشاهده شده فرضیه‌ای را ارائه می‌کند. البته، این فرایند استدلالی، برای تعیین یافته‌های محققان دیگر باید با بررسی تحقیقات قبلی همراه شود. روش استقرایی منبع سودمندی از فرضیه‌ها برای معلم است. او هر روز رفتار دانش‌آموزان خود را مشاهده و سعی می‌کند این رفتار را با رفتار خود، رفتار سایر دانش‌آموزان، با روش تدریس خود و با تغییرات محیط مدرسه و غیره مربوط کند. این معلم می‌تواند بر مبنای چنین تجربه‌ای و بر اساس دانش خود در زمینه رفتار در محیط مدرسه، به طور استقرایی تعمیمی را صورت‌بندی کند که رابطه مشاهده شده را توصیف می‌کند. به هر حال اعتبار این توصیف برای تبدیل شدن به فرضیه، باید از طریق یک بررسی علمی تعیین شود.

شاید آموزگاری مشاهده کند که آزمونهای کلاسی او موجب بروز اضطراب شدیدی در دانش‌آموزان می‌شود و فکر کند که این اضطراب بر عملکرد آنها تأثیر منفی می‌گذارد. بعلاوه به این مسئله نیز توجه کرده باشد که وقتی به دانش‌آموزان امکان نوشتن نظراتشان را در مورد سؤالاتها می‌دهد، به نظر می‌رسد که عملکرد آزمون آنها بهتر می‌شود. معلم استدلال می‌کند که آزادی ابراز

نظرات می‌باید به هر حال باعث کاهش اضطراب در دانش‌آموزان شده و در نتیجه نمرات آنها را افزایش داده باشد. این مشاهده یک فرضیه را صورت‌بندی می‌کند: نمره دانش‌آموزانی که به نوشتن نظرات خود در مورد سوالات آزمون بر روی پاسخنامه ترغیب شده‌اند، از نمره دانش‌آموزانی که این امکان را نداشته‌اند، بیشتر خواهد شد. سپس او می‌تواند برای آزمون اعتبار این فرضیه آزمایشی را طراحی کند. توجه کنید که فرضیه، اعتقاد معلم را در مورد رابطه بین دو متغیر نشان می‌دهد: ابراز نظر در مورد پرسشهای آزمون و عملکرد در آزمون. همچنین توجه کنید اضطراب، که حلقه‌ای از زنجیره قیاس برای ساخت فرضیه به شمار می‌آید، بخشی از فرضیه نهایی را به خود اختصاص نداده است. بنابراین، نتایج تحقیق تنها اطلاعاتی را در مورد رابطه بین ابراز نظر و عملکرد در آزمون فراهم خواهد آورد. رابطه بین اضطراب و نظرات و اضطراب و عملکرد آزمون می‌توانند موضوعاتی برای فرضیه‌های آتی جهت تحقیق باشند. محقق اغلب در خواهد یافت که ایده اصلی او مجموعه‌ای از روابط را دربردارد که به طور مستقیم قابل مشاهده نیستند. او سؤالش را برای تحقیق در مورد روابطی که قابلیت مشاهده مستقیم دارند، دوباره صورت‌بندی می‌کند.

در ادامه، چند مثال ساده دیگر از فرضیاتی که می‌توانند حاصل مشاهدات یک معلم باشند، ارائه می‌شوند: بین اضطراب در مورد ریاضی و اضطراب در مورد کامپیوتر در بین دانش‌آموزان دوره دبیرستان، رابطه مثبت وجود دارد؛ یادگیری کامپیوتر از طرف دانش‌آموزان در دوره اول دبیرستان، رشد تفکر منطقی را در آنها افزایش می‌دهد؛ استفاده از دفترچه یادداشتهای الکترونیکی پیشرفته<sup>(۱)</sup> یادگیری دانش‌آموزان دبیرستانی را از آموزش به کمک کامپیوتر در درس شیمی افزایش می‌دهد؛ آموزش خاصیت جابجایی، سرعت یادگیری ترکیبهای عمل جمع را افزایش می‌دهد؛ کودکانی که در گروههای کوچک آموزش می‌بینند، در آزمون نهایی خواندن در کلاس اول نمره بیشتری از آن دسته از دانش‌آموزان که در گروههای بزرگتر آموزش دیده‌اند، کسب می‌کنند؛ و رشد شناختی و احساسی کودکان کلاس اول، تحت تأثیر مجموعه تجربیات قبلی آنها در کودکان قرار دارد. محقق در فرآیند استقراء در وهله اول مشاهده می‌کند، سپس در مورد مسئله فکر می‌کند و در مرحله بعدی برای یافتن سرنخ به مآخذ موجود مراجعه می‌کند، بعد مشاهدات بیشتری انجام می‌دهد و پس از آن فرضیه‌ای را که سعی در تبیین علتی برای رفتار

مشاهده شده دارد، صورتبندی می‌کند. در مرحله بعدی فرضیه تحت شرایط کنترل شده آزمون می‌شود تا فرضیه‌های معلم دربارهٔ رابطه بین متغیرها به طور علمی اعتباریابی شوند. تحقیق در مورد فرضیه‌های استقرایی که از مشکلات روزمره استنتاج شده‌اند، اغلب می‌تواند در تعیین راه حل یک چنین مسائلی مفید باشد. به این دلیل که این گونه فرضیات از مسائل مشخص عملی سرچشمه می‌گیرند، به هر حال نتایج حاصل از آنها می‌تواند مجموعه‌ای از یافته‌های پرارزش، ولی ناپیوسته را فراهم آورد که از نظر توان توصیفی، محدودیت دارند.

### // فرضیه‌های قیاسی

بر خلاف فرضیاتی که با تعمیم از رابطه‌های مشاهده شده صورتبندی می‌شوند، فرضیات قیاسی از طریق قیاس از نظریه ساخته می‌شوند. اینگونه فرضیات دارای این مزیت هستند که می‌توانند به سیستم کلی تری از دانش، به مثابه چهارچوبی برای تلفیق معنادار آنها در پیکره معرفتی که از پیش وجود دارد، منتج شوند. اگر هر بررسی در حد یک کوشش مجزا باقی بماند، علم نمی‌تواند به طور مؤثری پیشرفت کند. یک علم با اتکاء به پیکره موجود واقعیها و نظریه‌ها می‌تواند به انباشتگی برسد. یک فرضیه استنتاج شده از نظریه به عنوان یک فرضیه قیاسی شناخته می‌شود.

شاید تشخیص بین نظریه و فرضیه مفید باشد. یک نظریه شامل مجموعه‌ای از مفاهیم، به همراه احکامی است که به چگونگی و چرایی روابط درونی موجود بین این مفاهیم می‌پردازد. نظریه‌ها به صورت تبیینهای کلی ارائه می‌شوند که در دامنه وسیعی از پدیده‌ها به کار می‌روند. در یک نظریه بیان روابط اغلب چیزی بیشتر از فرضیه‌هایی نیست که درست انگاشته شده‌اند؛ فرضیاتی که می‌توانند در بررسیهای بعد مورد آزمون قرار گیرند. فرد می‌تواند بر مبنای روابط درونی پیشنهادی در یک نظریه، پیامدهای خاص را که به طور منطقی می‌توان فرض کرد از آن روابط حادث می‌شوند، بیان کند. این پیامدهای فرض شده پایه‌هایی برای فرضیات هستند. یک نظریه علمی باید متضمن نتایجی باشد که بتوانند در جریان یک تحقیق تجربی تأیید شوند، یعنی فرد باید بر اساس نظریه قادر به پیش‌بینی رویدادهای خاصی باشد که قابل مشاهده یا غیر قابل مشاهده هستند. این پیامدهای قیاسی تبدیل به فرضیه‌هایی می‌شوند که موضوع بررسیهای تجربی هستند.

هنگامی که فرضیات مشتق شده از یک نظریه در تحقیقی مورد تأیید قرار می‌گیرند، به همان ترتیب نظریه نیز مورد تأیید قرار می‌گیرد. بنابراین، فرضیات شواهدی فراهم می‌کنند که نظریه‌ها را تأیید می‌کنند، بسط می‌دهند، نقض می‌کنند و یا به تجدید نظر در نظریه‌هایی منجر می‌شوند که از آن نشئت گرفته‌اند. باید بر این امر تأکید کرد که تحقیق هر چند می‌تواند نظریه‌ای را رد کند، ولی هرگز قادر به اثبات درستی یک نظریه نیست. دلیل این امر آن است که نظریه‌ها تعمیم‌هایی هستند که برای تمام موارد ممکن پدیده‌ای که سعی در تبیین آن دارند، به کار می‌روند و این امکان وجود ندارد که نظریه را در مورد تمام احتمالات، مورد آزمون قرار داد. اما، یک نظریه هر قدر تأیید بیشتری در تحقیقات مختلف به دست آورد، ما را بیشتر مطمئن می‌کند که معتبر و مفید است.

نظریه‌ها از نظر محتوا نسبت به فرضیه کلی‌تر هستند. یک نظریه ممکن است منشأ تعدادی فرضیه برای تحقیق در بررسیهای متعدد و مجزا باشد. یک محقق ممکن است مطالعه‌ای را با انتخاب یکی از نظریه‌های موجود در حیطه مورد علاقه‌اش آغاز کند. البته انتخاب نظریه طبق هدف تحقیق و سهمی که نظریه می‌تواند در فهم مسئله ایفا کند، صورت می‌گیرد. هنگامی که نظریه انتخاب شد، محقق برای آنکه به پیامدهای منطقی نظریه برسد، با استفاده از استدلال قیاسی، از آن نظریه، فرضیه استخراج می‌کند. این استدلال قیاسی سپس تبدیل به فرضیات بررسی می‌شوند. محققى که به بررسی انگیزه پیشرفت در بین کودکان علاقه‌مند است، باید نظریه کلاسیک مک کللند در مورد انگیزه پیشرفت را مطالعه کند.

یکی از فرضیه‌های نظریه مک کللند در باب انگیزه این است که شدت انگیزه پیشرفت تابعی است که به طور مستقیم با آموزش استقلال و خودکفایی تناسب دارد.<sup>(۱)</sup> محقق از این نظریه می‌تواند پیش‌بینی کند کودکانی که والدینشان استقلال و خودکفایی را در آنها ترغیب می‌کنند، در کلاس انگیزه بیشتری برای پیشرفت از خود بروز می‌دهند. محقق سپس برای رسیدن به پیامدهای منطقی نظریه که می‌تواند به طور تجربی به اثبات برسد، از استدلال قیاسی استفاده می‌کند. در بررسی تحقیقی فرضی زیر، این استنتاج قیاسی، فرضیه محسوب می‌شود. مسئله و فرضیه می‌توانند به شکل زیر بیان شوند:

مسئله: چه رابطه‌ای بین شیوه‌های پرورش کودکان در منزل و انگیزه پیشرفت در بین دانش‌آموزان دوره دبستان وجود دارد؟

فرضیه: کودکانی که والدینشان بشدت آزادی آنها را محدود می‌کنند، در مقایسه با کودکان دیگری که والدینشان آزادی آنها را محدود نمی‌کنند، در یک تکلیف درسی، که مقدار کار انجام شده برای آن تابعی از انگیزش آنهاست، نمره‌های پایین‌تری کسب می‌کنند. محقق سپس برای آزمون این فرضیه به طراحی یک بررسی، و به احتمال قوی از نوع علی - مقایسه‌ای، اقدام می‌کند. اگر نتایج پیش‌بینی شده مشاهده شوند، نظریه مک کلند در مورد انگیزه پیشرفت مورد تأیید قرار می‌گیرد. اگر هیچ تفاوتی در دو گروه از کودکان مشاهده نشود، محقق باید نظریه یا حداقل کاربرد آن در این شرایط را مورد سؤال قرار دهد.

محققی که به رشد تفکر منطقی در بین کودکان علاقه‌مند است، احتمالاً به نظریه پیازه مراجعه می‌کند. پیازه عقیده داشت که کودکان در جریان رشد ذهنی خود از مراحل مختلفی می‌گذرند. یکی از مراحل، مرحله عملیات عینی است که از ۷ یا ۸ سالگی آغاز می‌شود و مشخصه آن انتقال از مرحله وابستگی به ادراک، به مرحله بکارگیری برخی عملیات منطقی است. این عملیات در یک سطح عینی قرار دارند، ولی استدلال نمادین را نیز دربردارند.<sup>(۱)</sup> با استفاده از این نظریه به عنوان یک نقطه شروع، محقق می‌تواند این فرضیه را ارائه کند: نسبت کودکان ۹ ساله‌ای که قادر به پاسخگویی درست به این مسئله از نوع استدلال انتقالی هستند که «فرانک بلندتر از جورج است؛ جورج کوتاهتر از ابرت است؛ چه کسی از همه بلندتر است؟» از نسبت کودکان ۶ ساله‌ای که قادر به پاسخگویی درست به این سؤال هستند، بیشتر است.

در مطالعه‌ای که برای آزمون قیاس از یک نظریه طراحی شده است، واریسی هر یک از فواصل منطقی حایل بین نظریه و فرضیه از اهمیت فوق‌العاده‌ای برخوردار است. محقق باید از خود سؤال کند که آیا فرضیه به طور منطقی از نظریه پیروی می‌کند؟ اگر فرضیه به طور حقیقی از نظریه پیروی نکند، محقق قادر به دستیابی به نتایج معتبر در مورد شایستگی نظریه نخواهد بود. اگر فرضیه مورد تأیید قرار گرفته باشد، ولی به دقت از نظریه قیاس نشده باشد، محقق نمی‌تواند بگوید یافته او اعتباری برای نظریه فراهم آورده است. برعکس، اگر داده‌ها فرضیه را تأیید نکردند، نظریه‌ای که این فرضیه از آن نشئت گرفته است، لزوماً نامعتبر نخواهد بود. این یک حقیقت است که بسیاری از فرضیات که می‌توان آنها را از نظریه‌های شناخته شده تر قیاس کرد،

تاکنون مورد بررسی قرار گرفته‌اند، ولی بسیاری از چنین قیاس‌هایی هنوز ساخته و آزمون نشده‌اند. همچنین یک قیاس که در گذشته مورد تحقیق قرار گرفته است، می‌تواند برای ایجاد فرضیه‌هایی در شرایط وسیعتر و متنوعتر و برای وسعت بخشیدن به کاربرد نظریه مورد استفاده قرار گیرد.

### /// ویژگیهای فرضیه‌های قابل استفاده

بعد از آنکه فرضیه به طور آزمایشی صورتبندی شد و قبل از هرگونه تلاشی برای آزمون تجربی آن، توانایی فرضیه به عنوان ابزار تحقیق باید سنجیده شود. یک فرضیه باید معیارهای ویژه‌ای را دارا باشد. ارزش نهایی یک فرضیه نمی‌تواند قبل از آزمون تجربی آن مورد قضاوت قرار گیرد، اما معیارهای ویژه‌ای وجود دارند که فرضیه‌های ارزشمند را مشخص می‌سازند و محقق باید برای قضاوت در مورد شایستگی فرضیه پیشنهادی از آنها استفاده کند.

#### // یک فرضیه باید از قدرت تبیین برخوردار باشد

فرضیه باید امکان تبیین آنچه را که سعی در توصیف آن دارد، داشته باشد. این یک معیار واضح، ولی مهم است. برای روشن شدن مسئله، فرض کنید که شما سعی دارید اتومبیل خود را روشن کنید، ولی موفق نمی‌شوید. این فرضیه که اتومبیل روشن نمی‌شود چون شما شیر آب حمام را باز گذاشته‌اید، قابل قبول نیست، اما فرضیه‌ای که می‌گوید باطری اتومبیل تمام شده است، یک تبیین قابل قبول است و ارزش آزمون شدن را دارد.

#### // یک فرضیه باید رابطه مورد انتظار بین متغیرها را بیان کند

یک فرضیه باید رابطه مورد انتظار بین دو یا چند متغیر را گمانه زند. در مثال قبل، بیان اینکه «اتومبیل روشن نخواهد شد و یک سیستم سیم‌پیچی دارد»، به این دلیل که هیچ رابطه‌ای بین متغیرها مشخص نشده است و از اینرو هیچ رابطه‌ای پیشنهادی برای آزمون وجود ندارد، مفید نخواهد بود. یک فرضیه سودمند این خواهد بود: اتومبیل روشن نخواهد شد، به این دلیل که در سیستم سیم‌پیچی آن اشکالی وجود دارد. این معیار به طور آشکار روشن است، ولی تبیین زیر را ملاحظه کنید: اگر کودکان در خودآنگاره با یکدیگر متفاوت باشند، از نظر پیشرفت در درس علوم

اجتماعی نیز با یکدیگر متفاوت خواهند بود. از نگاه اول به نظر می‌رسد که این جمله یک فرضیه باشد، اما با کمی دقت روشن می‌شود که در آن هیچ رابطه مورد انتظاری بیان نشده است. یک رابطه مورد انتظار می‌توانست این گونه بیان شود: خودانگاره والاتر احتمالاً پیشابندی برای موفقیت در درس علوم اجتماعی است. این فرضیه را می‌توان به این نحو نیز بیان کرد: بین خودانگاره و موفقیت در درس علوم اجتماعی یک رابطه مثبت وجود دارد. اگر خلاف آن پیش‌بینی شده باشد، یعنی خودانگاره والاتر به موفقیت کمتر در درس علوم اجتماعی منجر شود، فرضیه این خواهد بود: بین خودانگاره و موفقیت در درس علوم اجتماعی یک رابطه منفی وجود دارد. هر دو بیان دومین معیار ما را برآورده می‌سازند.

### // یک فرضیه باید قابل آزمون باشد

گفته شده است که مهمترین ویژگی یک فرضیه «خوب»، داشتن قابلیت آزمون است. این بدان معنی است که فرضیه قابل اثبات است. بر این اساس است که قیاسها، نتیجه‌گیریها یا معانی ضمنی را به نحوی فراهم می‌آورند که مشاهدات تجربی، فرضیه را رد و یا تأیید کنند. اگر فرضیه درست باشد، پس باید نتایج قابل پیش‌بینی متجلی شوند. یک فرضیه قابل آزمون محقق را قادر می‌سازد تا از طریق مشاهده تعیین کند که پیامدهایی که به طور قیاسی استدلال شده است، در عمل رخ خواهند داد یا خیر. در غیر این صورت تأیید و یا رد فرضیه غیر ممکن خواهد بود. در مثال مابین فرضیه که «روشن نشدن ماشین، کیفی است برای گناهان من»، در دنیای ما به طور آشکار قابل آزمون نیست.

بسیاری از فرضیات و یا گزاره‌هایی که ممکن است به طور بنیادین نیز بیان شده باشند، اساساً قابل آزمون نیستند. برای مثال فرضیه «تجربیات پیش دبستانی، سازگاری همه جانبه کودکان را در دوره پیش از دبستان ارتقاء می‌بخشد»، به دلیل مشکل بودن تعریف و اندازه‌گیری مفهوم سازگاری همه جانبه، بسیار مشکل خواهد بود. مثال دیگر در مورد یک فرضیه غیر قابل آزمون این است: شیوه تقلید از آثار مشهور در درس هنر مدرسه، باعث خاموشی خلاقیت هنری دانش‌آموزان می‌شود. در این مورد یکی از مشکلات، تعریف و اندازه‌گیری خلاقیت هنری و همچنین تعیین معیار برای تشخیص وقوع خاموشی در خلاقیت است.

یک فرضیه قابل آزمون، باید متغیرهایی را به هم ربط دهد که قابلیت اندازه‌گیری داشته

باشند. اگر هیچ وسیله‌ای برای اندازه‌گیری متغیرها موجود نباشد، گردآوری داده‌های ضروری برای آزمون اعتبار فرضیه، غیر ممکن خواهد بود. در این مورد نمی‌توان تاکید بسیار کرد. بدون امکان تعریف دقیق شاخصهای هر متغیر و در پی آن اندازه‌گیری این متغیرها، فرضیه قابل آزمون نخواهد بود.

از شاخصهای متغیرها به عنوان تعریف کاربردی یاد می‌شود. تعریف کاربردی، همانطور که پیش از این توضیح داده شد، یک متغیر را با تبیین «عملیات» یا روشهای لازم برای اندازه‌گیری آن، مشخص می‌کند. برای مثال، به این فرضیه توجه کنید بین عزت نفس<sup>(۱)</sup> در کودک و پیشرفت در امر خواندن در کلاس اول دبستان رابطه مثبت وجود دارد. برای آنکه این فرضیه از معیارهای قابلیت پذیرش بهره‌مند شود، تعریف کاربردی متغیرها ضروری است. عزت نفس را می‌توان به عنوان نمره‌های به دست آمده از مقیاس عزت نفس کوپر سمیت<sup>(۲)</sup> و پیشرفت در امر خواندن را به عنوان «نمره‌های به دست آمده از آزمون خواندن کالیفرنیا» و یا رتبه‌بندی معلم کلاس اول در پیشرفت خواندن تعریف کرد.

یکی از ملاحظات اساسی در صورتبندی یک فرضیه، اطمینان یافتن از این امر است که می‌توان متغیرها را به طور کاربردی تعریف کرد. از بکارگیری سازه‌ها در تعریف متغیرها پرهیزید؛ چرا که یافتن اندازه‌های مناسب برای آنها مشکل و یا غیرممکن خواهد بود. توافق بر روی تعریفی کاربردی از سازه‌هایی مانند خلاقیت، استبدادگرایی، دموکراسی و امثالهم، از آنجا که این مفاهیم معانی متعددی به خود می‌گیرند، اگر غیرممکن نباشد، مشکل است. به خاطر داشته باشید که اصطلاحات باید به رفتار عینی قابل شناسایی اطلاق شوند.

پرهیز از تبیینات ارزشی در قالب یک فرضیه، امر مهمی است. عبارتی مانند «یک برنامه مشاوره‌ای در مدارس ابتدایی مطلوب است»، نمی‌تواند در یک بررسی تحقیقی مورد مطالعه قرار گیرد؛ هر چند که محقق می‌تواند این فرضیه را مورد آزمایش قرار دهد: دانش‌آموزان کلاس اول که از مشاوره برخوردار شده‌اند، به طور کلامی نسبت به دانش‌آموزانی که فاقد این امکان بوده‌اند، رضایت بیشتری از مدرسه ابراز خواهند کرد. محقق می‌تواند تعبیرات کلامی ناشی از رضایت را اندازه‌گیری کند، اما اینکه آیا آنها مطلوب هستند یا نه، یک قضاوت شخصی است.

1- Self- Esteem

2- Coopersmith, S. (1981). The Antecedents of Self- Esteem. San Francisco: Consulting Psychology.



## // یک فرضیه باید با پیکره دانش موجود سازگار باشد

فرضیه‌های جدید نباید ناقض فرضیه‌ها، نظریه‌ها و قوانین تثبیت شده قبلی باشند. فرضیه‌ای نظیر «ماشین من روشن نخواهد شد، چون آب باطری آن تبدیل به طلا شده است»، اگر چه معیار اول را داراست، ولی چنان با آگاهی از طبیعت مواد تناقض دارد که ارزش پیگیری ندارد. فرضیه «ماشین روشن نخواهد شد، چون آب باطری آن تا پایین‌ترین سطح تبخیر شده است»، با دانش موجود سازگار است و بنابراین ارزش پیگیری را دارد. فرض اینکه بین خود انگارهٔ پسران و دختران نوجوان و میزان رشد بدنی آنها رابطه‌ای وجود ندارد، بی‌فایده است؛ چرا که انبوه شواهد چنین رابطه‌ای را تأیید می‌کنند. در تاریخ علم، مردانی مانند انیشتین، نیوتن، داروین و کپرنیک را مشاهده می‌کنیم که حقیقتاً فرضیاتی انقلابی را ارائه کردند که با دانش آن عصر مغایرت داشته است، ولی باید به یاد داشت که کار چنین پیشتازانی تکذیب کامل دانش گذشته نبوده، بلکه سازماندهی مجدد آن، در نظریه‌های ابداع‌کننده تری بوده است. در بیشتر موارد و بخصوص برای محقق تازه کار، بهترین پیشنهاد این است که فرضیه، با آخرین آگاهی‌های حاصله در زمینه مورد نظر منطبق باشد. دوباره بر ضرورت بررسی کامل مآخذ و منابع موجود تأکید می‌کنیم، چون فرضیه‌ها بر اساس نتایج تحقیقات گزارش شده در زمینه مورد نظر صورتبندی خواهند شد.

## // فرضیه‌ها باید تا حد امکان ساده و موجز بیان شوند

بیان فرضیه به روشی ساده نه تنها آزمون آنرا بسیار آسان‌تر می‌کند، بلکه به هنگام نتیجه‌گیری نیز مبنایی برای ارائه گزارشی روشن و قابل درک فراهم می‌آورد. گاهی ممکن است لازم باشد که یک فرضیه کلی و گسترده را برای آنکه قابل آزمون و واضح باشد، به چند فرضیه کوچک مشخص تقسیم کنیم. برای مثال توما و لیوسن این فرضیه‌های کلی را مورد توجه قرار دادند: وضعیت اقتصادی - اجتماعی خانواده در تعیین میزان سازگاری نوجوان در موقعیتهای مختلف اجتماعی نقش دارد، یا اجزای گوناگون این وضعیت اجتماعی از نظر میزان تأثیر بر نگرش نوجوان نسبت به اقتدار تفاوت می‌کنند. توما و لیوسن برای آنکه این فرضیه‌های کلی را واضح و قابل آزمون سازند، آنها را به چند فرضیه مشخص تقسیم کردند: (۱) بین نگرش یک پسر نوجوان نسبت به اقتداری که در خانه تجربه می‌کند و وضعیت اقتصادی - اجتماعی خانواده‌اش، یک رابطه معنی‌دار منفی وجود دارد؛ (۲) بین نگرش یک پسر نوجوان نسبت به اقتداری که در

مدرسه تجربه می‌کند و موقعیت اقتصادی - اجتماعی خانواده‌اش یک رابطه معنی‌دار منفی وجود دارد، (۳) بین نگرش یک پسر نوجوان نسبت به اقتداری که در بین هم‌تایانش تجربه می‌کند و وضعیت اقتصادی - اجتماعی خانواده‌اش یک رابطه معنی‌دار منفی وجود دارد، (۴) بین نگرش پسران نوجوان نسبت به اقتدار و سطح تحصیلات پدران‌شان یک رابطه معنی‌دار منفی وجود دارد؛ (۵) بین نگرش پسران نوجوان نسبت به اقتدار و سطح تحصیلات مادران‌شان یک رابطه منفی معنی‌دار وجود دارد و غیره.<sup>(۱)</sup> برای انعکاس نگرشهای دختران نوجوان ضروری است که این فرضیه‌ها از نو بیان شوند. بیان فرضیات در این شکل بسیار مشخص و جزئی، گزارش یافته‌ها و نتایج را نیز تسهیل می‌کند. فرد آزمایشگر می‌تواند هر فرضیه را به طور جداگانه در نظر بگیرد و یافته‌ها و نتایج مربوط به هر یک را ارائه کند.

بنابراین، چنین به نظر می‌رسد که ممکن است در یک مطالعه نیاز به فرضیه‌های متعددی باشد. عموماً توصیه می‌شود که محقق برای هر یک از وجوه فرعی مسئله و یا هر ابزار گردآوری داده‌ها که به کار می‌برد، یک فرضیه بیان کند. برای مثال، یک محقق ممکن است این فرضیه را بیان کند: دانش‌آموزانی که ریاضی را از طریق PLATO (یک نرم افزار کامپیوتری آموزشی) آموخته‌اند، در مفاهیم ریاضی، یادگیری و یادداری بیشتری را نسبت به دانش‌آموزانی نشان خواهند داد که با کتابهای سنتی و متداول آموزش می‌بینند. از آنجا که گزارش نتایج، هم برای یادگیری و هم برای یادداری ضروری خواهد بود، فرضیه کلی باید به مثابه دو فرضیه مجزا مورد آزمون قرار گیرد. این فرضیات را می‌توان اینگونه بیان کرد: (۱) دانش‌آموزانی که ریاضیات را از طریق PLATO یاد گرفته‌اند، در مفاهیم ریاضی یادگیری بیشتری را نسبت به دانش‌آموزانی نشان خواهند داد که از طریق کتابهای سنتی و متداول آموزش دیده‌اند، و (۲) دانش‌آموزانی که ریاضیات را از طریق PLATO یاد گرفته‌اند، یاد داری بیشتری را نسبت به دانش‌آموزانی نشان خواهند داد که از طریق کتابهای سنتی و متداول آموزش دیده‌اند. به این ترتیب، امکان نشان دادن این مسئله که داده‌های به دست آمده در هر یک از موارد مجزا سؤال کلی را تأیید می‌کنند، وجود دارد. ممکن است داده‌ها تأثیر آموزش کامپیوتری را در یادگیری اولیه نمایان کنند، اما در مورد یادداری آنها چیزی نشان ندهند. محقق نباید از اطناب کلامی حاصل از بیان چندین فرضیه هراس داشته باشد. به یاد داشته باشید که دو اصل قابلیت آزمون و وضوح، با کمک فرضیه‌های

1- Tuma, and N. Livson, (1960) Family Socio Economic Status and Adolescent Attitude to Authority. Child Development, 31, 382-399.

مشخص تر به نحو بهتری تأمین می‌شود.

همچنین توصیه می‌شود که در تبیین فرضیه از ساده‌ترین واژه‌ها و اصطلاحات استفاده شود تا مفهوم مورد نظر انتقال یابد، از بکارگیری سازه‌های مبهم باید پرهیز شود. اصطلاحات را به شیوه‌ای به کار برید که به طور کلی برای اشاره به پدیده مورد نظر، قابل قبول باشند. وقتی دو فرضیه دارای قدرت توصیف یکسانی باشند، آنکه ساده‌تر است، باید مرجح باشد؛ زیرا تبیین ضروری را با مفروضه‌ها و متغیرهای کمتری که نیاز به تعریف داشته باشند، ارائه می‌کند. به یاد داشته باشید که اصل ایجاز در زارزشیابی فرضیه‌ها مهم است.

بسیاری از فرضیات که صورتبندی شده‌اند، بعد از آزمون تجربی رد خواهند شد. آنها پیش‌بینی‌هایی بوده‌اند که داده‌ها آنها را تأیید نکرده‌اند. در تاریخ تحقیق علمی، تعداد فرضیاتی که مورد تأیید قرار نگرفته‌اند، بیشتر از آنهایی بوده است که تأیید شده‌اند. محققان با تجربه عقیده دارند که فرضیات تأیید نشده، یک بخش مفید و مورد نظر در تجربیات علمی هستند. آنها قادرند تحقیق را به سمت بازبینی مجدد نظریه سوق دهند و اغلب ما را به توصیفی دقیق از وضعیت امور نزدیکتر کنند. داروین گفته است: «من دائماً تلاش کرده‌ام تا مغزم را از هر گونه فرضیه رها کنم، هر چند بسیار به آن علاقه‌مند بوده‌ام (و نمی‌توانم در مقابل ساختن آن در مورد هر موضوعی مقاومت کنم)، اما واقعیات سرعت خلاف این را نشان دادند. برآستی هیچ چاره‌ای جز حرکت در این راستا نداشتم. به استثنای فرضیه صخره‌های مرجانی، من هیچ فرضیه جدید مجرد را به یاد نمی‌آورم که بعد از مدتی یا کاملاً کنار گذاشته نشود یا کاملاً تغییر نیابد.»<sup>(۱)</sup> حتی یک فرضیه تأیید نشده نیز می‌تواند مفید باشد؛ زیرا ضرورت توجه به سایر جنبه‌های یک مسئله را مسجل می‌کند. از این روی می‌تواند محقق را در رسیدن به توصیفی قابل قبول، یک گام جلو تربرد. در صورتبندی فرضیه باید از ابهام پرهیز شود.

به جز موارد استقرای کامل، حتی اگر محقق تأییدی برای فرضیه خود بیابد، به معنای اثبات درستی آن نخواهد بود. یک فرضیه هرگز اثبات و یا رد شده نیست، فرضیه تنها تأیید می‌شود و یا تأیید نمی‌شود. فرضیات در ماهیت خود اساساً احتمالی هستند، شواهد تجربی می‌توانند محقق را به سوی این نتیجه‌گیری هدایت کنند که تبیین او احتمالاً حقیقت دارد یا پذیرفتن فرضیه منطقی است، ولی هرگز فرضیه را ثابت نمی‌کنند. نمونه‌ای از یک بررسی که فرضیه آن از نظریه قیاس شده است، در اینجا ارائه می‌شود.

1- Darwin F.(ed). (1919). Life and Letters of Charles Darwin. New York: Appleton.

تأثیر ساعت آموزش طی روز در پیشرفت تحصیلی<sup>(۱)</sup>

## مسئله تحقیق

این که آموزش در چه ساعتی از روز صورت گرفته است، در پیشرفت درس زبان انگلیسی و ریاضیات دانش آموزان کلاس هشتم چه تأثیری دارد؟

## نظریه

نظریه‌ای در مورد حافظه کوتاه مدت (STM)<sup>(۲)</sup> و حافظه بلند مدت (LTM)<sup>(۳)</sup> می‌گوید که تکالیفی که عمدتاً به استفاده از حافظه کوتاه مدت مربوط می‌شوند، در صبح بهتر یاد گرفته و انجام می‌شوند و تکالیفی که عمدتاً به استفاده از حافظه بلند مدت مربوط می‌شوند، در بعدازظهر بهتر یاد گرفته و انجام می‌شوند. دو عامل روان شناختی دلیل تفاوت‌های یادگیری را در اوقات روز بیان می‌کنند: (الف) انگیزتگی پایه از سطح نازل در صبح به بالاترین حد آن در عصر می‌رسد و استفاده از حافظه کوتاه مدت و بلندمدت به طور قابل توجهی با آن همبسته هستند؛ (ب) به همین ترتیب ریتم‌های روزانه در سطوح پلازما و هورمون، در شیوه‌ای که افراد اطلاعات را رمزگشایی، بایگانی و بازیابی می‌کنند، مؤثر است و با پردازش بهتر حافظه بلند مدت در اواخر روز و پردازش بهتر حافظه کوتاه مدت در ساعت اولیه روز همبسته است.

## فرضیه

به دلیل آنکه مهارت خواندن، اطلاعات چاپی را با آگاهی و تجربه قبلی مرتبط می‌کند، این فرضیه صورتبندی می‌شود: دانش آموزان هنگامی در خواندن پیشرفت خواهند کرد که آموزش آنها در بعدازظهر انجام شود. و چون آموزش مهارت‌های ریاضی نیازمند دستکاری بیشتر داده‌ها در ذهن و استفاده از حافظه کوتاه مدت است، فرضیه اینگونه صورتبندی می‌شود: دانش آموزان هنگامی در ریاضی پیشرفت بیشتری خواهند کرد که آموزش آنها به جای عصر، در صبح صورت گیرد.

1- Davis. Z.T.(1988). The Effect of Time- of- day of Instruction on Eight-Grade Student' English and Mathematics Achievement. The High School Journal, 71, 78- 80.

2- Short- Term Memory

3- Long- Term Memory

## روشها

هشتاد دانش‌آموز کلاس هشتم به طور تصادفی در گروه‌های تیمار ذیل مشخص شدند: (۱) دوره اول انگلیسی، (۲) آخرین دوره انگلیسی، (۳) دوره اول ریاضی، (۴) آخرین دوره ریاضی. یک معلم انگلیسی واحد و یک معلم ریاضی واحد، بخشهای دوره اول و دوره آخر درسها را با استفاده از روشها و مواد درسی یکسان آموزش دادند. یک آزمون جامع مهارتهای پایه‌ای (CTBS)<sup>(۱)</sup> در شروع (پیش آزمون) و پایان (پس آزمون) سال تحصیلی برای اندازه‌گیری میزان پیشرفت به اجرا در آمد. آزمونهای مربوط به هر محتوا در یک روز واحد داده شد و به موازات آن در ساعات میانه روز برای کنترل تأثیر بالقوه ساعات روز در میزان بازدهی، به همه گروههای دیگر نیز این آزمون داده شد.

## تحلیل

مجموعه نمرات آزمونهای مربوط به انگلیسی و ریاضی هر کدام به وسیله تحلیل کوواریانس<sup>(۲)</sup> و با استفاده از نمرات پیش آزمون به عنوان متغیر کمکی، تحلیل شد. تحلیل کوواریانس تغییرات موجود بین نمرات پیش آزمون دانش‌آموزان را در نظر می‌گیرد و نمرات پس آزمون را در رابطه با آن تطبیق می‌دهد، به گونه‌ای که می‌توان مقایسه‌های معتبری صورت داد.

## نتایج

گروه انگلیسی بعد از ظهر به نسبت گروه انگلیسی صبح پیشرفت معنی دارتری داشتند. هیچ تفاوت معنی داری بین پیشرفت گروههای ریاضی صبح و بعد از ظهر وجود نداشت.

## نتیجه‌گیری

نتیجه آن که دانش‌آموزان در درس انگلیسی کلاس هشتم، از آموزش بعد از ظهر، بیش از آموزش صبح بهره می‌گیرند. ساعت آموزش روزانه هیچ تفاوتی در پیشرفت درس ریاضی دانش‌آموزان کلاس هشتم نداشت. محقق نتیجه گرفت که فرضیه مربوط به درس ریاضی باید مورد بررسی بیشتری قرارگیرد. شاید کارکرد حافظه‌های کوتاه مدت و بلند مدت در تحقیق قبلی، به طور کامل معرف آن چیزی نبود که در استدلال ریاضی مورد نیاز است، یا شاید استدلال ریاضی استفاده نسبتاً متعادل از حافظه‌های کوتاه مدت و بلند مدت را در بردارد.

## /// انواع فرضیه

## // فرضیه تحقیق

فرضیاتی را که پیش از این مورد بحث قرار دادیم، در نهایت فرضیه تحقیق و یا اساساً فرضیات می‌نامیم. آنها فرضیاتی هستند که از طریق مشاهده، پیشینه تحقیق و یا نظریه‌ای که در تحقیق توصیف شده است، ساخته می‌شوند. فرضیه تحقیق گزاره‌ای است درباره رابطه‌ای که محقق انتظار دارد آن را به عنوان نتیجه تحقیق پیدا کند. فرضیه می‌تواند بیانی درباره رابطه یا تفاوت مورد انتظار بین متغیرهای موجود در بررسی باشد. یک فرضیه درباره هوشبهر و اضطراب کودکان در کلاس درس را می‌توان اینگونه بیان کرد: یک رابطه مثبت بین هوشبهر و اضطراب کودکان در دوره اول دبستان وجود دارد، یا کودکان با هوشبهر بالا نسبت به کودکان با هوشبهر پایین در کلاس درس اضطراب بیشتری از خود نشان می‌دهند. فرضیات تحقیق را می‌توان به شکل جهت‌دار یا غیر جهت‌دار بیان کرد. یک فرضیه جهت‌دار ماهیت رابطه یا تفاوت مورد انتظار را مشخص می‌کند. دو نمونه فرضیه در مورد هوشبهر و اضطراب، که در بالا ذکر شد، از نوع فرضیات جهت‌دار هستند. از سوی دیگر، یک فرضیه غیر جهت‌دار، وجود یک رابطه و یا تفاوت را بیان می‌کند، ولی ماهیت یافته مورد انتظار را مشخص نمی‌کند. برای مثال، در کودکان بین هوشبهر و اضطراب رابطه وجود دارد.

## // فرضیه صفر

فرضیات تحقیق نمی‌توانند از طریق شیوه‌های آماری موجود مستقیماً مورد آزمون قرار گیرند. در بررسیهای تجربی که در آنها آزمونهای آماری باید مورد استفاده قرار گیرند، فرضیه تحقیق باید به نوع دیگری از فرضیه که به فرضیه صفر معروف است (با  $H_0$  نشان داده می‌شود)، تبدیل شود. این فرضیه را از اینرو فرضیه صفر می‌نامند که بیان‌کننده «هیچ تفاوت»، «هیچ اثر» یا «هیچ رابطه‌ای» می‌باشد. یک فرضیه صفر، نقیض آنچه را بیان می‌کند که آزمایشگر انتظار دارد و یا پیش‌بینی می‌کند. یک آزمایشگر ممکن است مشتاق نشان دادن این مسئله باشد که بعد از یک تیمار آزمایشی، دو جامعه آماری میانگین‌های متفاوتی خواهند داشت، اما فرضیه صفر می‌گوید که میانگین‌های جامعه‌های آماری با یکدیگر تفاوت نخواهند داشت.

یک فرضیه صفر به این دلیل به کار می‌رود که محققان را قادر می‌سازد تا یافته‌های خود را با

کمک روشهای آماری در مقابل انتظارات اتفاقی مقایسه کنند. فرضیه صفر فرض می‌کند از آنجایی که تفاوت‌های مشاهده شده صرفاً از روی اتفاق رخ می‌دهند، اصلاً نشان‌دهنده تفاوت‌های واقعی نیستند. آزمونهای آماری برای تعیین احتمال صحیح بودن فرضیه صفر به کار می‌روند. اگر آزمونها نشان دهند که احتمال وقوع تفاوت‌های مشاهده شده به طور اتفاقی بسیار کم است، فرضیه صفر به یک فرض بسیار غیر محتمل تبدیل می‌شود. بنابراین، می‌توان آن را به نفع فرضیه مقابل رد کرد. شواهد برای این نتیجه‌گیری غیر قطعی که تفاوت واقعی است، معتبر است. اگر تفاوت‌های مشاهده شده پسادگی ناشی از اتفاق باشند، شواهد معتبر نیست و فرضیه صفر حفظ می‌شود. آزمون یک فرضیه صفر شبیه کار دادستان در یک دادگاه جنایی است. دادستان به منظور اثبات جرم (در سیستم حقوقی ایالات متحده) باید شواهد معتبری فراهم کند تا هیأت منصفه بتواند از ورای تردیدهای منطقی، بی‌گناهی مجرم را رد کنند. برای دادستان این امکان وجود ندارد که به طور قطع جرم را ثابت کند، و به همین ترتیب محقق نیز نمی‌تواند برای یک فرضیه تأیید بی‌چون و چرا به دست آورد. وکیل مدافع، چنین فرض می‌کند که تا وقتی شواهد معتبری مبنی بر مجرم نبودن متهم وجود دارد، وی بی‌گناه است، و تا زمانی که شواهد معتبر خلاف آن را ثابت کند، فرضیه صفر صحیح فرض می‌شود.

اجازه بدهید که به تحقیق فرضی سطوح اضطراب بین کودکان با هوشبهر بالا و پایین برگردیم. فرضیه تحقیق را می‌توان اینگونه بیان کرد: کودکان با هوشبهر بالا نسبت به کودکان با هوشبهر پایین اضطراب بیشتری از خود نشان خواهند داد. فرضیه صفر می‌گوید: سطح اضطراب کودکانی که هوشبهر بالا دارند با سطح اضطراب کودکان با هوشبهر پایین متفاوت نیست.

فرضیه صفر از طریق روشهای آماری متعددی مورد آزمون قرار می‌گیرد که در فصل ششم تشریح شده است. اگر به هنگام تکمیل بررسی، داده‌های تجربی نشان دهند که تفاوت بین گروههای نمونه بیشتر از آن است که صرفاً ناشی از اتفاق باشد، پس می‌توان فرضیه صفر را رد کرد. رد فرضیه صفر نشان می‌دهد که برابری سطح اضطراب دو گروه از کودکان محتمل نیست. محقق سپس یک فرضیه مقابل (با  $H_1$  نشان داده می‌شود) را می‌پذیرد که اینگونه بیان می‌شود: میانگین نمرات اضطراب در جامعه آماری برابر نیست. گفته می‌شود که این فرضیه مقابل غیر جهت‌دار است، چون نشان نمی‌دهد که متوسط نمرات کدام گروه بیشتر خواهد بود. احتمالاً بیشتر تحقیقات در زمینه تعلیم و تربیت با فرضیات غیر جهت‌دار آغاز می‌شوند. هر چند در

بعضی از موقعیتها، ممکن است محقق برای بیان یک فرضیه جهت‌دار (همانطور که از نام آن پیداست، تعیین جهت یافته‌های مورد انتظار) یک دلیل صریح داشته باشد. یک نمونه از فرضیه جهت‌دار این است: میانگین نمرات اضطراب کودکان با هوشبهر بالا به نسبت کودکان با هوشبهر پایین بیشتر خواهد بود.

### /// آزمون فرضیه

یک بررسی با یک فرضیه تحقیقی آغاز می‌شود، که باید بانی ساده و روشن، از رابطه مورد انتظار بین متغیرها باشد. پیش از این توضیح دادیم که فرضیه باید قابل آزمون باشد؛ به این معنا که در عمل قابل اثبات باشد. زمانی که محققان از آزمون یک فرضیه صحبت می‌کنند، در هر حال منظورشان فرضیه صفر است. متنها فرضیه صفر را می‌توان از طریق روشهای آماری، مستقیماً مورد آزمون قرارداد. آزمون یک فرضیه شامل این-مراحل است:

- ۱- در صورت درستی فرضیه تحقیق، روابطی را که باید مورد مشاهده قرار گیرند، در قالب اصطلاحات کاربردی، بیان کنید.
- ۲- فرضیه صفر را بیان کنید.
- ۳- روش تحقیقی انتخاب کنید که امکان مشاهده و یا انجام آزمایش لازم برای نشان‌دادن وجود یا عدم این روابط را برای شما فراهم آورد.
- ۴- داده‌های تجربی را گردآوری و تحلیل کنید.
- ۵- معین کنید که آیا شواهد برای رد فرضیه صفر کافی هستند یا خیر.

### // یک مثال از آزمون فرضیه

یک مثال ممکن است در روشن‌تر شدن فرآیند آزمون تجربی یک فرضیه کمک کند. فرض کنید که یک محقق، مشتاق بررسی نظریه تقویت در کلاس درس است. معلم بر مبنای درک خود از نظریه تقویت، ممکن است فرض کند که تحسین یا تشویق باعث افزایش انگیزه در بین پاره‌ای از کودکان می‌شود. اگر این فرضیه درست باشد، منطقی خواهد بود اگر فرض کنیم که پیامد نظرات تشویق آمیز معلم بر روی ورقه‌های آزمون شاگردان باید موجب بهبود عملکرد دانش‌آموزان شود. این فرض ایجاب می‌کند که افزایش انگیزه با بهبود نمرات آزمون دانش‌آموزان نمایان شود.



(مرحله اول) این معنای قیاسی را می‌توان اینگونه بیان کرد: اظهار نظر معلم بر ورقه‌ها موجب بهبود عملکرد دانش آموزان در آزمون‌ها می‌شود. بین دو متغیر، نظرات معلم و عملکرد دانش آموزان، یک رابطه وجود دارد که باید مورد آزمون قرار گیرد.

(مرحله دوم) برای آزمون آماری، فرضیه تحقیقی بالا باید به یک فرضیه صفر تبدیل شود. فرضیه صفر اینگونه بیان می‌شود: نظرات معلم بر روی ورقه‌های دانش آموزان به بهبود عملکرد دانش آموزان در آزمون‌ها منجر نخواهد شد.

(مرحله سوم) این نوع فرضیه را می‌توان با یک آزمایش، مورد آزمون قرار داد. محقق می‌تواند به طور تصادفی تعدادی از کلاسها را برای بررسی انتخاب کند. در هر کلاس دانش آموزان باید به طور تصادفی به دو گروه تقسیم شوند: برای دانش آموزانی که در گروه الف قرار دارند، معلم درباره نحوه انجام آزمون نظرات تشویق آمیزی را می‌نویسد. این نظرات می‌تواند شامل کلمات تشویق آمیزی چون «عالی است»، «کار خود را به همین خوبی ادامه بده» و یا «کار تو پیشرفت داشته است»، باشد. این نظرات نباید با محتوای ورقه دانش آموزان یا تصحیح اشتباهات ویژه او مرتبط باشند. در غیر این صورت، بهبود عملکرد دانش آموزان را می‌توان به سود بخشی آموزشی این نظرات نسبت داد، نه به افزایش انگیزه. دانش آموزانی که در گروه ب قرار دارند، در ورقه آزمون خود هیچ اظهار نظری دریافت نمی‌کنند.

(مرحله چهارم) آموزگار، یک آزمون عینی که بخش معینی از محتوا را در بر می‌گیرد، به اجرا در می‌آورد. ورقه‌ها نمره گذاری می‌شوند و تیمار آزمایشی، به شیوه‌ای که در بالا توصیف شد، ارائه می‌شود. سپس معلم آزمون دوم را که از نظر سطح دشواری همانند آزمون اول و شامل محتوایی است که بعد از آزمون اول و تیمار آزمایشی تدریس شده است، به اجرا در می‌آورد. تغییر نمرات هر دانش آموز از آزمون اول به آزمون دوم باید مشخص شود و میانگین این تغییر برای گروه نیز به دست آید. سپس تعیین اینکه آیا متوسط افزایش نمره‌ها در آزمون دوم با تیمار آزمایشی (نظرات معلم بر روی ورقه‌ها رابطه دارد، از طریق تجزیه و تحلیل داده‌ها امکان پذیر خواهد بود. (مرحله پنجم) اگر این نتیجه به دست آید که دانش آموزانی که نظر معلم را دریافت کرده‌اند (گروه الف)، مشخصاً نمرات بالاتری از گروهی که این نظرات را دریافت نکرده‌اند (گروه ب)، به دست آورده‌اند، آن وقت است که می‌توان فرضیه صفر را رد کرد. رد فرضیه صفر به این معنا خواهد بود که اثرات دو روش همسان نیستند. محقق می‌تواند به طور غیرقطعی نتیجه بگیرد که نظرات معلم بر ورقه دانش آموزان منجر به بهبود عملکرد دانش آموزان در آزمون‌ها می‌شود.

// بررسی مقدماتی<sup>(۱)</sup>

قبل از آماده‌سازی طرح تحقیق، آزمایش کردن روشهای پیشنهادی بر روی تعداد محدودی از آزمودنیها می‌تواند مفید واقع شود. این اجرای آزمایشی یا بررسی مقدماتی بیش از همه به محقق کمک می‌کند تا در مورد عملی بودن و ارزش ادامه بررسی تصمیم‌گیری کند، این آزمایش امکان ارزیابی متناسب و عملی بودن وسایل گردآوری اطلاعات را فراهم می‌آورد. این آزمایش همچنین امکان آزمون مقدماتی فرضیهی را فراهم می‌کند و دلایلی جهت تثبیت و یا اصلاحات بعدی مورد نیاز فرضیه ارائه می‌کند.

بررسی مقدماتی، همچنین کنایت روشهای تحقیقی و اندازه‌های انتخاب شده برای متغیرها را نشان می‌دهد. مسائل پیش‌بینی نشده می‌توانند در این مرحله حل شوند، در نتیجه از صرف وقت و انرژی در مراحل بعدی جلوگیری به عمل می‌آید. یک بررسی مقدماتی فرصت مغتنمی است که بخصوص برای محقق تازه کار توصیه می‌شود.

## /// برنامه تحقیق

بعد از آنکه سؤال و فرضیه صورتبندی شدند، محقق برای تکمیل برنامه تحقیق آزمایشی خود آماده است. این نیاز وجود دارد که هدف تحقیق و چگونگی برنامه‌ریزی برای رسیدن به آن با جزئیات نوشته شود. تهیه برنامه تحقیق امری اساسی است و محقق را وادار می‌سازد که ایده‌های خود را به شیوه‌ای عینی بنویسد. بسیاری از ایده‌های اولیه تا وقتی که فرد آنها را بر روی کاغذ نیاورده است، نویدبخش هستند، اما بعد از آن که بر روی کاغذ آمدند، اشکالات و نارساییهای آنها آشکار می‌شود.

برنامه تحقیق مدون را می‌توان برای دریافت نظرات و انتقادات به دیگران ارائه کرد. یافتن نقایص و خطاها در یک برنامه مدون بسیار ساده‌تر از زمانی است که بخواهیم آنها را از طریق ارتباط کلامی بیان کنیم.

در این مرحله برنامه تحقیق نوعاً تنها یک طرح مقدماتی است و تا نگارش طرح نهایی و رسمی احتمالاً نیاز به تغییرات زیادی خواهد داشت. به هر حال بهتر است به خاطر داشته باشیم که هر چه طرح پیشنهادی اولیه کاملتر و با جزئیات بیشتر بیان شود، برای محقق مفیدتر است و از اتلاف وقت او نیز جلوگیری می‌کند. برنامه تحقیق شامل قسمتهای ذیل است.

## // مسئله

برنامه با یک بیان روشن از سؤالی که محقق قصد پاسخگویی به آنرا دارد، آغاز می‌شود. بیان مسئله در مورد روابط بین متغیرها سؤال و جامعه آماری مورد نظر را مشخص می‌کند. یک شرح کوتاه از زمینه نظری مسئله و پیشینه تحقیقات نیز باید در این بخش گنجانده شود.

## // فرضیه

سؤال با تبیینی موجز از فرضیه و یا فرضیاتی که باید مورد آزمون قرار گیرند، دنبال می‌شود. فرضیه جهت تحقیق را روشن می‌سازد. همه طرحهای آتی برای پروژه تحقیقی به بیان فرضیه بستگی دارند. همچنین بیان فرضیه و دلایل تأیید آن با روشنی هر چه تمامتر از طرف محقق ضروری است. این بخش از برنامه باید شامل تعاریف کاربردی از متغیرهای بررسی باشد.

## // طرح تحقیق

مرحله بعدی برنامه، شامل توصیفی در مورد طرح تحقیق - یعنی توصیفی از روشهایی است که در آزمون فرضیه دنبال می‌شوند. انتخاب یک روش مناسب برای آزمون فرضیه بسیار مهم است. یک سؤال آزمایشی را نمی‌توان از طریق روشهای توصیفی پاسخ گفت و همچنین عکس آن نیز صادق است. این بخش همچنین باید شامل فهرستی از اندازه‌ها یا وسایل مورد استفاده در گردآوری داده‌ها باشد. محققان باید آزمونها، مقیاسها و سایر وسایل مورد نیاز برای اندازه‌گیری متغیرها را تعیین کنند و اعتبار و پایایی این ابزار را بسنجند. هدف، انتخاب اندازه‌هایی است که تا حد امکان عینی و پایا باشند، بدون آنکه شایستگی آنها، یعنی متناسب بودن آنها با مفاهیمی که قرار است معرف آن باشند، قربانی شود.

## // نمونه

برنامه باید شامل توصیفی از جامعه آماری مورد نظر، یعنی نوع آزمودنیهای مورد مطالعه باشد. محقق باید توجه خود را به چگونگی دسترسی به این آزمودنیها نشان داده باشد. همچنین ضروری است که روشهای نمونه‌گیری خود را توصیف کند. جامعه آماری که از آن نمونه‌گیری می‌شود و علاوه بر آن شیوه‌هایی که در انتخاب نمونه دنبال خواهد شد و اندازه پیشنهادی نمونه نیز باید مشخص شود.

## // تجزیه و تحلیل آماری

این بخش شامل خط مشی محقق برای تجزیه و تحلیل آماری داده‌هاست. پیش از آنکه محقق شروع به گردآوری داده‌ها کند، باید روشهای آماری را که امکان پاسخگویی به سؤال تحقیق و یا آزمون فرضیه را فراهم می‌آورند، تعیین کند. در مرحله نخست محقق نیازمند توصیف یا خلاصه کردن داده‌های گردآوری شده از نمونه مورد بررسی است. سپس باید بتواند پایایی یا دقت استنباطها و تعمیم‌هایی را که از یافته‌های مربوط به گروه نمونه به دست آمده و به کل جامعه آماری عمومیت داده شده است، تخمین بزند. روشهای آماری به هر دوی این ویژه کارها کمک می‌کند. کار خلاصه کردن داده‌های به دست آمده از طریق آمار توصیفی صورت می‌گیرد. آمار استنباطی محقق را قادر می‌سازد تا با کمک داده‌های گروه نمونه، دست به استنتاج بزند.

بسیاری از محققان با تجربه، و همچنین افرادی که به تازگی فرایند تحقیق را آموخته‌اند، مشورت با یک کارشناس آمار را قبل از تکمیل طرح بررسی خود ضروری می‌یابند. در فصل ششم بحث کوتاهی در مورد نقش تجزیه و تحلیل آماری در آزمون فرضیه، ارائه خواهد شد.

## /// خلاصه

برای رسیدن به مراحل منطبق با بررسی تحقیقی، داشتن فرضیه‌هایی که بروشنی بیان شده‌اند، دارای اهمیت است. فرضیه، پیش‌بینی محقق در مورد نتیجه بررسی است. فرضیه‌ها به شیوه استقرایی از مشاهده و یا به شیوه قیاسی از یک نظریه شناخته شده استنتاج می‌شوند. تجربه، دانش و آشنایی با تحقیقات قبلی در حیطه مورد نظر در صورتبندی یک فرضیه رضایت‌بخش، عوامل مهمی به شمار می‌روند.

فرضیه در تحقیق، به عنوان عملکردی چند منظوره ایفای نقش می‌کند؛ زیرا توصیفی ارائه می‌کند که به طور تجربی قابل آزمون است و به گسترش دانش خدمت می‌کند. فرضیه با مشخص کردن روش تحقیق و نوع داده‌های مرتبط با حل مسئله، به تلاشهای محقق جهت می‌دهد. همچنین فرضیه چهارچوبی برای تفسیر و تبیین نتایج بررسی مهیا می‌سازد.

یک فرضیه خوب، باید این معیارهای مشخص را دارا باشد: (۱) باید قدرت توصیف داشته باشد، (۲) باید قابل آزمون باشد؛ به عبارتی باید متغیرهایی را با هم مرتبط کند که بتوان آنها را

اندازه‌گیری کرد، (۳) باید با داده‌های مسلط موجود سازگار باشد، و (۴) باید بروشنی و ایجاز هر چه تمامتر بیان شود.

بعد از صورتبندی یک فرضیه، مرحله بعدی تدوین یک برنامه برای انجام بررسی، شامل بیان مسئله و فرضیه و توصیف طرح تحقیق، گروه نمونه و تجزیه و تحلیل آماری مورد استفاده است. این برنامه بدون اولیه، امکانی برای محقق و دیگران فراهم می‌کند تا تعیین کنند آیا یک برنامه عملی برای آزمون فرضیه را می‌توان به اجرا درآورد؟

پس از صورتبندی و ارزشیابی فرضیه براساس معیارهای گفته شده، محقق آماده است تا فرضیه تحقیق را مورد آزمون تجربی قرار دهد. فرضیه صفر، نقیض آنچه که مورد انتظار محقق است، بیان می‌شود. مهم است که به یاد داشته باشیم که یک فرضیه، اثبات شدنی و یا اثبات ناشدنی نیست و تنها تأیید و یا رد می‌شود. یک فرضیه حتی اگر تأیید هم نشود، هنوز می‌توان مفید باشد، چون محقق را به سوی ارزشیابی مجدد استدلالها و رشوها و توجه به رویکردهای دیگر برای حل مسئله هدایت می‌کند.

### // مفاهیم اصلی

فرضیه استقرایی	آزمون فرضیه
فرضیه غیرجهت‌دار	فرضیه جهت‌دار
فرضیه صفر	فرضیه قیاسی
معیار برای ارزشیابی فرضیه	کارکردهای فرضیه
	نظریه

### /// تمرینها

- ۱- مقصود از ارائه فرضیه چیست؟
  - ۲- چه تفاوتی بین یک فرضیه استقرایی و قیاسی وجود دارد؟
  - ۳- بر مبنای هر یک از سؤالات تحقیقی که در ذیل فهرست شده است، یک فرضیه بیان کنید؟
- الف - اثر استفاده از روش کوئیزر در تدریس حساب ابتدایی چیست؟
- ب - آیا بین جنسیت معلم و افزایش مهارت خواندن در بین دانش‌آموزان دختر سیاه‌پوست

در مدارس ابتدایی رابطه‌ای وجود دارد؟

ج - آیا زندگی در محلات چندنژادی در نگرش فرد نسبت به مردم نژادهای دیگر تأثیر می‌گذارد؟

د - آیا بین نوع تشویق (مادی یا غیرمادی) و میزان یادگیری کودکان دارای محرومیت اقتصادی رابطه‌ای وجود دارد؟

ه - آیا آموزش پیش‌دبستانی شکاف آموزشی بین کودکان دارای محرومیت اقتصادی و کودکان بدون محرومیت اقتصادی را کاهش می‌دهد؟

و - آیا انتظارات معلم از شاگردان برای داشتن عملکردی خردمندانه تأثیری در عملکرد واقعی آنها دارد؟

۴ - فرضیه ذیل را در قالب فرضیه صفر بازنویسی کنید: کودکانی که مهارت خواندن آنها از سطح کلاس پایین‌تر است، نسبت به کودکانی که مهارت خواندن آنها از سطح کلاس بالاتر است، رضایت کمتری از مدرسه نشان می‌دهند.

۵ - شایستگی هر یک از فرضیات ذیل را ارزیابی کنید. اگر یک فرضیه شایسته است، دلیل شایستگی آن را بیان کنید و یک فرضیه شایسته بنویسید.

الف - سزاوار است که معلمان بیش از مسئولان مدارس حقوق دریافت کنند.

ب - دانش‌آموزان دوره اول دبستان که درس علوم اجتماعی را می‌گذرانند، نسبت به کسانی که این درس را نمی‌گذرانند، توان قضاوت روشن‌بینانه‌تری را در مورد امور سیاسی محلی خواهند داشت.

ج - برای تدریس ترکیبات ضرب به دانش‌آموزان کندآموز، دادن تکالیف و تجربیات کامپیوتری بهتر از استفاده از کارتهای کمک آموزشی است.

د - اگر دانش‌آموزان از نظر وضعیت اجتماعی - اقتصادی متفاوت باشند، از نظر نمرات درس انگلیسی نیز متفاوت خواهند بود.

ه - کودکانی که هیجان زایدالوصفی نشان می‌دهند، براساس مقیاس اندازه‌گیری اضطراب آشکار، اضطراب بیشتری را نیز نشان خواهند داد.

و - تقویت مثبت پاسخهای دانش‌آموز از طرف معلم در کلاس، احتمال پاسخهای آتی را کاهش می‌دهد.

- ۶ - یک فرضیه غیرجهت‌دار و یک فرضیه جهت‌دار مبتنی بر این سؤال تحقیقی بنویسید که چه رابطه‌ای بین سرعت بلوغ پسران نوجوان و خوددنگاره آنها وجود دارد.
- ۷ - چرا یک فرضیه باید پیش از شروع تحقیق بروشنی بیان شده باشد؟
- ۸ - فرضیات ذیل را به عنوان فرضیات تحقیقی و یا فرضیه صفر نام‌گذاری کنید:
- الف - دانش‌آموزانی در آزمونهای پیشرفت تحصیلی نمرات پایین خواهند گرفت که به سطوح بالای طبقه‌بندی بلوم رسیده باشند.
- ب - هیچ تفاوتی در عملکرد دانش‌آموزانی که ریاضی را با روش الف آموخته‌اند با دانش‌آموزانی که ریاضیات را با روش ب آموخته‌اند، وجود ندارد.
- ج - هیچ تفاوتی بین میانگین نمرات یادداری کودکانی که داروی آزمایشی را خورده‌اند با میانگین نمرات کودکانی که این دارو را نخورده‌اند، وجود ندارد.
- د - دانش‌آموزانی که تحت آموزش معلمان آزادمنش قرار داشته‌اند، نسبت به دانش‌آموزانی که تحت آموزش معلمان مقتدر بوده‌اند، مهارتهای بیشتری در حل مسئله نشان خواهند داد.
- ۹ - یک بررسی تحقیقی، که فرضیه‌ای را بیان می‌کند، پیدا کنید و سپس سعی کنید نظریه‌ای را که فرضیه از آن منتج شده است، شناسایی کنید.
- ۱۰ - فرضیه ذیل را نقد کنید: برای آنکه بعدها شهروندان بهتری داشته باشیم، در کلاس علوم اجتماعی دوم دبیرستان استفاده از روش بحث، سودمندتر از روش سخنرانی است.

### /// پاسخها

- ۱ - هدف فرضیه تهیه یک طرح پیشنهادی موقت (غیرقطعی) به عنوان راه‌حلی برای یک مسئله و یا تبیین برخی پدیده‌هاست.
- ۲ - با یک فرضیه استقرایی، محقق در مورد روابط دست به مشاهده می‌زند و سپس تبیینی برای رفتارهای مشاهده شده فرض می‌کند. از طریق یک فرضیه قیاسی، محقق فرضیه‌ای مبتنی بر نظریه‌ای شناخته شده به همراه دلیل منطقی برای فرضیه خود صورتبندی می‌کند.
- ۳ - الف - دانش‌آموزان ابتدایی که از طریق روش کوئیزر حساب را آموخته‌اند، نسبت به آنهایی که حساب را از این طریق نیاموخته‌اند، در آزمون ریاضی نمرات بیشتری خواهند گرفت.
- ب - دانش‌آموزان پسر سیاهپوست دوره ابتدایی که از طریق معلم مرد آموزش دیده‌اند،

نسبت به دانش‌آموزان پس‌سیاهپوست دوره ابتدایی که از طریق معلم زن آموزش دیده‌اند، نمرات بهتری در خواندن می‌گیرند.

ج - افرادی که در محلات چندنژادی زندگی می‌کنند، نسبت به افرادی که در محلات تفکیک شده زندگی می‌کنند، در برخورد با افراد دیگر نژادها نگرش مطلوب‌تری از خود نشان می‌دهند.

د - دانش‌آموزان ضعیفی که از طریق پاداشهای مادی تقویت شده‌اند، نسبت به دانش‌آموزان ضعیفی که با پاداشهای غیرمادی تقویت شده‌اند، در خواندن پیشرفت بیشتری را نشان می‌دهند.

ه - کودکان محروم و غیرمحرومی که در سنین کودکستان از آموزش پیش‌دستانی بهره‌مند بوده‌اند، نسبت به کودکان محروم و غیرمحرومی که در سنین کودکستان از آموزش پیش‌دستانی بی‌بهره بوده‌اند، از نظر آموزشی با فاصله کمتری از یکدیگر متمایز می‌شوند.

و - کودکانی که معلمشان از آنها انتظار زیادی برای عملکرد عاقلانه دارد، نسبت به کودکانی که معلمشان از آنها انتظار کمی برای عملکرد عاقلانه دارد، رفتار سطح بالاتری را نشان می‌دهند.

۴- بین کودکانی که مهارت خواندن آنها از سطح کلاس پایین‌تر است و کودکانی که مهارت خواندن آنها از سطح کلاس بالاتر است، هیچ تفاوتی در ابراز رضایت از مدرسه وجود ندارد.

۵ - الف - فرضیه نامناسب است؛ چون یک تبیین ارزشی است و نمی‌تواند در یک بررسی تحقیقاتی مورد پژوهش قرار گیرد. یک فرضیه معقول این است: معلمانی که دستمزد بیشتری از مسئولان مدرسه دریافت می‌کنند، نسبت به معلمانی که دستمزد کمتری دریافت می‌کنند، رضایت شغلی بیشتری نشان می‌دهند.

ب - فرضیه نامناسب است؛ چون قضاوت‌های روشن‌بینانه یک اصطلاح ارزشی است. یک فرضیه قابل قبول می‌تواند این باشد: دانش‌آموزانی که درس علوم اجتماعی دوره اول دبیرستانی را می‌گذرانند، نسبت به دانش‌آموزانی که این درس را نگذرانده‌اند، دانش بیشتری نسبت به امور سیاسی محلی از خود نشان می‌دهند.

ج - فرضیه نامناسب است؛ چون یک تبیین ارزشی و فاقد تعاریف کاربردی روشن و دقیق است. یک فرضیه قابل آزمون می‌تواند این باشد: دانش‌آموزانی که در ریاضی عملکردی



ضعیف دارند و ترکیبات ضرب را از طریق تکالیف کامپیوتری تمرین می‌کنند، در مجموع نسبت به دانش‌آموزانی که در ریاضی عملکردی ضعیف دارند، با صرف وقت یکسان ترکیبات ضرب را از طریق کارتهای کمک آموزشی تمرین می‌کنند، در یک آزمون ملاکی، به نسبت بیشتری، پاسخ درست می‌دهند.

د - فرضیه نامناسب است؛ چون در آن بیان یک رابطه مورد انتظار بین متغیرها وجود ندارد.

ه - کودکانی که انگیزه پیشرفت زیادی از خود نشان می‌دهند؛ بیش از کودکانی که انگیزه پیشرفت کمی دارند، در مقیاس اضطراب آشکار کودکان، امتیاز کسب خواهند کرد.  
و - فرضیه نامناسب است؛ چون با دانش موجود در مورد تقویت مثبت و اثرات آن در پاسخهای دانش‌آموز، متضاد است.

۶ - فرضیه جهت‌دار: پسران زودرس، خودانگاره مثبت‌تری از پسران دیررس نشان می‌دهند. فرضیه غیرجهت‌دار: بین خودانگاره نوجوان پسر زودرس و دیررس تفاوت وجود دارد.

۷ - فرضیه به گردآوری و تفسیر داده‌ها جهت می‌دهد. بیان روشن فرضیه می‌تواند نقطه ضعفهایی را نشان دهد که هنگام تکوین ایده گنگ بررسی در ذهن، آشکار نیست.

۸ - الف - تحقیق

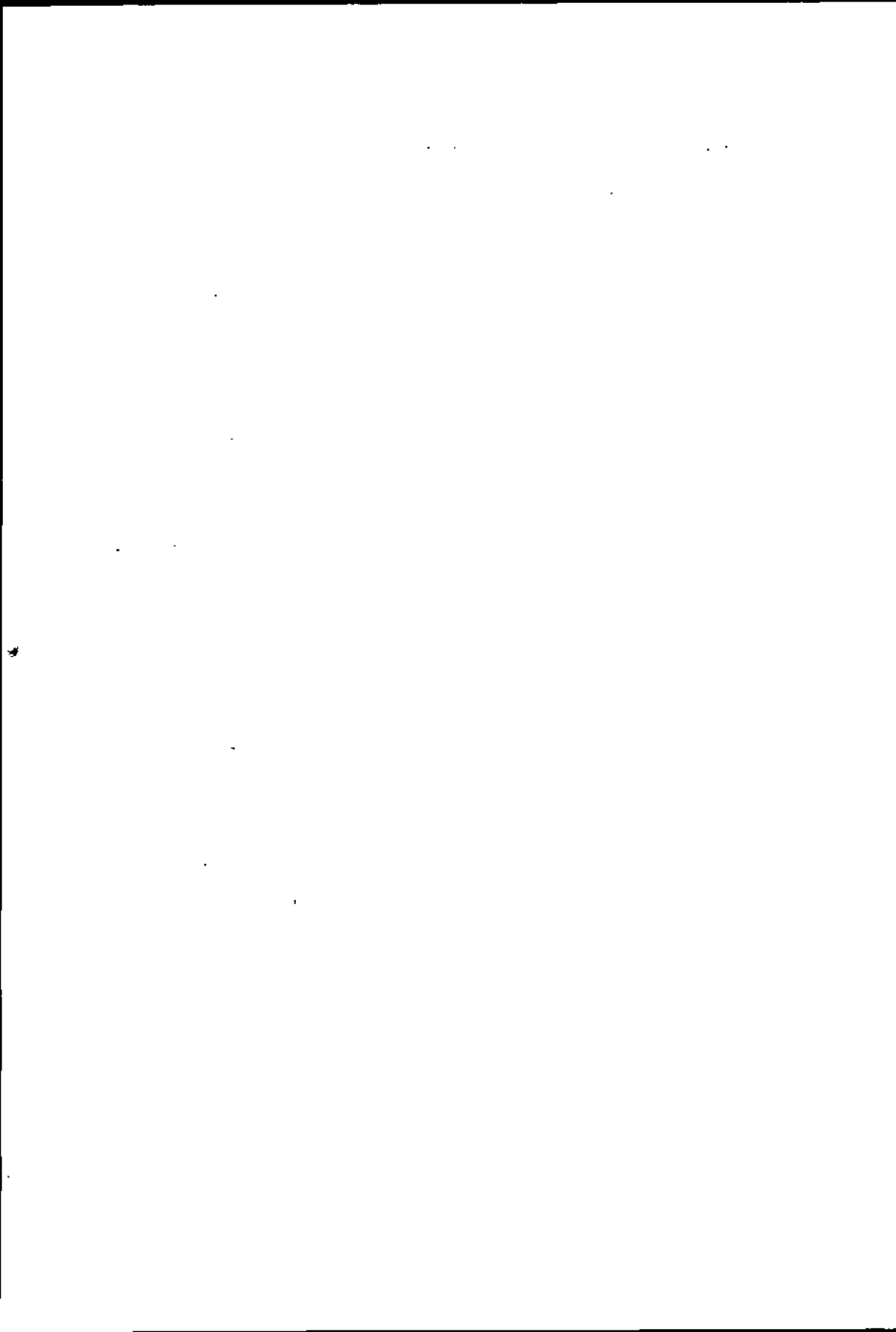
ب - صفر

ج - صفر

د - تحقیق

۹ - پاسخها می‌توانند متعدد باشند.

۱۰ - فرضیه در یک مدت زمان منطقی قابل آزمون نیست. بعلاوه، تعریف و اندازه‌گیری اصطلاح شهروند بزرگسال بهتر، مشکل خواهد بود.



# بخش سوم

## تحلیل آماری

### فصل پنجم

#### آمار توصیفی

##### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- ویژگیها و محدودیتهای چهار نوع مقیاس اندازه گیری: اسمی، رتبه ای، فاصله ای و نسبی را تشخیص دهد.
- ۲- داده های تحقیق را به صورت توزیع فراوانی طبقه بندی کند و آنها را به شکل نمودار چند ضلعی فراوانی و نمودار ستونی ارائه کند.
- ۳- تفاوت اندازه های گرایش مرکزی را تشخیص دهد و موقعیتهای قابل استفاده از هر یک را بازشناسد و برای هر داده مفروض میانگین، میانه و نما را محاسبه کند.
- ۴- کاربردهای مناسب اندازه های تغییرپذیری را توصیف کند و برای هر مجموعه از داده های مفروض واریانس، انحراف معیار، انحراف چارکی و دامنه را محاسبه کند.
- ۵- نمره های خام را به نمره های استاندارد تبدیل و جایگاه نسبی نمره  $Z$  را در یک منحنی نرمال تعیین کند.
- ۶- کاربردهای مناسب شاخصهای مختلف همبستگی را برای توصیف رابطه بین متغیرها بشناسد.

- ۷- ضریب همبستگی بین دو مجموعه از داده‌های فاصله‌ای یا نسبی را محاسبه کند.
- ۸- ضریب همبستگی بین دو مجموعه از داده‌ها را که به صورت رتبه‌ای منظم شده‌اند، محاسبه کند.
- ۹- معنای ضریب تعیین و کاربرد آن را در تفسیر ضریب همبستگی توصیف کند.
- ۱۰- معنا، ویژگیها و کاربردهای فرا تحلیل<sup>(۱)</sup> را بشناسد.

شیوه‌های آماری، اساساً روشهایی هستند که اطلاعات کمی را به گونه‌ای به کار می‌گیرند که آن اطلاعات معنادار می‌شوند. این شیوه‌ها دو مزیت اصلی دارند. نخست آنکه این شیوه‌ها ما را قادر می‌سازند که مشاهده‌های خود را طبقه‌بندی، توصیف و خلاصه کنیم. دوم، این شیوه‌ها به ما کمک می‌کنند تا دریابیم که با چه اعتباری می‌توان نتیجه‌گیری کرد که پدیده مشاهده شده در یک گروه محدود، یعنی نمونه‌ای در جامعه بزرگ مورد نظر که نمونه از آن انتخاب شده است، به وقوع خواهد پیوست - به سخن دیگر، ما با چه دقتی می‌توانیم از طریق استدلال استقرایی، نتیجه بگیریم که آنچه در جزء مشاهده کرده‌ایم، در کل نیز مشاهده خواهیم کرد. ما برای مسائلی که چنین ماهیتی دارند، نیازمند به کارگیری آمار استنباطی هستیم.

شناخت برخی از شیوه‌های آماری برای کسانی که به کار پژوهش می‌پردازند، از آنرو اساسی است که آنها می‌توانند داده‌های خود را تحلیل و تفسیر کنند و از طریق یافته‌های خود با دیگران ارتباط برقرار نمایند. بعلاوه، شایسته است که کارشناسان آموزشی، که نیازمند آگاهی از امر پژوهش و استفاده از یافته‌های پژوهشی هستند، به منظور درک و ارزشیابی مطالعات پژوهشی دیگران، با شیوه‌های آماری آشنا باشند. همچنین آگاهی از شیوه‌های آماری برای اجرا و تفسیر مناسب آزمونهایی که در مدارس ما معمول است، ضرورت دارد. آموزگاران که با این شیوه‌ها ناآشنا هستند، ممکن است در ارزشیابی تواناییها و پیشرفت دانش‌آموزان خود با اشکال مواجه شوند. همچنین، آنها بررسی پژوهشهای حیطة تخصصی خود و کسب اطلاعات جدید را نیز دشوار خواهند یافت.

### /// مقیاسهای اندازه‌گیری

یک گام اساسی در انجام تحقیق، اندازه‌گیری است؛ فرآیندی که از طریق آن مشاهده‌ها به اعداد برگردانده می‌شوند. اس. اس. استیونس گفته است، «اندازه‌گیری در گسترده‌ترین معنای آن، تخصیص اعداد به اشیاء یا رویدادها، بنابر قانونمندیهاست.»<sup>(۱)</sup> محققان با متغیرها آغاز می‌کنند و سپس برای تعیین اینکه این متغیرها چگونه به شکل عددی در می‌آیند، از قوانین استفاده می‌کنند. متغیر رحجان مذهبی را می‌توان بر مبنای شمارش پاسخهای دانشجویان به پرسشی که

1- Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens, (ed.). Handbook of Experimental psychology (p.1). New York: Wiley.

از آنها می‌خواهد تا از میان چهار گزینه (۱) کاتولیک، (۲) یهودی، (۳) پروتستان، یا (۴) دیگر مذاهب، یکی را انتخاب کنند، اندازه‌گیری کرد. متغیر وزن را می‌توان با مشاهده ارقامی که با گام نهادن آزمودنیها بر روی ترازو به دست می‌آید، اندازه‌گیری کرد. متغیر بلوغ اجتماعی را می‌توان بر مبنای نمره‌هایی که از مقیاس بلوغ اجتماعی واینلند<sup>(۱)</sup> به دست می‌آید، اندازه‌گیری کرد.

ماهیت فرآیند اندازه‌گیری که اعداد را خلق می‌کند، تفسیری را که می‌توان از آنها به عمل آورد و شیوه‌های آماری را که می‌توان به طور معنی‌دار در مورد آنها به کار برد، تعیین می‌کند. رایج‌ترین طبقه‌بندی شیوه‌های اندازه‌گیری، مقیاسهای اندازه‌گیری استیونس است، که اندازه‌گیری را به انواع اسمی، رتبه‌ای، فاصله‌ای و نسبی دسته‌بندی کرده است.

### // مقیاس اسمی

ابتدایی‌ترین مقیاس اندازه‌گیری، مقیاس اسمی است. اندازه‌گیری اسمی مستلزم جای‌دادن اشیاء یا افراد در طبقاتی است که نه از نظر کمی، بلکه از نظر کیفی با هم متفاوت هستند. اندازه‌گیری این سطح، تنها نیازمند آن است که هر فرد بتواند دو یا چند طبقه مرتبط را از هم تشخیص دهد و معیارهای لازم را برای جای‌دادن افراد یا اشیاء در این یا آن طبقه بشناسد. سطح فعالیت تجربی ضروری، مستلزم تشخیص این است که آیا یک شیء یا فرد فرضی به طبقه نامتناخل (مانعة‌الجمع) تعلق دارد یا ندارد. تنها رابطه‌ی مابین طبقات، این است که آنها با یکدیگر متفاوت هستند، و هیچ دلالتی وجود ندارد که این طبقات گویای «بیشتر» یا «کمتر» بودن ویژگی اندازه‌گیری شده باشند. دسته‌بندی کردن دانشجویان برحسب جنس، یک اندازه‌گیری اسمی خواهد بود.

در سطح اندازه‌گیری اسمی، اعداد، اغلب مورد استفاده قرار می‌گیرند، اما این امر تنها برای مشخص کردن طبقات صورت می‌گیرد. اعداد، که برحسب سلیقه به طبقات اختصاص می‌یابند، صرفاً به عنوان یک برحسب عمل می‌کنند. به همهٔ اعضای یک طبقه‌بندی عدد یکسانی تخصیص داده می‌شود و هرگز به دو طبقه‌بندی، عددی یکسان تخصیص نمی‌یابد. برای مثال، در آماده‌سازی داده‌ها برای کامپیوتر، می‌توان رقم صفر را برای نمایاندن مردان و رقم ۱ را برای نمایاندن زنان مورد استفاده قرار داد. مابین اعدادی که برای اندازه‌گیری اسمی به کار می‌روند،

هیچ رابطه تجربی‌ای که با روابط ریاضی بین اعداد مطابقت کند، وجود ندارد. عدد ۱ گویای آن نیست که چیزی بیشتر از عدد صفر دارد. اعداد می‌توانند با هم جابجا شوند، بدون اینکه برجیزی به جز طرح نامگذاری مورد استفاده، تأثیر بگذارند.

اعداد بکار رفته در یک مقیاس اسمی، معرف مقدار مطلق یا نسبی هیچ ویژگی‌ای نیستند. آنها صرفاً برای شناساندن اعضای یک طبقه‌بندی فرضی عمل می‌کنند. برای مثال، اعداد تخصیص یافته به بازیکنان فوتبال، یک مقیاس اسمی تشکیل می‌دهند. ما نمی‌خواهیم بگوییم بازیکنی که شماره ۴۸ را روی پیراهن خود دارد، لزوماً بهتر از بازیکن شماره ۳۶ است. همچنین نمی‌توانیم بگوییم که بین بازیکنان شماره ۴۰ و ۴۸ همان تفاوتی در توانایی بازی وجود دارد که بین بازیکنان شماره ۵۰ و ۵۸.

البته، اعداد شناسایی در یک مقیاس اسمی هرگز نمی‌توانند از نظر محاسباتی، از طریق جمع، منها، ضرب، یا تقسیم دستکاری شوند. تنها می‌توان شیوه‌های آماری مبتنی بر شمارش صرف، نظیر گزارش تعداد مشاهده‌ها در هر طبقه‌بندی را مورد استفاده قرار داد.

### // مقیاس رتبه‌ای

دومین مقیاس مهم اندازه‌گیری، مقیاس رتبه‌ای است، که در آن با در نظر گرفتن یک ویژگی، موقعیت نسبی اشیاء یا افراد، بدون نشان دادن فاصله بین آن موقعیتها، مشخص می‌شود. برای اندازه‌گیری در این سطح، یک معیار تجربی برای نظم و ترتیب دادن به اشیاء یا رویدادها، با توجه به ویژگی مورد نظر، ضرورتی اساسی دارد، یعنی به شیوه‌ای نیاز است که ترتیب افراد یا اشیاء را بر حسب دارا بودن ویژگی مورد نظر تعیین کند. برای مثال، هنگامی که آموزگاران دانش‌آموزان خود را بر مبنای برخی ویژگیها، نظیر بلوغ اجتماعی، تواناییهای رهبری، حس همکاری و غیره رتبه‌بندی می‌کنند، یک اندازه‌گیری رتبه‌ای صورت پذیرفته است. دانش‌آموزان اغلب بر حسب پیشرفت تحصیلی خود یا عملکردشان در یک مسابقه ورزشی یا رقابت در زمینه موسیقی رتبه‌بندی می‌شوند.

در اندازه‌گیری رتبه‌ای، شیوه تجربی مورد استفاده برای مرتب کردن اشیاء باید معیاری به نام اصل رابطه غیر مستقیم<sup>(۱)</sup> را داشته باشد. این اصل چنین نوشته می‌شود: اگر  $(a > b)$  و  $(b > c)$ ،

پس ( $a > c$ )، و معنای آن این است که رابطه باید به گونه‌ای باشد که اگر شیء  $a$  بزرگتر از شیء  $b$  است، و شیء  $b$  بزرگتر از شیء  $c$  است، پس شیء  $a$  بزرگتر از شیء  $c$  خواهد بود. البته، ممکن است کلمات دیگری جانشین بزرگتر از شوند، این کلمات می‌توانند شامل قوی‌تر از، جلوتر از، از یک ویژگی مقدار بیشتری دارد، و غیره باشند.

عملکرد تجربی در اندازه‌گیری رتبه‌ای تنها شامل مقایسه مستقیم اشیاء یا افراد برحسب مقدار ویژگی مورد نظر در آنهاست. بنابراین، هنگامی که اعداد به اشیاء تخصیص می‌یابند، تنها اطلاعاتی که مورد توجه قرار می‌گیرد، ترتیب اشیاء است. از اینرو، تنها ویژگی اعداد، که دارای معناست، ترتیب آنهاست. اعداد تخصیص یافته در اندازه‌گیری رتبه‌ای تنها مرتبه جایگاه را نشان می‌دهند و دیگر هیچ. نه تفاوت بین اعداد و نه نسبت آنها، معنایی ندارند. بکارگیری اعداد ۱ و ۲ و ۳ در اندازه‌گیری رتبه‌ای به این معنا نیست که رتبه ۱ همان قدر بالاتر از رتبه ۲ است که رتبه ۲ از رتبه ۳ و الی آخر. فاصله بین کودک رتبه ۱ و کودک رتبه ۲ ممکن است با فاصله دو کودک رتبه ۲ و ۳، یکی باشد، کمتر باشد و یا بیشتر باشد. بسادگی باید گفت که هیچ مبنایی برای تفسیر اندازه تفاوت بین اعداد یا نسبت اعداد وجود ندارد. در یک مسابقه دو زمان‌گیری نشده، ما می‌دانیم که چه کسانی اول، دوم، سوم و... به خط پایان رسیده‌اند، اما نخواهیم دانست که یک دوندۀ چقدر از دوندۀ دیگر سریعتر دویده است. تفاوت بین اول و دوم لزوماً با تفاوت بین دوم و سوم و یا سوم و چهارم یکسان نخواهد بود. همچنین نمی‌توانیم بگوییم که دوندۀ ای که به عنوان نفر دوم به خط پایان رسیده، دو بار سریعتر از دوندۀ چهارم بوده است.

مقیاس سختی مواد معدنی (کانیها) یک مثال خوب برای مقیاس رتبه‌ای است. مواد معدنی برحسب توانایی آنها در خراشیدن یکدیگر رده‌بندی می‌شوند. اگر ماده معدنی  $A$  بتواند بر ماده معدنی  $B$  خراش بیندازد، گفته می‌شود که ماده معدنی  $A$  سخت‌تر از ماده معدنی  $B$  است. بر این مبنا، الماس به عنوان سخت‌ترین ماده معدنی رده‌بندی می‌شود؛ زیرا الماس می‌تواند همه مواد معدنی شناخته شده دیگر را بخراشد، اما هیچ ماده معدنی نمی‌تواند بر آن خراش بیندازد. مجموعه‌ای از ده ماده معدنی که دامنه آنها از نرمترین تا سخت‌ترین است، به عنوان استاندارد (معیار - ملاک - استاندارد) انتخاب شده و اعداد ۱ تا ۱۰ به آنها تخصیص یافته است، که ۱ نشانه نرمترین و ۱۰ نشانه سخت‌ترین ماده معدنی است. به مواد معدنی دیگر نیز بر اساس آزمایش خراش‌پذیری، اعدادی تخصیص یافته است. بنابراین، ما ترتیب سختی مواد معدنی را می‌دانیم،



ولی نمی‌دانیم یک ماده معدنی چقدر از دیگری سخت‌تر است. ما نمی‌توانیم فرض کنیم که یک ماده معدنی که ارزش ۴ به آن تخصیص یافته است، دو بار سخت‌تر از ماده‌ای با ارزش ۲ است، یا تفاوت سختی بین مواد معدنی ۲ و ۴ به اندازه تفاوت سختی بین موادی با ارزشهای ۳ و ۴ است. آماره‌های مناسب برای یک مقیاس رتبه‌ای محدود هستند. از آنجا که اندازه فاصله بین طبقات نامشخص است، نمی‌توانیم از هیچ شیوه آماری‌ای که فرض آن فواصل مساوی است، استفاده کنیم. آماره‌هایی که نشان‌دهنده نقاطی هستند که درصدهای معینی از موارد در زیر آن نقاط قرار می‌گیرند، برای یک مقیاس رتبه‌ای مناسب هستند.

### // مقیاس فاصله‌ای

مقیاس فاصله‌ای مقیاسی است که فواصل مساوی از یک مبدا اختیاری را تأمین می‌کند. یک مقیاس فاصله‌ای نه تنها اشیاء یا رویدادها را برحسب مقدار ویژگی‌ای که دارا هستند، رتبه‌بندی می‌کند، بلکه همچنین، بین واحدهای اندازه‌گیری، فواصل مساوی بنا می‌نهد. تفاوت‌های مساوی در اعداد، معرف تفاوت‌های مساوی در ویژگی اندازه‌گیری شده است. دماسنج‌های فارنهایت و سانتیگراد نمونه‌هایی از مقیاس فاصله‌ای هستند.

در یک مقیاس فاصله‌ای، هم رابطه‌های ترتیبی و هم رابطه‌های فاصله‌ای بین اعداد، دارای معنا هستند. ما می‌توانیم بگوییم که تفاوت بین ۵۰ و ۵۱ درجه سانتیگراد با تفاوت بین ۳۰ و ۳۱ درجه سانتیگراد برابر است، اما به هر روی نمی‌توانیم بگوییم که ۵۰ درجه دو بار گرمتر از ۲۵ درجه است. این امر به دلیل آن است که در مقیاس فاصله‌ای نقطه صفر واقعی وجود ندارد. نقطه صفر قراردادی است، مثل مقیاس سانتیگراد، که ارزش صفر درجه را برای نقطه انجماد آب تعیین کرده است.

همچنین، در یک آزمون روان‌شناختی یا آموزشی، نقطه صفر اختیاری است. برای مثال، هوش صفر وجود ندارد. در آزمون‌های هوش استاندارد شده ما هیچ راهی برای شناسایی فرد دارای هوش صفر وجود ندارد. گاهی یک دانشجو ممکن است در آزمون آمار نمره صفر بگیرد، اما این به معنای آن نیست که دانش آماری او صفر است. اگر ما سه دانشجو داشته باشیم که در یک آزمون آمار نمره‌های ۱۵، ۳۰ و ۴۵ گرفته باشند، نمی‌توانیم بگوییم که نمره ۳۰ معرف دو برابر دانش آماری نمره ۱۵ یا نمره ۴۵ معرف سه برابر دانش آماری نمره ۱۵ است. برای درک این

مطلب، اجازه دهید فرض کنیم ۱۵ پرسش بسیار ساده به آزمون اضافه می‌شود؛ به گونه‌ای که هر سه دانشجو قادرند آنها را بدرستی پاسخ گویند. اکنون سه نمره قبلی به ۳۰، ۴۵ و ۶۰ تبدیل خواهد شد. اگر تلاش کنیم در این مقیاس از نوع فاصله‌ای، نسبت‌هایی بین نمره‌ها بسازیم، به اشتباه گزارش خواهیم کرد دانشجویی که نمره ۶۰ گرفته، دو برابر دانشجویی که نمره ۳۰ گرفته است، دانش آماری دارد، در حالی که در نسبت قبلی ما به خطا فرض کردیم که همان دانشجو سه برابر دانشجوی دیگر آمار می‌داند.

بنابراین، از آنجا که صفر اختیاری است، ضرب و تقسیم اعداد درست نیست و همان‌گونه که دیدیم نسبت‌های بین اعداد در یک مقیاس فاصله‌ای بی‌معنا هستند. به هر حال، در یک مقیاس فاصله‌ای می‌توان تفاوت بین جایگاهها (موقعیتها)<sup>(۱)</sup> را گزارش داد یا اعداد را با هم جمع کرد. همه شیوه‌های آماری مبتنی بر عمل جمع و شیوه‌های آماری مناسب برای مقیاس‌های سطح پایین‌تر را می‌توان در این مقیاس مورد استفاده قرار داد. این امر اکثر شیوه‌های رایج آماری را در بر می‌گیرد.

نکتهٔ حائز اهمیت آن است که در موارد بسیاری که از مقیاس‌های فاصله‌ای استفاده می‌کنیم، فواصل مقیاس از لحاظ خود ابزار اندازه‌گیری، مساوی هستند، اما لزوماً از لحاظ توانایی مورد اندازه‌گیری، مساوی نیستند. برای روشن شدن موضوع، یک آزمون هجی کردن را با این کلمات در نظر بگیرید: گربه، بشقاب، توپ، پرحرف، اسکیزوفرنی و سینه پهلوی. در اینجا فاصله بین یک پاسخ درست و سه پاسخ درست با فاصله بین سه پاسخ درست و پنج پاسخ درست همسان است. به هر حال، هنگامی که توانایی هجی کردن را در نظر بگیریم، تفاوت بین ۳ و ۵ پاسخ درست در مقایسه با ۱ و ۳ پاسخ درست گویای توانایی بیشتری است؛ مگر اینکه کسی بتواند بگوید که در آزمون هجی کردن، فاصله بین ۳ و ۵ و فاصله بین ۱ و ۳ توانایی هجی کردن را به یک میزان نشان می‌دهند، که در این صورت این نمره‌ها رتبه‌بندی دانشجویان را بیان می‌کنند.

با این همه، با دقت و تلاش زیاد می‌توان آزمونی تهیه کرد که فواصل بین نمرات آن یک تخمین منطقی از فاصله بین تواناییها باشد. آزمونهای بهترهوش، شاهد این گفته هستند. تفاوت در توانایی بین هوشبهر (IQ) ۹۰ و هوشبهر (IQ) ۹۵، شاید دقیقاً همان تفاوت بین هوشبهر (IQ) ۱۰۵ و ۱۱۰ نباشد، اما اگر هم فرض کنیم که این دو تفاوت تقریباً یکسان هستند، خیلی به خطا نرفته‌ایم.

## // مقیاس نسبی

یک مقیاس نسبی، یعنی بالاترین سطح اندازه‌گیری، مقیاسی است که هم نقطه صفر واقعی و هم فواصل مساوی را تأمین می‌کند. در این مقیاس، بین هر دو مقدار فرضی می‌توان نسبتها را تشکیل داد. خط‌کشی که برای اندازه‌گیری طول در واحد اینچ یا فوت به کار می‌رود، یک مقیاس نسبی است؛ زیرا مبدأ مقیاس یک صفر مطلق و مطابق با طول صفر است. بنابراین، می‌توانیم بگوییم که طول یک تکه چوب ۶ فوتی دو برابر تکه چوب ۳ فوتی است. در یک مقیاس نسبی، این امکان وجود دارد که هر یک از مقادیر را در یک عدد معین ضرب یا تقسیم کنیم، بدون آنکه خصوصیات مقیاس تغییر کند. برای مثال، می‌توانیم ۲ پوند را در ۱۶ ضرب کنیم، برای آنکه واحد اندازه‌گیری را به ۳۲ اونس تبدیل کنیم یا می‌توانیم ۶ فوت را در ۱۲ ضرب کنیم، برای آنکه واحد را به اینچ تبدیل کنیم. ما می‌توانیم عمل ضرب را انجام دهیم و همان نسبتهای قبل از ضرب را حفظ کنیم. برای مثال، ما می‌توانیم ۴ کوارت<sup>(۱)</sup> شیر و ۲ کوارت شیر را در ۲ ضرب و واحد اندازه‌گیری را به پینت<sup>(۲)</sup> تبدیل کنیم. در واحد پینت هنوز ۸ پینت شیر دو برابر ۴ پینت است.

در تعلیم و تربیت تنها چند متغیر مورد نظر، ماهیت نسبی دارند. این متغیرها بیشتر به عملکرد حرکتی و دیگر اندازه‌های فیزیولوژیکی منحصر هستند. با آنکه ما می‌توانیم بگوییم که قد شخصی با ۶ فوت دو برابر قد شخص دیگری با ۳ فوت است، زیرا ارتفاع، یک داده نسبی است، اما نمی‌توانیم بگوییم که شخصی با هوشبهر (IQ) ۱۵۰ دو برابر شخص دیگری با هوشبهر (IQ) ۷۵، باهوشتر است؛ زیرا نمره‌های هوشبهر (IQ) داده‌های فاصله‌ای هستند. همه انواع شیوه‌های آماری برای یک مقیاس نسبی مناسب هستند.

## // سازماندهی داده‌های پژوهش

توصیف داده‌هایی که به نوعی مرتب شده باشند، اگر ناممکن نباشد، بسیار مشکل است. بنابراین، سازمان دادن داده‌های پژوهش در آمار توصیفی یک گام اساسی است. دو روشی که اغلب در سازمان دادن داده‌ها مورد استفاده قرار می‌گیرند، عبارتند از (۱) مرتب کردن اندازه‌ها به صورت توزیع فراوانی و (۲) ارائه آنها به شکل نگاره‌ای.

## // توزیع فراوانی

آرایش نظام‌مند اندازه‌های فردی، از کمترین تا بالاترین آنها، توزیع فراوانی نامیده می‌شود. استفاده از این روش صرفاً مستلزم ساخت فهرستی از اندازه‌های فردی به صورت ستونی است که بالاترین اندازه در بالای ستون و پایین‌ترین اندازه در انتهای ستون قرار گرفته باشد. اغلب دیده می‌شود که نمرات یکسان متعددی در یک توزیع وجود دارد. به جای آنکه این نمرات جداگانه فهرست شوند، مرسوم است که ستون دومی اضافه شود و در آن فراوانی هر یک از اندازه‌ها ثبت شود. در جدول ۵-۱، نمرات ۱۰۵ دانش‌آموز در آزمون آمار نشان داده شده است. در قسمت الف این جدول، نمرات به شکل سازمان‌یافته فهرست شده‌اند. در قسمت ب همان نمرات بعد از مرتب شدن در یک توزیع فراوانی ارائه شده‌اند.

شکل کلی یک توزیع را می‌توان از توزیع فراوانی آن دریافت. با سازمان دادن نمرات در یک توزیع فراوانی، می‌توان گستردگی آنها را تعیین کرد که آیا این نمرات به طور متعادل توزیع شده‌اند یا گرایش به تجمع دارند و این تجمع‌ها در چه نقاطی از توزیع صورت می‌پذیرند. برای مثال، با نگاهی به توزیع فراوانی نمرات جدول ۵-۱، بسادگی روشن می‌شود که محدوده این نمرات از ۲۱ تا ۳۵ است، و نمره ۲۹ بیشتر از دیگر نمره‌ها تکرار شده است، و این نمرات گرایش دارند که در بالای توزیع تجمع یابند، تا در پایین آن. اگر نمره‌ها سازمان داده نشده بودند، هیچ یک از این موارد آشکار نمی‌شد. سازماندهی داده‌ها در توزیعهای فراوانی، همچنین، محاسبه آماره‌های مفید مختلف را تسهیل می‌کند.

## // ارائه نگاره‌ای داده‌ها

اغلب ارائه داده‌های پژوهش به شکل نگاره‌ای مفید و مناسب است. در میان انواع مختلف نگاره‌ها، نمودار ستونی و نمودار چندضلعی فراوانی، بیشترین مورد استفاده را دارند.

مراحل اساسی ساخت نمودار ستونی و نمودار چند ضلعی فراوانی یکسان است:

۱- بر روی یک محور افقی (محور طولی)، از چپ به راست، از پایین‌ترین مقدار تا بالاترین مقدار، نقاطی را برای نمره‌ها مشخص کنید. در دو انتهای توزیع، فضای کافی برای یک نمره اضافی در نظر بگیرید.

۲- بر روی محور عمودی، فراوانی نمره‌ها یا فاصله‌ها را مشخص کنید. از صفر شروع کنید و رو به بالا ادامه دهید.

۳- بر بالای مرکز هر نمره در سطح فراوانی آن یک نقطه بگذارید.

از این قسمت به بعد می‌توان هر دو نمودار ستونی یا چند ضلعی را ساخت. برای ساخت نمودار ستونی، همانطور که در شکل ۵-۱ نشان داده شده است، باید بر هر نقطه، خطی افقی به پهنای هر نمره کشید.

برای ساخت نمودار چند ضلعی، نقاط مجاور به هم متصل می‌شوند و دو انتهای شکل به دست آمده نیز در نقاطی که یکی کمتر از پایین‌ترین نمره و یکی بیشتر از بالاترین نمره باشد، به محور افقی (خط صفر)، همانطور که در شکل ۵-۲ نشان داده شده است، وصل می‌شوند.

### /// اندازه‌های گرایش مرکزی

یکی از راههای ساده خلاصه کردن داده‌ها، یافتن یک شاخص منفرد است که بتواند معرف مجموعه‌ای از اندازه‌ها باشد. برای مثال، یافتن یک نمره منفرد، که بتواند نشانه‌ای از عملکرد یک گروه ۳۰۰ نفری دانش‌آموزان را در آزمون IQ به دست دهد، برای مقاصد مقایسه‌ای مفید خواهد

جدول شماره ۱-۵ نمرات ۱۰۵ دانش‌آموز در آزمون آمار

#### الف. نمرات سازمان نیافته

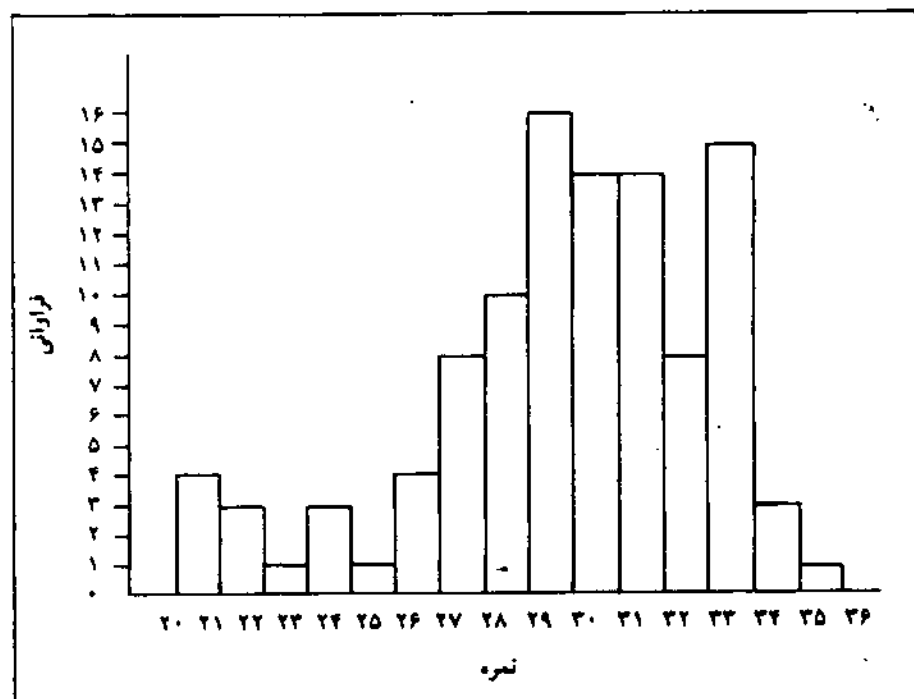
۲۲	۲۹	۳۰	۳۰	۳۳	۲۹	۳۳	۳۲	۲۸	۲۴	۲۴	۳۱	۲۷	۲۹	۲۲
۲۵	۲۹	۲۴	۲۷	۲۶	۳۳	۳۳	۲۶	۳۰	۲۸	۲۶	۲۹	۳۲	۳۲	۳۱
۲۸	۳۴	۳۰	۳۱	۳۲	۲۱	۲۹	۳۱	۳۰	۳۲	۳۵	۳۰	۳۱	۲۷	۲۹
۲۶	۲۹	۳۳	۳۲	۲۹	۲۸	۲۸	۳۰	۲۸	۲۷	۳۰	۳۱	۳۴	۳۳	۲۲
۳۰	۲۹	۲۷	۲۹	۲۴	۳۰	۲۱	۳۱	۳۱	۳۳	۲۸	۲۱	۳۱	۲۹	۳۱
۳۱	۳۳	۲۲	۲۹	۳۱	۳۲	۳۲	۳۱	۲۸	۲۹	۳۰	۲۲	۳۳	۳۰	۳۰
۳۲	۳۳	۳۱	۳۳	۲۸	۲۹	۲۷	۳۳	۲۷	۲۱	۳۰	۲۹	۲۸	۲۷	۳۳

#### ب. توزیع فراوانی

نمرات (X)	خط نشانها	فراوانی (f)
۲۵	I	۱
۳۴	III	۳
۳۳	III III III	۱۵
۳۲	III III	۸
۳۱	III III III	۱۴
۳۰	III III III	۱۴
۲۹	III III III	۱۶
۲۸	III III	۱۰
۲۷	III III	۸
۲۶	III	۴
۲۵	I	۱
۲۴	III	۳
۲۳	I	۱
۲۲	III	۳
۲۱	III	۴
		N = ۱۰۵

بود. در آمار، سه شاخص برای چنین منظوری در دسترس است. این شاخصها اندازه‌های گرایش مرکزی یا متوسطها خوانده می‌شوند. برای بسیاری از مردم عادی، معنای واژه متوسط، جمع نمره‌ها و تقسیم آن بر تعداد نمره‌هاست. برای یک آمارشناس، واژه متوسط همین معنا را دارد، که آن را میانگین، یا یکی از دو اندازه دیگر گرایش مرکزی، یعنی نما یا میانه می‌نامند. هر یک از این سه می‌توانند به عنوان شاخصی که نماینده یک گروه به صورت یک کل باشند، عمل کنند.

شکل ۱-۵ نمودار ستونی ۱۰۵ نمره درس آمار از جدول ۱-۵



// نما

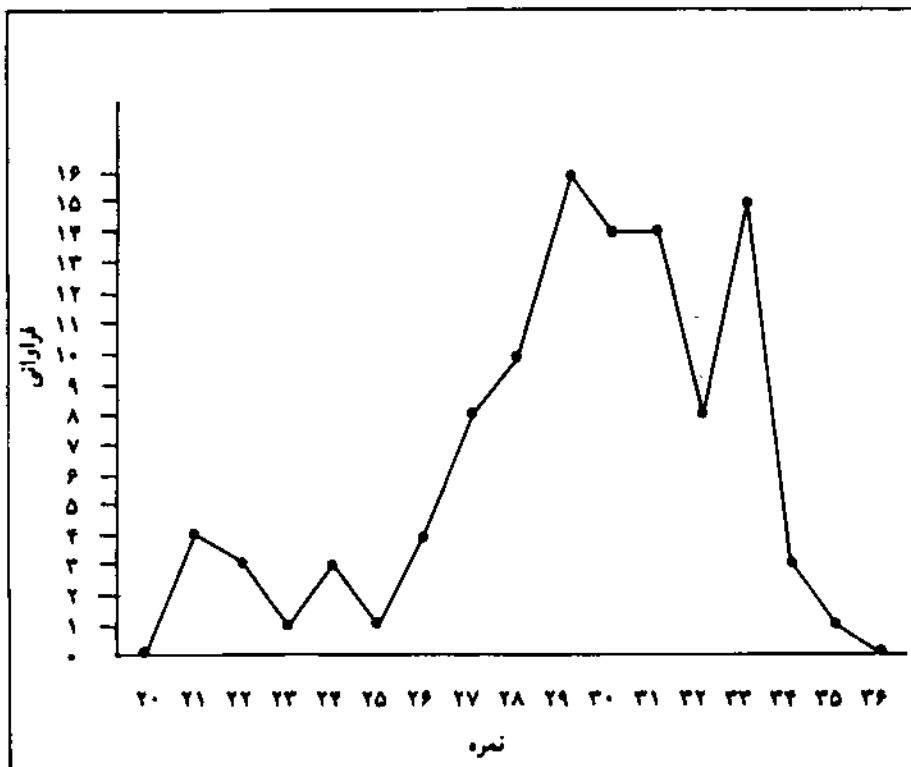
در یک توزیع، نما مقداری است که بیشترین تکرار را داراست. از میان سه اندازه گرایش مرکزی، یافتن نما ساده‌تر از دو اندازه دیگر است، زیرا استفاده از آن تنها به منظور بازیابی صورت می‌گیرد. نما را می‌توان برای هر یک از مقیاسهای اندازه‌گیری گزارش کرد، اما تنها اندازه گرایش مرکزی است که کاربرد آن در مقیاس اسمی درست است.

// میانه

میانه در یک توزیع به نقطه‌ای گفته می‌شود که ۵۰ درصد موارد در زیر آن قرار گیرند. (به این معنا که ۵۰ درصد دیگر در بالای آن نقطه قرار می‌گیرند). برای مثال، اگر توزیع نمره‌های زیر را داشته باشیم:

۱۴    ۱۶    ۱۶    ۱۷    ۱۸    ۱۹    ۱۹    ۱۹    ۲۱    ۲۲

شکل ۲-۵ نمودار چند ضلعی ۱۰۵ نمره درس آمار از جدول ۱-۵

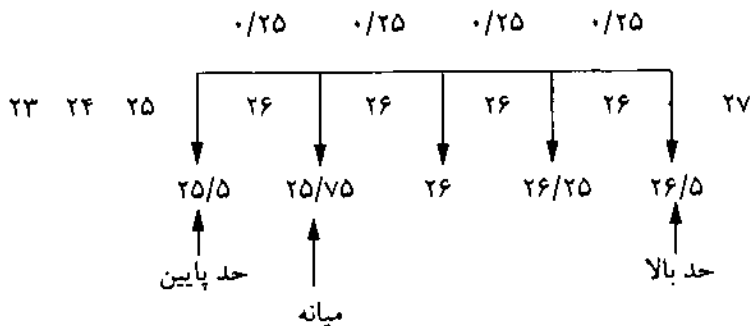


نقطه‌ای که ۵۰ درصد از موارد در زیر آن قرار می‌گیرند، در وسط ۱۸ و ۱۹ واقع شده است. بنابراین، میانه این توزیع ۱۸/۵ است. توجه کنید که برای یافتن این مقدار، ده نمره این توزیع ابتدا از کم به زیاد مرتب شده‌اند و سپس نقطه‌ای که نیمی از نمره‌ها زیر آن قرار می‌گیرند، پیدا شده است. این نقطه (۱۸/۵)، که دو مقدار ۱۸ و ۱۹ را دقیقاً جدا می‌کند، در الفبای آماری، حد بالای نمره ۱۸ و حد پایین نمره ۱۹ خوانده می‌شود. در محاسبه میانه، چنین تصور می‌شود که هر نمره، معرف دامنه یا فاصله‌ای است که از نیمه بین آن نمره و نمره پایین‌تر از خود تا نیمه بین آن نمره و نمره بالاتر از خود امتداد دارد. بنابراین، در این مثال، تصور بر آن است که ۱۸ معرف فاصله از ۱۷/۵ تا ۱۸/۵ است، در حالی که ۱۹ معرف فاصله از ۱۸/۵ تا ۱۹/۵ است.

نکته مهم آنکه میانه همیشه در خط مرزی بین دو مقدار قرار نمی‌گیرد. در واقع، میانه، اغلب در مکانی بین حد بالا و حد پایین یک فاصله جای می‌گیرد. به منظور تعیین میانه، ما باید یک نمره معرف دامنه حد پایین و بالای آن تصور کنیم، نه به صورت یک نقطه منفرد. مثال زیر را در نظر بگیرید:

۲۳      ۲۴      ۲۵      ۲۶      ۲۶      ۲۶      ۲۶      ۲۷

در این توزیع، میانه، بین دو نمره ۲۶ قرار می‌گیرد. به فراوانی نمره ۲۶ توجه کنید. در توزیع، چهار نمره ۲۶ وجود دارد، که یکی از آنها زیر نقطه میانی و سه عدد دیگر بالای نقطه میانی جای می‌گیرند. در این حالت برای یافتن میانه ما باید فاصله ۲۶/۵ و ۲۵/۵ را به چهار بخش تقسیم کنیم. پس فاصله بین هر یک از این چهار نمره، ۰/۲۵ خواهد بود. بنابراین تصور می‌شود که هر یک از این نمره‌ها معرف دامنه‌ای است که ۰/۲۵ از فاصله بین ۲۵/۵ و ۲۶/۵ را که حد پایین و بالای نمره ۲۶ هستند، می‌پوشاند. این مفهوم در زیر به تصویر در آمده است و موقعیت و مقدار میانه را نشان می‌دهد.





استفاده از یک فرمول برای یافتن میانه، ساده‌تر است تا اینکه هر بار چنین فرآیندی را طی

کنیم.

$$Md = L + \left( \frac{\frac{N}{2} - cfb}{fw} \right) i \quad (5-1)$$

$Md$  = میانه

$L$  = حد پایین فاصله‌ای که میانه در آن قرار می‌گیرد

$N$  = تعداد موارد موجود در توزیع

$cfb$  = کل فراوانی در همه فاصله‌هایی که زیر فاصله حاوی میانه قرار دارند

$fw$  = فراوانی موارد در داخل فاصله‌ای که حاوی میانه است

$i$  = اندازه فاصله

در مثال قبل مقدار  $L$  برابر  $۲۵/۵$ ،  $N$  برابر  $۸$ ،  $cfb$  برابر  $۳$ ،  $fw$  برابر  $۴$  و  $i$  برابر  $۱$  می‌باشد.

$$\begin{array}{ccccccc} \frac{۲۳}{cfb = ۳} & \frac{۲۴}{۳} & \frac{۲۵}{۳} & \frac{۲۶}{۴} & \frac{۲۶}{۴} & \frac{۲۶}{۴} & \frac{۲۷}{۴} \\ & & \downarrow & & & & \\ & & L = ۲۵/۵ & & & & \end{array}$$

این مقادیر را در فرمول می‌گذاریم و میانه را به دست می‌آوریم.

$$Md = ۲۵/۵ + \left( \frac{\frac{۸}{2} - ۳}{۴} \right) ۱ = ۲۵/۷۵$$

مقدار  $i$  تنها زمانی ضرورت پیدا می‌کند که اندازه فاصله،  $۱$  نباشد، برای مثال، اگر با آزمون

پیشرفت تحصیلی، که نمره‌های آن با واحدهای ده دهی نظیر  $۳/۴$  و  $۵/۹$  ارائه می‌شوند، سروکار

داشته باشیم، اندازه فاصله  $۰/۱$  خواهد بود.

برای محاسبه ساده‌تر میانه می‌توان از توزیع فراوانی استفاده کرد. جدول ۵-۲ همان داده‌های

جدول ۵-۱ را نشان می‌دهد و شامل یک ستون اضافی، به نام فراوانی تجمعی ( $cf$ ) است که از آن

برای یافتن میانه استفاده می‌شود. فراوانیهای تجمعی، فراوانی مقادیر را در یک توزیع، تا هر فاصله به طوری که شامل همان فاصله هم باشد، معین می‌کنند. برای مثال، در جدول ۵-۲، فراوانی نمره‌ها، تا فاصله‌ای که شامل نمره ۳۰ است، همان‌طور که در ستون cf (فراوانی تجمعی) مشخص شده، ۶۴ است. این ستون ما را قادر می‌سازد تا فاصله‌ای را که میانه در آن جای دارد، مشخص کنیم. از آنجا که در یک توزیع، میانه نقطه‌ای است که ۵۰ درصد از موردها در زیر آن قرار می‌گیرند، فاصله‌ای که میانه را در خود دارد، دارای فراوانی تجمعی برابر با  $\frac{N}{2}$  خواهد بود. به عبارت دیگر، N توزیع را بر ۲ تقسیم می‌کنیم تا فاصله‌ای را که میانه در آن قرار دارد، پیدا کنیم.

در توزیع نمره‌هایی که در جدول ۵-۲ نشان داده شده است، تعداد موردها، یعنی N برابر ۱۰۵ است، بنابراین مقدار  $\frac{N}{2}$  برابر ۵۲/۵ می‌باشد. با نگاه کردن به ستون متوجه می‌شویم که فراوانی تجمعی برای نمره ۲۹، برابر ۵۰ و برای نمره ۳۰، برابر ۶۴ است، بنابراین، میانه در فاصله‌ای که نمره ۳۰ معرف آن است، قرار دارد. اکنون، امکان به کاربردن فرمول ۵-۱ و یافتن میانه توزیع وجود دارد. در این مثال مقدار  $\frac{N}{2}$ ، ۲۹/۵ است؛ زیرا فاصله‌ای که میانه را در خود دارد، معرف دامنه‌ای است که حد پایین آن ۲۹/۵ و حد بالای آن ۳۰/۵ است. مقدار cf (یعنی، فراوانی مقادیر زیر فاصله‌ای که میانه در آن است) ۵۰ است. فراوانی مقادیر موجود در فاصله مربوط به نمره ۳۰ (یعنی، مقدار fw) ۱۴ است. اندازه فاصله، یا مقدار  $\Delta$  در این توزیع، ۱ است؛ زیرا هر نمره، گستره فاصله را ۱ نشان می‌دهد. با استفاده از فرمول ۵-۱ در می‌یابیم که میانه توزیع، ۲۹/۶۷۸ است.

توجه داشته باشید که اندازه، نمره‌های فردی را به حساب نمی‌آورد. ما برای یافتن میانه، داده‌های خود را از کم به زیاد مرتب می‌کنیم و نقطه‌ای را که توزیع را به دو نیمه مساوی تقسیم می‌کند، پیدا می‌کنیم. میانه، یک آماره رتبه‌ای است؛ زیرا بنای آن بر رتبه‌بندی است. ما می‌توانیم از داده‌های فاصله‌ای یا نسبی نیز میانه را محاسبه کنیم، اما در چنین مواردی از خصیصه فاصله‌ای بودن داده‌ها استفاده نمی‌شود.

یکی از موافقی که میانه اندازه مطلوب گرایش مرکزی است، هنگامی است که چند نمره انتهایی در توزیع وجود داشته باشد. در چنین حالتی استفاده از یک اندازه گرایش مرکزی که مقدار هر نمره را به حساب می‌آورد، منتج به برآورد بالا یا برآورد پایین نمره معرف می‌شود.

میان به لحاظ عدم حساسیت به نمره‌های انتهایی مناسبترین شاخص برای یافتن نمره معرف است. برای روشن شدن مسئله، توزیع زیر را در نظر بگیرید:

۴۹ ۵۰ ۵۱ ۵۳ ۵۴ ۵۵ ۵۶ ۷۰ ۸۹

جدول ۲-۵ محاسبه میان به توزیع فراوانی نمرات ۱۰۵ دانش آموز در

آزمون آمار

نمره (X)	فراوانی (f)	فراوانی تجمعی (cf)
۳۵	۱	۱۰۵
۳۴	۳	۱۰۴
۳۳	۱۵	۱۰۱
۳۲	۸	۸۶
۳۱	۱۴	۷۸
۳۰	۱۴	۶۴
۲۹	۱۶	۵۰
۲۸	۱۰	۳۴
۲۷	۸	۲۴
۲۶	۴	۱۶
۲۵	۱	۱۲
۲۴	۳	۱۱
۲۳	۱	۸
۲۲	۳	۷
۲۱	۴	۴

$c f b = 50$

$$Md = L_{29/50} + \left( \frac{\frac{105}{2} - 50}{14} \right) 1 = 29/678$$

نمره ۵۴ که میان این توزیع است، بهترین معرف است. شاخصی که نمره‌های منفرد ۷۰ و ۸۹ را در نظر بگیرد، حتماً برآورد بالایی از نمره معرف به دست خواهد داد.

// میانگین

اندازه گرایش مرکزی که بیشترین مورد استفاده را دارد، میانگین است که معمولاً به نام

متوسط یا متوسط حسابی شناخته می‌شود. میانگین، عبارت است از مجموع تمام مقادیر در یک توزیع، تقسیم بر تعداد موردها. میانگین به زبان فرمول عبارت است از:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} \quad (5-2)$$

که معمولاً به این صورت نوشته می‌شود:

$$\bar{X} = \frac{\sum X}{N} \quad (5-3)$$

$\bar{X}$  = میانگین

$\Sigma$  = مجموع

$X$  = نمره خام

$N$  = تعداد موردها

با کاربرد فرمول ۵-۲ در مورد نمره‌های هوشیهر (IQ) زیر، در می‌یابیم که میانگین آنها ۱۱۱ است.

IQ نمره: ۱۱۲ ۱۲۱ ۱۱۵ ۱۰۱ ۱۱۹ ۱۰۹ ۱۰۰

$$\bar{X} = \frac{112 + 121 + 115 + 101 + 119 + 109 + 100}{7} = \frac{777}{7} = 111$$

توجه داشته باشید که در این محاسبه نمره‌ها مرتب نشدند. برای یافتن میانگین یک مجموعه از اندازه‌ها، نیازی به مرتب کردن نمره‌ها نیست.

از آنجا که میانگین یک متوسط حسابی است، آن را یک آماره فاصله‌ای به شمار می‌آورند. استفاده از آن برای داده‌های فاصله‌ای یا نسبی مناسب، اما برای داده‌های اسمی یا رتبه‌ای نامناسب است.

محاسبه میانگین از یک توزیع فراوانی

اگر داده‌ها به صورت یک توزیع فراوانی مرتب شوند، مجموع نمره‌ها را می‌توان با ضرب کردن در فراوانی آن، جمع کردن حاصلضربها و سپس تقسیم نتیجه بر تعداد موردها به دست آورد. فرمول محاسبه میانگین از یک توزیع فراوانی عبارت است از:

$$\bar{X} = \frac{\sum fX}{N} \quad (۵-۴)$$

جدول ۳-۵ محاسبه میانگین را برای داده‌های ارائه شده در جدول ۱-۵ نشان می‌دهد. آشکار است که کاربرد فرمول ۴-۵ برای آن توزیع فراوانی، کار محاسبه میانگین آن توزیع را آسان کرده است.

// مقایسه سه شاخص گرایش مرکزی

از آنجا که میانگین، یک آماره فاصله‌ای یا نسبی است، به طور کلی از میانه (آماره رتبه‌ای) یا نما (آماره اسمی) اندازه‌ای دقیق‌تر است. میانگین، مقدار هر نمره را به حساب می‌آورد. همچنین،

جدول ۳-۵ محاسبه میانگین برای توزیع فراوانی نمره‌های ۱۰۵ دانش‌آموز در آزمون آمار

x	f	fx
۳۵	۱	۳۵
۳۴	۳	۱۰۲
۳۳	۱۵	۴۹۵
۳۲	۸	۲۵۶
۳۱	۱۴	۴۳۴
۳۰	۱۴	۴۲۰
۲۹	۱۶	۴۶۴
۲۸	۱۰	۲۸۰
۲۷	۸	۲۱۶
۲۶	۴	۱۰۴
۲۵	۱	۲۵
۲۴	۳	۷۲
۲۳	۱	۲۳
۲۲	۳	۶۶
۲۱	۴	۸۴
	$N=105$	$\sum fX=3076$

$$\bar{X} = \frac{3076}{105} = 29/295$$

همچنین، میانگین در بین سه اندازه گرایش مرکزی، ثابت‌ترین آنهاست؛ زیرا اگر از یک جامعه آماری، به طور تصادفی، چند نمونه انتخاب شود، میانگین‌های این نمونه‌ها، کمتر از میان‌ها و نماهای آنها با هم تفاوت خواهند داشت. به همین دلیل میانگین بیشتر از دو شاخص دیگر، در پژوهشها مورد استفاده قرار می‌گیرد.

میانگین، بهترین شاخص عملکرد ترکیبی یک گروه یکپارچه است. با این همه، میانه بهترین شاخص عملکرد نوعی است. برای مثال، هیأت مدیره یک دبستان را در نظر بگیرید که در آمد سالانه آنها ۷۰۰۰۰، ۳۰۰۰۰، ۲۵۰۰۰، ۲۰۰۰۰ و ۲۰۰۰۰ دلار است. میانگین ۳۳۰۰۰ دلار، کل درآمد گروه را نسبت به تعداد اعضا نشان می‌دهد، اما این میانگین از درآمد همه اعضای هیأت مدیره بجز یک نفر، بیشتر است. در واقع میانه، ۲۵۰۰۰ دلار، تصویر بهتری از درآمد نوعی گروه را به دست می‌دهد.

### / شکل توزیعها

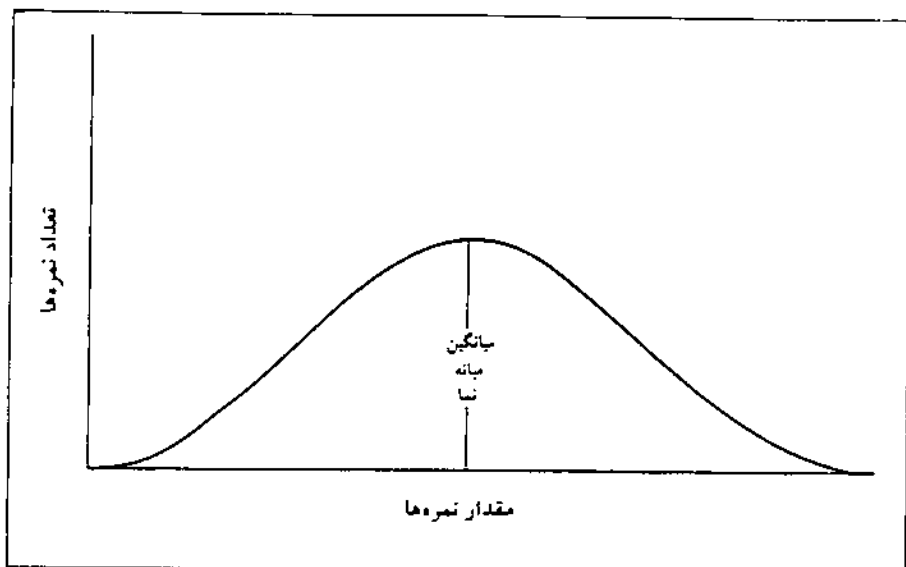
توزیعهای فراوانی می‌توانند شکل‌های گوناگون داشته باشند. یک توزیع، هنگامی متقارن است که دو نیمه آن، تصویر انعکاسی (تصویر آینه‌ای)<sup>(۱)</sup> یکدیگر باشند. در یک توزیع متقارن، مقادیر میانگین و میانه بر روی هم منطبق می‌شوند. اگر چنین توزیعی، به جای دو یا چند نما، تنها یک نما داشته باشد، سه شاخص گرایش مرکزی، همان طور که در شکل ۳-۵ نشان داده شده است، بر هم منطبق خواهند شد.

اگر توزیعی متقارن نباشد، آنرا توزیع دارای چولگی<sup>(۲)</sup> توصیف می‌کنند. در توزیعهای دارای چولگی مقادیر اندازه‌های گرایش مرکزی با هم تفاوت دارند. در چنین توزیعهایی، مقدار میانگین، به دلیل آنکه از نمره‌های انتهایی تأثیر می‌پذیرد، همان طور که در شکل ۴-۵ و ۵-۵

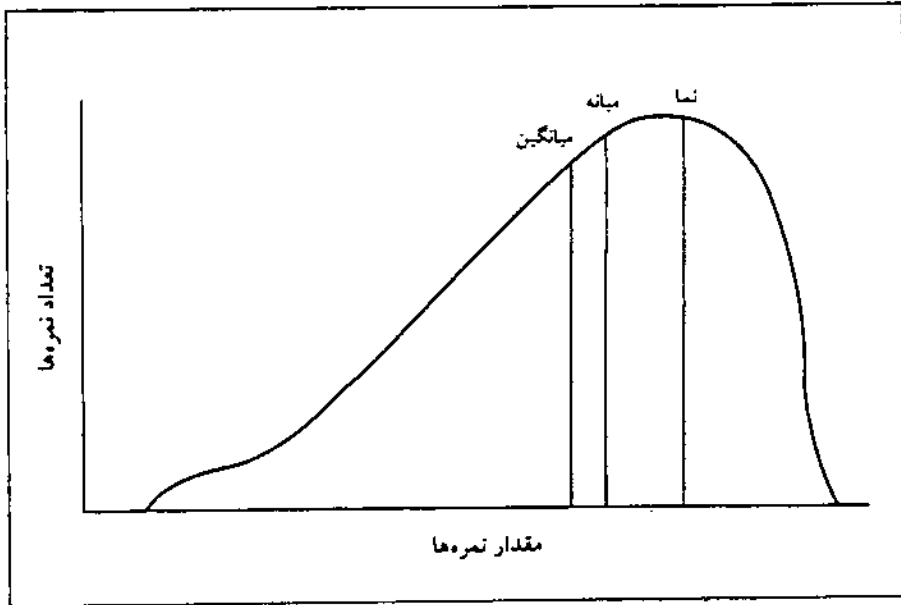
نشان داده شده است، به سمت انتهای توزیع، که نمره‌های افراطی در آن جای دارند، کشیده می‌شود. تأثیر نمره‌های افراطی بر میانه کمتر است؛ زیرا این شاخص از مقدار نمره تأثیر نمی‌پذیرد، بلکه از جایگاه نمره‌ها تأثیر می‌گیرد.

مقادیر افراطی هیچ تأثیری بر نما ندارند، زیرا این شاخص، هیچ رابطه‌ای با دو انتهای توزیع ندارد. زمانی که چولگی توزیع به سمت انتهای پایین‌تر باشد، یا چولگی منفی باشد، همیشه میانگین، کوچکتر از میانه است و معمولاً کوچکتر از نما است (شکل ۴ - ۵). زمانی که چولگی توزیع به سمت انتهای بالاتر باشد، یا چولگی مثبت باشد، میانگین همیشه بیشتر از میانه و میانه معمولاً بیشتر از نما است (شکل ۵ - ۵). چولگی یک توزیع را می‌توان از مقایسه میانگین و میانه تشخیص داد، بدون آنکه لازم باشد که نمودار چند ضلعی یا ستونی آن را ترسیم کرد.

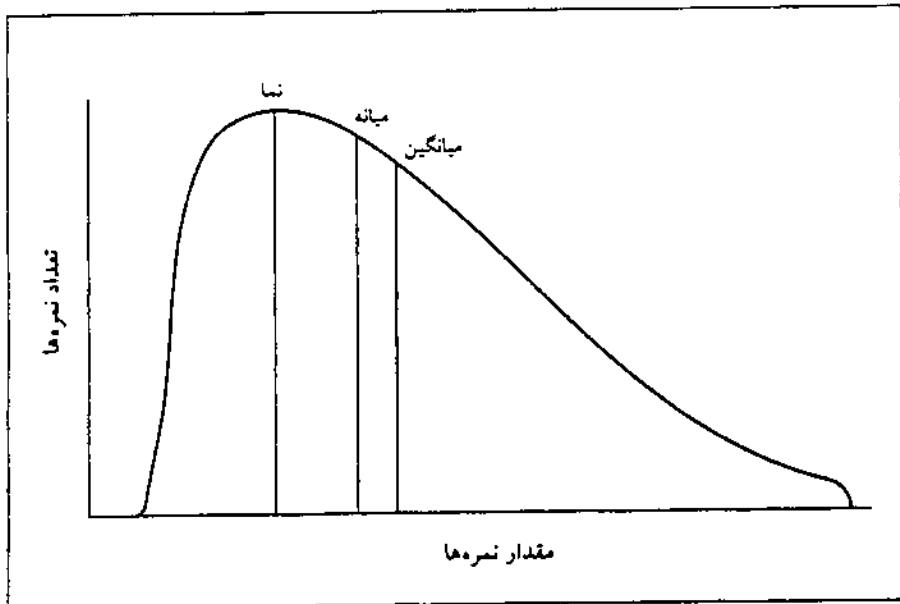
شکل ۳-۵ یک توزیع متقارن



شکل ۴ - ۵ یک توزیع دارای چولگی منفی



شکل ۵ - ۵ یک توزیع دارای چولگی مثبت





### /// اندازه‌های تغییرپذیری

اگر چه شاخصهای گرایش مرکزی در توصیف داده‌ها به بیان مقدار متوسط یا اندازه معرف کمک می‌کنند، اما تصویر کاملی از توزیع را ارائه نمی‌کنند. امکان دارد میانگینهای دو توزیع یکسان باشند، در حالی که درجه پراکنش<sup>(۱)</sup> یا تغییرپذیری نمره‌های آنها کاملاً با هم فرق کنند. امکان دارد در یک توزیع، نمره‌ها حول یک مقدار مرکزی گرد آیند، و در توزیع دیگر پراکنده باشند. برای روشن شدن موضوع، توزیع این نمره‌ها را در نظر بگیرید:

$$\bar{X} = \frac{175}{V} = 25 \quad \text{(الف) } 24, 24, 25, 25, 25, 26, 26$$

$$\bar{X} = \frac{175}{V} = 25 \quad \text{(ب) } 16, 19, 22, 25, 28, 30, 35$$

مقدار میانگین در هر دو توزیع ۲۵ است، اما درجه پراکنندگی نمره‌ها به طور قابل ملاحظه‌ای با هم تفاوت دارد. نمره‌های توزیع الف، آشکارا، بسیار همگن‌تر از نمره‌های توزیع ب هستند. روشن است که ضرورت شاخصی که توزیع‌ها را بر مبنای تغییرپذیری نمره‌ها توصیف کند، وجود دارد. در آمار شاخص‌های متعددی برای این منظور وجود دارند. چهار شاخصی که بیشتر از بقیه استفاده می‌شوند، عبارتند از: دامنه، انحراف چارکی، واریانس و انحراف معیار.

#### // دامنه

ساده‌ترین شاخص تغییرپذیری، دامنه است. دامنه، تفاوت بین بالاترین و پایین‌ترین نمره در یک توزیع است و با کسر کوچکترین مقدار از بیشترین، به دست می‌آید. دامنه، به شکل فرمول عبارت است از:

$$R = X_b - X_i \quad (5-5)$$

$R$  = دامنه

$X_b$  = بیشترین مقدار در یک توزیع

$X_i$  = کمترین مقدار در یک توزیع

برای مثال، در توزیعهای زیر، دامنه به ترتیب ۱۴ و ۶ است.

۲	۱۰	۱۱	۱۲	۱۳	۱۴	۱۶
۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵

دامنه، شاخص تغییرپذیری ناپایداری است؛ زیرا اساس آن تنها بر دو مقدار، بیشترین و کمترین، بنا شده است. همانطور که بسادگی در دو توزیع بالا دیده می‌شود، دامنه، نشانگر استواری برای ماهیت پراکندگی اندازه‌ها، حول یک مقدار مرکزی نیست. به همین دلیل، استفاده از دامنه بیشتر به مقاصد بازیانه محدود شده است. در برخی از گزارشهای تحقیقاتی، دامنه توزیع به عنوان یک آماره مورد استفاده قرار می‌گیرد، اما معمولاً استفاده از آن به همراه سایر اندازه‌های تغییرپذیری، نظیر انحراف چارکی و انحراف معیار صورت می‌پذیرد.

### // انحراف چارکی

انحراف چارکی عبارت از نصف تفاوت بین چارکهای بالا و پایین در یک توزیع است. چارک بالا ( $Q_3$ ) نقطه‌ای است که ۷۵ درصد از موارد در زیر آن قرار می‌گیرند. چارک پایین ( $Q_1$ ) نقطه‌ای است که ۲۵ درصد از موارد در زیر آن قرار می‌گیرند. همچنین، چارک بالا را صدک هفتاد و پنجم و چارک پایین را صدک بیست و پنجم به شمار می‌آورند. شیوه محاسبه  $Q_3$  و  $Q_1$  مشابه روشی است که برای محاسبه میانه به کار می‌رود. میانه، در واقع، چارک دوم است. در مورد  $Q_3$ ، فرمول محاسبه عبارت است از:

$$Q_3 = L + \left( \frac{\frac{3N}{4} - cfb}{fw} \right) i \quad (5-6)$$

و برای  $Q_1$  فرمول محاسبه عبارت است از:

$$Q_1 = L + \left( \frac{\frac{N}{4} - cfb}{fw} \right) i \quad (5-7)$$

$Q_+$  = چارک بالا

$Q_1$  = چارک پایین

$L$  = حد پایین فاصله‌ای که چارک در آن قرار می‌گیرد

$N$  = تعداد موردها در توزیع

$cfb$  = فراوانی تراکمی، در زیر فاصله‌ای که حاوی چارک است

$fiw$  = فراوانی موارد، در داخل فاصله‌ای که حاوی چارک است

$i$  = اندازه فاصله

هنگامی که مقادیر چارکهای اول و سوم پیدا شد، انحراف چارکی به این صورت محاسبه

می‌شود:

$$QD = \frac{Q_+ - Q_1}{2} \quad (5-8)$$

برای مثال، اگر در توزیع نمره‌ها، چارکهای بالا و پایین، به ترتیب، ۳۵ و ۱۵ باشند، مقدار

انحراف چارکی، ۱۰ خواهد بود.

$$QD = \frac{35 - 15}{2} = 10$$

اندازه‌ای که انحراف چارکی به دست می‌دهد، عبارت است از نصف دامنه نمره‌هایی که در آن ۵۰ درصد میانی موردها قرار دارند. اگر گستردگی نمره‌ها زیاد باشد، مقدار انحراف چارکی بیشتر از زمانی خواهد بود که گستردگی نمره‌ها کوچک باشد. هر چند انحراف چارکی، نظیر دامنه، مبتنی بر دو مقدار یک توزیع ( $Q_+$  و  $Q_1$ ) است، اما اندازه تغییرپذیری مفیدتری از دامنه است. این امر به دلیل آن است که  $Q_+$  و  $Q_1$  بسیار پایدارتر از بالاترین و پایین‌ترین مقادیر یک توزیع هستند، که دامنه مبتنی بر آنهاست.

انحراف چارکی به خانواده آماری میانه تعلق دارد؛ زیرا یک آمارهٔ رتبه‌ای است. از این آماره غالباً به همراه میانه استفاده می‌شود. آن را همچنین دامنه نیمه چارکی نیز می‌نامند. انحراف چارکی، نظیر میانه، بویژه هنگامی مفید است که ما به یک اندازه نیاز داشته باشیم که تحت تأثیر چند نمره انتهایی قرار نگیرد. اگر انحراف چارکی برای درآمد خانوار در دهکده A، ۱۰۰۰۰ دلار و در دهکده B، ۸۰۰۰ دلار باشد، در می‌یابیم که درآمدهای خانوار در دهکده A بر اساس اندازه‌گیری پراکندگی در وسط توزیع، ناهمگن‌تر هستند. برنده شدن یک خانوار در یک

بخت آزمایی در دهکده B، تغییر بسیار اندکی در میانه ایجاد خواهد کرد و یا اصلاً تغییری ایجاد نخواهد کرد؛ هر چند این امر افزایش قابل ملاحظه‌ای در میانگین و انحراف معیار به وجود می‌آورد.

### // واریانس و انحراف معیار

واریانس و انحراف معیار مفیدترین اندازه‌های تغییر پذیری هستند. هر دوی آنها بر پایه نمره انحراف از میانگین بنا شده‌اند. نمره‌هایی که تفاوت بین یک نمره خام و میانگین را نشان می‌دهند. فرمول یک نمره انحراف از میانگین، عبارت است از  $x = X - \bar{X}$ . نمره‌های خام زیر میانگین، نمره انحراف از میانگین منفی و نمره‌های خام بالای میانگین، نمره انحراف از میانگین مثبت خواهند داشت. بنا به تعریف، مجموع نمره‌های انحراف از میانگین در یک توزیع، همیشه صفر است. بنابراین اگر کسی بخواهد از نمره‌های انحراف از میانگین برای محاسبه اندازه‌های تغییر پذیری استفاده کند، باید برای این واقعیت که  $\sum x = 0$  است، چاره‌ای بیندیشد. شیوه مورد استفاده، مربع کردن همه نمره‌های انحراف از میانگین است تا همه آنها به اعداد مثبت تبدیل شوند. اگر ما پس از آن، مربع نمره‌های انحراف از میانگین را با هم جمع کنیم و به تعداد نمره‌ها تقسیم کنیم، میانگین مربع نمره‌های انحراف از میانگین، یا واریانس را به دست خواهیم آورد. واریانس به زبان ریاضی عبارت است از:

$$\sigma^2 = \frac{\sum x^2}{N} \quad (5-9)$$

$\sigma^2$  = واریانس

$\Sigma$  = مجموع

$x$  = انحراف هر نمره از میانگین،  $(X - \bar{X})$  یا به عبارت دیگر نمره انحراف از میانگین

$N$  = تعداد موردها در توزیع

از آنجا که هر نمره انحراف از میانگین به توان دوم می‌رسد، لزوماً واریانس با واحدی بیان می‌شود که مربع واحد اصلی اندازه است. برای مثال، امکان دارد که دریاپیم که واریانس قد شاگردان یک کلاس ۹ اینچ مربع است. این واریانس به ما می‌گوید که این کلاس از نظر قد ناهمگن تر از کلاسی است که واریانس آن ۴ اینچ مربع است و همگن تر از کلاسی است که واریانس آن ۱۶ اینچ مربع است.

در بسیاری از موارد کارشناسان آموزش و پرورش شاخصی را ترجیح می‌دهند که داده‌ها را به

همان واحد اندازه‌گیری داده‌های اصلی خلاصه کند. انحراف معیار ( $\sigma$ )، ریشه دوم (جذر) واریانس، چنین شاخصی را به دست می‌دهد. بنا به تعریف، انحراف معیار، ریشه دوم (جذر)، میانگین نمره‌های انحراف از میانگین است. اگر این تعریف را به زبان نمادین ریاضی بنویسیم، خواهیم داشت:

$$\sigma = \sqrt{\frac{\sum x^2}{N}} \quad (5-10)$$

برای وضوح بیشتر، جدول ۴-۵ را در نظر بگیرید. ستون اول این جدول توزیع نمره‌های آزمون ده دانش‌آموز را نشان می‌دهد. میانگین این توزیع ۶ است. ستون دوم، انحراف از میانگین هر نمره را ارائه می‌کند. برای مثال، انحراف نمره ۸ از میانگین، ۲+ است، انحراف نمره ۵ از میانگین، ۱- است، و الی آخر. ستون سوم، مربع هر یک از نمره‌های انحراف از میانگین را نشان می‌دهد. مجموع مربع نمره‌های انحراف از میانگین ۹۰ است. با جایگذاری این مقدار در فرمول و تقسیم آن بر ۱۰ (تعداد موردها) به عدد ۹ می‌رسیم، که میانگین مربع نمره‌های انحراف از میانگین است. ریشه دوم (جذر) این عدد ۳ است، که انحراف معیار این توزیع می‌باشد.

جدول ۴-۵. محاسبه انحراف معیار

(۱) $X$	(۲) $x = X - \bar{X}$	(۳) $x^2 = (X - \bar{X})^2$
۱۰	+۴	+۱۶
۹	+۳	+۹
۹	+۳	+۹
۸	+۲	+۴
۷	+۱	+۱
۶	۰	۰
۵	-۱	+۱
۳	-۳	+۹
۲	-۴	+۱۶
۱	-۵	+۲۵
$\sum X = ۶۰$		$\sum x^2 = ۹۰$

$$\bar{X} = \frac{۶۰}{۱۰} = ۶$$

$$\sigma = \sqrt{\frac{۹۰}{۱۰}} = \sqrt{۹} = ۳$$

روش بالا، وقتی مناسب است که میانگین توزیع، یک عدد سر راست (گرد شده) باشد، اما در بیشتر موارد چنین نیست. بنابراین برای اجتناب از زحمت کارکردن با اعداد کسری (اعشاری) از بسط فرمول اصلی، فرمول زیر به دست آمده است. فرمول زیر همان نتیجه را با کار کمتر به دست می‌دهد.

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} \quad (5-11)$$

$\sigma$  = انحراف معیار

$\sum X^2$  = مجموع مربعات هر نمره (یعنی، ابتدا  $\sum X^2$ )

نمره‌ها به توان دوم می‌رسد و سپس

با هم جمع می‌شوند)

مربع مجموع نمره‌ها (نمره‌ها ابتدا  $(\sum X)^2$ )

با هم جمع می‌شوند و سپس جمع

کل به توان دوم می‌رسد)

$N$  = تعداد موردها

جدول ۵ - ۵ محاسبه انحراف معیار را بر اساس داده‌های جدول ۴ - ۵ با این فرمول نشان می‌دهد. در این جدول، ستون اول، نمره‌ها و مجموع آنها را نشان می‌دهد. ستون دوم، مربع هر نمره و جمع مربع نمره‌ها را نشان می‌دهد. بقیه جدول، کاربرد فرمول ۵-۱۱ را برای یافتن مقدار انحراف معیار نشان می‌دهد. توجه داشته باشید که مقدار حاصله، همان مقداری است که با استفاده از فرمول ۵ - ۱۰ در مورد همان داده‌ها به دست آمده است.

انحراف معیار به خانواده آماری میانگین تعلق دارد، یعنی نظیر میانگین، یک آماره فاصله‌ای یا نسبی است و محاسبه آن بر اندازه نمره‌های فردی در توزیع بنا شده است. انحراف معیار، یکی از اندازه‌های تغییر پذیری است که بیشترین مورد استفاده را دارد و به همراه میانگین به کار می‌رود.

جدول ۵-۵ محاسبه انحراف معیار با استفاده از فرمول ۱۱-۵

$X$	$X^2$	
۱۰	۱۰۰	$\sigma = \sqrt{\frac{450 - \frac{(60)^2}{10}}{10}}$ $= \sqrt{\frac{450 - 360}{10}}$ $= \sqrt{9}$ $= 3$
۹	۸۱	
۹	۸۱	
۸	۶۴	
۷	۴۹	
۶	۳۶	
۵	۲۵	
۳	۹	
۲	۴	
۱	۱	
$\sum X = 60$	$\sum X^2 = 450$	

### /// نمره‌های استاندارد

ما اغلب دوست داریم مقایسه‌ای بین موقعیت‌های نسبی یک فرد در دو آزمون متفاوت صورت دهیم. مقایسه درست دو نمره تنها در صورتی امکان دارد که دو آزمون، میانگین و انحراف معیار یکسان داشته باشند، اما این امر در عمل کمتر اتفاق می‌افتد. برای غلبه بر این مشکل، می‌توانیم اندازه‌ها (نمره‌های متفاوت) را به نمره‌های استاندارد برگردانیم. یک نمره استاندارد، که موارد استفاده گسترده‌ای دارد و در تحلیل‌های آماری نقش مهمی بازی می‌کند، نمره  $Z$  است، و تعریف آن عبارت است از فاصله یک نمره از میانگین که بر اساس واحدهای انحراف معیار اندازه‌گیری شده باشد. فرمول محاسبه نمره  $Z$  عبارت است از:

$$z = \frac{X - \bar{X}}{\sigma} = \frac{x}{\sigma} \quad (5-12)$$

$X$  = نمره خام

$\bar{X}$  = میانگین توزیع

$\sigma$  = انحراف معیار توزیع

$x$  = نمره انحراف از میانگین

با کاربرد این فرمول، نمره‌ای که دقیقاً یک انحراف معیار بالاتر از میانگین باشد، +۱، و نمره‌ای که دقیقاً یک انحراف معیار زیر میانگین باشد، -۱ خواهد شد. نمره  $z$  برای نمره‌ای که مقدار عددی آن برابر میانگین باشد، صفر خواهد بود. برای روشن شدن مطلب، فرض کنید نمره دانشجویی در آزمون روان‌شناسی ۷۲ باشد، در حالی که میانگین توزیع آن ۷۸ و انحراف معیار آن ۱۲ است. همچنین، فرض کنید همان دانشجو در آزمون آمار ۴۸ گرفته باشد، در حالی که میانگین و انحراف معیار توزیع نمرات آزمون آمار ۵۱ و ۶ است. اگر اعداد مربوط را در فرمول ۱۲ - ۵ بگذاریم، برای هر آزمون یک نمره  $z$  خواهیم داشت.

$$\begin{array}{l} \text{آمار} \\ z_2 = \frac{48 - 51}{6} = -0.5 \\ \text{روان‌شناسی} \\ z_1 = \frac{72 - 78}{12} = -0.5 \end{array}$$

هر دو این نمره‌های استاندارد به توزیع  $z$  تعلق دارند، که بنا به تعریف، میانگین و انحراف معیار آن صفر و یک است، و بنابراین به طور مستقیم قابل مقایسه‌اند. در این مثال روشن است که نمره ۷۲ در آزمون روان‌شناسی و نمره ۴۸ در آزمون آمار با هم برابر هستند، یعنی هر دوی این نمره‌ها نمایانگر سطح نسبی عملکرد همسان می‌باشند. به عبارت دیگر، موقعیت دانشجویی که این نمره‌ها را کسب کرده است، در مقایسه با عملکرد همکلاسانش، در هر دو آزمون همسان است. بدون استفاده از روش نمره  $z$ ، انجام چنین مقایسه‌ای بسیار دشوار خواهد بود. اکنون مثال دیگری می‌زنیم: فرض کنید دانشجوی دیگری که همان آزمونها را گذرانده است، در آزمون روان‌شناسی نمره ۸۱ و در آزمون آمار نمره ۵۳ کسب کرده است. مانند مثال قبل، مقایسه این نمره‌های خام برای نشان دادن اینکه این دانشجو در کدام آزمون بهتر عمل کرده، مشکل است. تبدیل این نمره‌ها به نمره  $z$  مقایسه را آسان می‌سازد. با استفاده از فرمول ۱۲ - ۵، مقادیر  $z_1$  و  $z_2$  را برای این مورد، پیدا می‌کنیم.

$$\begin{array}{l} \text{آمار} \\ z_2 = \frac{53 - 51}{6} = +0.33 \\ \text{روان‌شناسی} \\ z_1 = \frac{81 - 78}{12} = +0.25 \end{array}$$



این نتیجه شگفت‌آور نشان می‌دهد که نمره ۵۳ در آزمون آمار، نمایانگر عملکرد نسبی اندک بهتری از نمره ۸۱ در آزمون روان‌شناسی است. این دانشجو، در مقایسه با همتایانش، در آمار کار بهتری نسبت به روان‌شناسی داشته است.

یکی از اشکالات نمره  $Z$  آن است که می‌باید با مقادیر منفی و کسرهای اعشاری سروکار داشت. ما می‌توانیم به هنگام مقایسه بین نمره‌هایی از توزیعهای مختلف، برای غلبه بر این دشواریها نمره  $Z$  را به مقیاس نمره استاندارد دیگری تغییر دهیم که اعداد منفی و اعشاری در بر ندارد.

یکی از رایج‌ترین نمره‌های استاندارد، توزیع نمره  $Z$  است که دارای میانگین ۵۰ و انحراف معیار ۱۰ است. برای تبدیل نمره  $Z$  به نمره  $Z$  مقدار  $Z$  را در ۱۰ ضرب می‌کنیم و با ۵۰ جمع می‌کنیم.

فرمول نمره  $Z$  چنین است:

$$Z = 10z + 50 = 10 \left( \frac{X - \bar{X}}{\sigma} \right) + 50 \quad (5-13)$$

فرض کنید نمره دانشجویی در آزمون زبان اسپانیولی ۲۱ است. اگر میانگین نمره‌ها در این آزمون ۲۷ و انحراف معیار ۶ باشد، نمره  $Z$  عبارت خواهد بود از  $(21 - 27) / 6$ ، که می‌توان آن را مستقیماً در فرمول  $Z$  قرار داد:

$$Z = 10 = \left( \frac{21 - 27}{6} \right) + 50 = 40$$

تبدیل نمره  $Z$  به نمره  $Z$  نه تنها ما را قادر می‌سازد که با اعداد صحیح کار کنیم، بلکه همچنین، سبب می‌شود که از نکات منفی روان‌شناختی که در توصیف عملکرد آزمودنیها با اعداد منفی وجود دارد، اجتناب کنیم.

آموزگاران می‌خواهند موقعیت شاگردان خود را در آزمونهای متوالی مقایسه کنند، یا می‌خواهند تمام نمراتی را که شاگردان در آزمونهای مختلف یک دوره درسی کسب کرده‌اند، با هم جمع و یک توزیع کلی فراهم کنند، می‌توانند نمرات خام دانشجویان را به نمره  $Z$  یا نمره  $Z$

برگردانند تا به هر مجموعه از نمره‌ها وزن یکسانی داده باشند. اکنون باید روشن شده باشد که جمع کردن و میانگین گرفتن از نمره‌هایی که به توزیعهای مختلفی تعلق دارند و دارای میانگین و انحراف معیار متفاوتی هستند، بدون تبدیل (برگرداندن) آنها به نوعی از نمره‌های استاندارد، از نظر آماری تأیید نمی‌شود.

علاوه بر  $Z$  توزیعهای نمره استاندارد تبدیل یافته دیگری نیز وجود دارند. برای تبدیل یک توزیع نمره‌ها به یک توزیع استاندارد شده جدید، تنها لازم است که نمره  $Z$  را در انحراف معیار دلخواه ضرب و با میانگین دلخواه جمع کنیم. فرمول کلی عبارت است از:

$$A = \mu_A + \sigma_A (z) \quad (5 - 135)$$

$A$  = نمره استاندارد در مقیاس جدید

$\mu_A$  = میانگین مقیاس استاندارد جدید

$\sigma_A$  = انحراف معیار برای مقیاس استاندارد جدید

برای مثال، هر قسمت از امتحان ورودی بعد از لیسانس (GRE)<sup>(۱)</sup> یک میانگین ۵۰۰ و یک انحراف معیار ۱۰۰ برای توزیع تبدیل یافته خود دارد. اگر شما در بخش کلامی GRE، ۱/۵ انحراف معیار بالای میانگین ( $z = 1/5$ ) قرار گرفته باشید، نمره شما ۶۵۰ [  $500 + 100(1/5) = 650$  ] گزارش خواهد شد. اگر نمره کمی شما ۵۰۰ باشد، شما نمره‌ای به دست آورده‌اید که دقیقاً برابر میانگین است.

مقیاس هوش استانفورد - بینه یک مقیاس نمره استاندارد با میانگین ۱۰۰ و انحراف معیار ۱۶ است. بنابراین برای تبدیل یک نمره خام آزمون استانفورد - بینه به مقیاس استاندارد، ابتدا باید  $Z$  آن را تعیین و سپس آن را در ۱۶ ضرب و با ۱۰۰ جمع کرد.

تبدیل مجموعه‌ای از نمره‌ها به نمره‌های استاندارد، شکل توزیع اصلی را تغییر نمی‌دهد. اگر توزیع نمره‌ها چولگی داشته باشد، نمره‌های استاندارد حاصله نیز یک توزیع دارای چولگی تولید خواهند کرد. تنها اگر توزیع اصلی نرمال باشد، نمره‌های استاندارد یک توزیع نرمال تولید خواهند کرد.

## /// منحنی نرمال

دیده شده است که توزیع بسیاری از اندازه‌های فیزیکی و روانی، وقتی که به صورت چند ضلعی فراوانی ترسیم شوند، شکل زنگوله را به خود می‌گیرند. برای مثال، اگر قد پسران آمریکایی را در دهمین روز تولد اندازه بگیریم، در خواهیم یافت که قد اکثر کودکان نزدیک میانگین است، قد تعداد کمتری نیز اندکی بالاتر و پایین تر از میانگین می‌باشد. البته هر چقدر از میانگین دور شویم، در هر سطح بلندی قد، پسران کمتری یافت خواهند شد. یک نمودار چند ضلعی که این توزیع را نشان بدهد، کاملاً شبیه چندضلعی نظری است که منحنی نرمال خوانده می‌شود. جداول عرض و سطوح منحنی نرمال پیش از این محاسبه شده‌اند. سطوح منحنی نرمال در جدول ۱ - A ضمیمه، نشان داده شده‌اند. این نمودار چند ضلعی فرضی، فراوانیهای مورد انتظار (یعنی نظری) همه نمره‌های ممکن  $Z$  را نشان می‌دهد. این چند ضلعی نشان می‌دهد که انتظار به وقوع پیوستن نمره‌های  $Z$  نزدیک به صفر، بیشتر از مقادیر دیگر نمره‌های  $Z$  است و هر چقدر یک نمره  $Z$  دورتر از صفر باشد، انتظار کمتری برای به وقوع پیوستن آن وجود دارد.

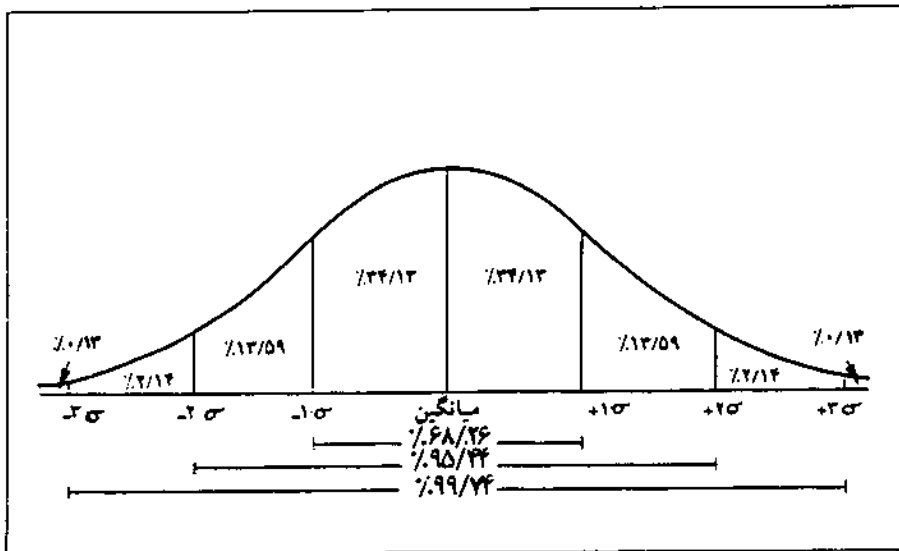
از آنجا که بسیاری از توزیعهای رخدادهای طبیعی به منحنی نرمال شباهت دارند، ثابت شده که این مدل نظری، بسیار مفید است. هر گاه دریابیم یا بر این باور باشیم که داده‌های واقعی به صورت توزیع شبیه منحنی نرمال هستند، می‌توانیم با استنتاج قیاسی برآوردهای مفیدی از خواص نظری منحنی نرمال به عمل آوریم. منحنی نرمال، توزیع متقارنی از اندازه‌ها است که در فواصل مشخص آن، در زیر میانگین و بالای میانگین، تعداد یکسانی جای گرفته‌اند. میانگین نقطه‌ای است که در زیر آن ۵۰ درصد از موردها و بالای آن ۵۰ درصد دیگر از موردها جای دارند. در چنین توزیعی میانه و نما مقادیر یکسانی هستند و بر میانگین منطبق می‌شوند. در یک منحنی نرمال، اکثر موردها نزدیک میانگین متمرکز می‌شوند. همچنانکه از میانگین در دو جهت دور می‌شویم، مقدار موردها کاهش می‌یابد. در منحنی نرمال تقریباً ۳۴ درصد از موردها بین میانگین و یک انحراف معیار بالای میانگین یا یک انحراف معیار زیر میانگین قرار می‌گیرند. سطح بین یک انحراف معیار از میانگین و دو انحراف معیار از میانگین، در هر دو سوی توزیع، شامل حدود ۱۴ درصد از موردها است. تنها حدود ۲ درصد از موردها بین دو انحراف معیار و سه انحراف معیار از میانگین و تنها حدود ۰/۱ درصد از موردها بالا یا پایین سه انحراف معیار از میانگین قرار دارند. این امر در شکل ۵-۶ به تصویر در آمده است.

می‌توان با مراجعه به جدول ۱ - A در ضمیمه، که سطوح منحنی نرمال را ارائه می‌کند، درصد موردهای پایین و بالای هر نمره Z را در یک منحنی نرمال تعیین کرد. در این جدول، ستون اول شامل مقادیر مختلف Z است.

در ستون دوم سطح زیر منحنی بین میانگین و هر مقدار Z داده شده است. سرانجام، ستون سوم سطح باقیمانده از هر نمره Z تا انتهای منحنی را نشان می‌دهد. بنابراین حاصل جمع سطوحی که در ستون دوم و سوم آمده است، به  $0/5000$  می‌رسد. برای مثال، نمره Z برابر با  $+0/70$  را در نظر بگیرید. سطح بین این مقدار Z و میانگین را می‌توان در ستون دوم یافت، این سطح  $0/2580$  است. این رقم می‌گوید که حدود ۲۶ درصد از موردها بین این مقدار Z و میانگین توزیع قرار می‌گیرند.

چون میانگین منحنی نرمال بر میانه منطبق می‌شود، ۵۰ درصد از موردها زیر میانگین قرار می‌گیرند.  $0/5$  را به  $0/2580$  اضافه می‌کنیم و نتیجه به ما می‌گوید می‌توانیم انتظار داشته باشیم که  $75/8$  درصد از موردها زیر نمره Z  $+0/70$  قرار بگیرند. ستون سوم نشان می‌دهد که  $24/2$  درصد از موردها بالای نمره Z  $+0/70$  قرار می‌گیرند.

شکل ۵-۶ درصد موردهای بین انحراف معیارهای متوالی در یک منحنی نرمال



زمانی که مقدار  $z$  منفی باشد، این روش بر عکس می‌شود، یعنی به جای آنکه درصد موردهای بین میانگین و مقدار  $z$  داده شده را با ۵۰ درصد جمع کنیم، آن را از ۵۰ درصد کم می‌کنیم. فرض کنید می‌خواهیم درصد موردهای زیر مقدار  $z = ۰/۷۰ -$  را پیدا کنیم. سطح بین میانگین و نمره  $z = ۰/۷۰ -$ ،  $۰/۲۵۸۰$ ، یا به زبان درصد،  $۲۵/۸$  درصد از موردها است. با کم کردن  $۲۵/۸$  از  $۵۰$ ، عدد  $۲۴/۲$  را به دست خواهیم آورد.

این نتیجه نشان می‌دهد که در یک منحنی نرمال تنها در حدود ۲۴ درصد از نمره‌ها زیر نمره  $z = ۰/۷۰ -$  قرار می‌گیرند. این مقدار را همچنین می‌توان در ستون سوم جدول سطوح زیر منحنی نرمال یافت. در این ستون مقدار داده شده برای یک نمره  $z = ۰/۷۰$  برابر با  $۰/۲۴۲۰$  است. سطوح زیر منحنی نرمال بین دو نمره  $z$  دقیقاً تنها برای یک توزیع نرمال کاربرد دارد. بهره‌گیری از ویژگیهای یک منحنی نرمال در مورد توزیعهای دارای چولگی، در بهترین حالت، فقط یک تخمین به بار می‌آورد. بنابراین، ضرورت دارد که قبل از بکارگیری این ویژگیها، شکل توزیع بررسی شود.

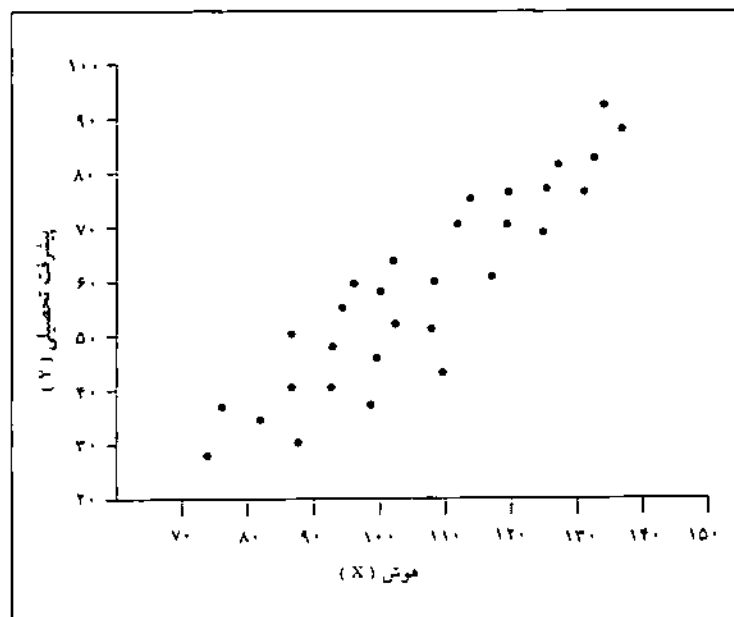
### /// همبستگی

بحث ما در مورد روشهای آماری، تاکنون، درباره توصیف نمردهای یک توزیع منفرد بوده است. اکنون می‌خواهیم درباره روشی که رابطه بین زوجهایی از نمره‌ها را نشان می‌دهد، بحث کنیم. روشهای آماری برای تعیین رابطه بین زوجهایی از نمره‌ها، شیوه‌های همبستگی نامیده می‌شوند. به طور نوعی، اندازه‌گیری دو متغیر برای هر عضو یک گروه معمول است و می‌توان وجود رابطه بین این اندازه‌گیرهای زوج شده را تعیین کرد. شیوه‌های همبستگی نشان می‌دهند که تغییر در یک متغیر، تا چه اندازه با تغییر در متغیر دیگر پیوستگی دارد. برای مثال، می‌دانیم که پیشرفت تحصیلی و هوش با هم رابطه دارند و بنابراین انتظار داریم دانشجویانی که از  $IQ$  بالایی برخوردارند، در آزمونهای پیشرفت تحصیلی نمرات بالاتر از میانگین کسب کنند. یک راه ساده نمایش این رابطه، ترسیم نمرات آزمون هوش و نمرات آزمون پیشرفت تحصیلی عده‌ای از افراد در یک جدول دو بعدی است، که آن را نمودار پراکنندگی می‌نامند. یک مجموعه از نمرات بر محور افقی، به ترتیب از کمترین به بیشترین و از چپ به راست، رسم می‌شود. مجموعه دیگر نمرات بر محور عمودی به ترتیب از پایین به بالا رسم می‌شود. سپس موقعیت هر فرد در دو

آزمون با یک نقطه منفرد بر روی نمودار پراکندگی نشان داده می‌شود. در شکل ۷ - ۵ نمرات آزمون پیشرفت تحصیلی ۳۰ دانش‌آموز کلاس دهم در برابر نمرات آزمون هوش آنها ترسیم شده است. بررسی این شکل آشکار می‌کند که با بالا رفتن نمره هوش، نمره آزمون پیشرفت تحصیلی نیز به سمت بالا گرایش پیدا می‌کند.

ترسیم یک نمودار پراکندگی ما را قادر می‌سازد که هم جهت و هم شدت رابطه را مشاهده کنیم. جهت رابطه به مثبت یا منفی بودن رابطه اطلاق می‌شود. در شکل ۷ - ۵ نقطه‌ها طرحی را تشکیل می‌دهند که از سمت چپ پایین آغاز و به سمت راست بالا ختم می‌شود، به گونه‌ای که نمره‌های کم در یک متغیر (هوش) با نمره‌های کم در متغیر دیگر (پیشرفت تحصیلی) و نمره‌های زیاد در یک متغیر با نمره‌های زیاد در متغیر دیگر پیوستگی دارند (بنا به قرار داد، نمره‌های متغیر مستقل  $[X]$  بر روی محور افقی و نمره‌های متغیر وابسته  $[Y]$  بر روی محور عمودی ترسیم می‌شوند). یک چنین رابطه‌ای بین متغیرها مثبت نامیده می‌شود؛ زیرا نمره‌های زیاد با نمره‌های زیاد و نمره‌های کم با نمره‌های کم پیوستگی دارند.

شکل ۷ - ۵ یک نمودار پراکندگی نمایانگر رابطه بین هوش و پیشرفت تحصیلی



رابطه بین دو متغیر همیشه مثبت نیست. برخی از متغیرها با هم رابطه منفی دارند. برای مثال، روشن شده است که نرخ تولد و پایگاه اجتماعی - اقتصادی با هم رابطه منفی دارند، یعنی نرخ تولد با بالا رفتن پایگاه اجتماعی - اقتصادی کاهش می‌یابد. در یک رابطه منفی نمره‌های زیاد در یک متغیر با نمره‌های کم در متغیر دیگر پیوستگی می‌یابند و نقطه‌های موجود در نمودار پراکندگی از سمت چپ بالا به سمت راست پایین امتداد پیدا می‌کند. همچنین، یک نمودار پراکندگی نمره‌های  $z$  شدت رابطه بین متغیرها را آشکار می‌کند. اگر نقطه‌های موجود در نمودار پراکندگی، یک نوار باریک را تشکیل دهند؛ به گونه‌ای که اگر یک خط مستقیم در طول نوار رسم شود، نقطه‌ها نزدیک خط قرار گیرند، یک رابطه قوی بین متغیرها وجود دارد. به هر حال اگر نقطه‌های نمودار پراکندگی نمره‌های  $z$  کاملاً متفرق باشند، رابطه بین متغیرها نسبتاً ضعیف است. نمودار پراکندگی در شکل ۸ - ۵ رابطه‌های مثبت و منفی و قوی و ضعیف گوناگونی را نشان می‌دهد.

### /// ضریب‌های همبستگی

دسته‌ای از شاخصهای آماری برای این تهیه شده‌اند که هم، جهت (منفی یا مثبت) و هم، شدت رابطه بین متغیرها را نشان دهند. این شاخصها ضرایب همبستگی نامیده می‌شوند. محاسبه ضریب همبستگی بین دو متغیر، مقداری را به دست می‌دهد که دامنه آن  $-1$  تا  $+1$  است. ضریب همبستگی  $-1$  - رابطه منفی کامل را نشان می‌دهد، مقدار  $+1$  متضمن یک رابطه مثبت کامل است و نقطه وسط این دامنه، صفر، نمایانگر آن است که هیچ رابطه‌ای بین دو متغیر وجود ندارد. یک ضریب همبستگی مثبت کامل، هنگامی به دست می‌آید که نمره  $z$  هر فرد در یک متغیر، از نظر اندازه و علامت با نمره  $z$  او در متغیر دیگر یکسان باشد. از سوی دیگر، یک ضریب همبستگی منفی کامل، هنگامی به دست می‌آید که نمره‌های  $z$  هر فرد از نظر اندازه یکسان و از نظر علامت متضاد باشند. یک همبستگی صفر، هنگامی به دست می‌آید که هیچ یک از این دو روند وجود نداشته باشند، یعنی جایگاه فرد در یک متغیر با جایگاه او در متغیر دیگر همبسته نباشد. ضریب همبستگی نزدیک به یک، چه  $-1$  و چه  $+1$ ، رابطه بسیار بالا را نشان می‌دهد. چنین رابطه‌های بالایی انجام پیش‌بینی‌های دقیق را درباره یک متغیر، بر مبنای اطلاعاتی از یک متغیر دیگر، ممکن می‌سازند. از ضریب همبستگی منفی، همانند ضریب همبستگی مثبت، می‌توان برای پیش‌بینی استفاده کرد. در شکل ۸ - ۵ ضرایب همبستگی برای مجموعه‌هایی از داده‌ها به صورت نمره  $z$  آمده است. توجه کنید که در حالت همبستگی کامل

تمام نمره‌ها بر روی یک خط مستقیم قرار می‌گیرند. هر چه ضریب همبستگی به صفر نزدیکتر باشد، انحراف نمره‌ها از خط مستقیم بیشتر است. در حالت همبستگی صفر (۰)، نمره‌ها بر سطح نگاره پراکنده هستند و در هیچ جهتی شکل معینی به خود نمی‌گیرند. ضرایب همبستگی در اندازه‌های روان‌شناختی و آموزش، به دلیل پیچیدگی این پدیده‌ها، ندرتاً به بیشترین حد، یعنی +۱ یا -۱ می‌رسند. در مورد این اندازه‌ها، هر ضریبی که بیشتر از به‌اضافه یا منهای ۰/۹ باشد، معمولاً همبستگی بسیار بالا محسوب می‌شود.

### // همبستگی گشتاوری

ضریب همبستگی گشتاوری، که آماردان انگلیسی کارل پیرسون آن را ارائه کرده است و  $r$  پیرسون نامیده می‌شود، یک شاخص همبستگی است که بیشترین مورد استفاده را دارد. این ضریب هنگامی که مقیاس اندازه‌گیری فاصله‌ای یانسی باشد، به کار می‌رود. این شاخص را میانگین حاصل ضرب نمرات  $Z$  تعریف می‌کنند، یعنی نمره  $Z$  هر فرد برای متغیر  $(X)$  در نمره  $Z$  همان فرد برای متغیر دیگر  $(Y)$  ضرب می‌شود. حاصلضرب این نمره‌های  $Z$  زوج شده با هم جمع و به تعداد آنها تقسیم می‌شود. تعریف  $r$  پیرسون به زبان فرمول عبارت است از:

$$r = \frac{\sum Z_x Z_y}{N} \quad (۵-۱۴)$$

ضریب همبستگی گشتاوری پیرسون  $r =$

مجموع حاصلضرب نمره‌های  $Z$   $\sum Z_x Z_y =$

تعداد نمره‌های زوج شده  $N =$

به دلیل نحوه تعریف نمره‌های  $Z$  به زبان ریاضی، هنگامی که نمره  $Z$  همه افراد در دو متغیر  $X$  و  $Y$  همسان باشد، مجموع حاصلضربهای  $Z_x Z_y$  با تعداد زوجهای نمره‌های  $Z$  برابر خواهد بود و میانگین حاصلضرب نمره‌های  $Z$  (همبستگی گشتاوری پیرسون) مساوی +۱ خواهد شد. اگر همخوانی کامل مثبت بین نمره‌های  $Z$  وجود داشته باشد، همبستگی گشتاوری +۱ خواهد شد. اگر نمره‌های  $Z$  از نظر عددی یکسان، اما از نظر علامت متضاد باشند، حاصلضرب منفی می‌شود و همبستگی گشتاوری پیرسون -۱ خواهد شد. در هر دو حالت وقتی که همه نمره‌ها بر روی یک نمودار پراکنده‌گی ترسیم شوند، بر روی یک خط مستقیم قرار می‌گیرند. اکنون بیایید با استفاده از



فرمول ۱۴-۵ همبستگی بین نمره‌های ۱۴ آزمودنی را در دو آزمون  $X$  (آمار توصیفی) و  $Y$  (آمار استنباطی)، آن گونه که در جدول ۶-۵ نشان داده شده است، محاسبه کنیم.

ستونهای (۲) و (۳)، به ترتیب، نمره‌های خام آزمودنیها ( $X$ ) و نمره‌های انحراف از میانگین ( $x$ ) را در آزمون آمار توصیفی نشان می‌دهند. ستونهای (۵) و (۶) نمره‌های خام آزمودنیها ( $Y$ ) و نمره‌های انحراف از میانگین ( $y$ ) را در آزمون آمار استنباطی ارائه می‌کنند. ستونهای (۴) و (۷) مربع نمره‌های انحراف از میانگین را که برای محاسبه انحراف معیار مورد استفاده قرار گرفته‌اند، نشان می‌دهند. ستونهای (۸) و (۹) نمره‌های  $Z$  مربوط به نمره‌های  $X$  و  $Y$  را که با استفاده از فرمول ۱۲-۵ محاسبه شده‌اند، معرفی می‌کنند. ستون (۱۰) حاصلضربهای مقادیر  $z_x z_y$  را نشان می‌دهد. مجموع این حاصلضربها  $۱۱/۵۰$  است. اکنون امکان یافتن ضریب همبستگی گشتاوری پیرسون، بین دو مجموعه از نمره‌ها، با کاربرد فرمول ۱۴-۵ فراهم آمده است.

$$r = \frac{۱۱/۵۰}{۱۴} = +۰/۸۲$$

جدول ۶-۵ محاسبه  $r$  پیرسون بین دو مجموعه از نمره‌ها ( $Y$  و  $X$ )

(۱)	(۲)	(۳)	(۴)	(۵)	(۶)	(۷)	(۸)	(۹)	(۱۰)
آزمودنیها	نمره‌های $X$	$x$	$x^2$	نمره‌های $Y$	$y$	$y^2$	$z_x$	$z_y$	$z_x z_y$
۱	۱۸	-۳	۹	۲۸	۴	۱۶	-۱/۵	۰	۰/۵
۲	۱۸	-۳	۹	۳۰	۶	۳۶	-۱/۵	-۱/۵	-۲/۲۵
۳	۱۷	-۴	۱۶	۳۰	۶	۳۶	-۱	-۱/۵	-۱/۵
۴	۱۷	-۴	۱۶	۲۶	۲	۴	-۱	-۱/۵	-۱/۵
۵	۱۶	-۵	۲۵	۲۸	۴	۱۶	-۱/۵	۰	۰/۵
۶	۱۶	-۵	۲۵	۲۴	۰	۰	-۱/۵	۰	۰
۷	۱۵	-۶	۳۶	۲۲	-۲	۴	۰	-۱/۵	۰
۸	۱۵	-۶	۳۶	۲۰	-۴	۱۶	۰	-۱	۰
۹	۱۴	-۷	۴۹	۲۶	۲	۴	-۱/۵	-۱/۵	-۱/۲۵
۱۰	۱۴	-۷	۴۹	۲۲	-۲	۴	-۱/۵	-۱/۵	-۱/۲۵
۱۱	۱۳	-۸	۶۴	۲۴	۰	۰	-۱	۰	۰
۱۲	۱۳	-۸	۶۴	۱۸	-۶	۳۶	-۱	-۱/۵	-۱/۵
۱۳	۱۲	-۹	۸۱	۲۰	-۴	۱۶	-۱/۵	-۱	-۱/۵
۱۴	۱۲	-۹	۸۱	۱۸	-۶	۳۶	-۱/۵	-۱/۵	-۱/۲۵
	۲۱۰		۵۶	۲۳۶		۲۲۴			۱۱/۵

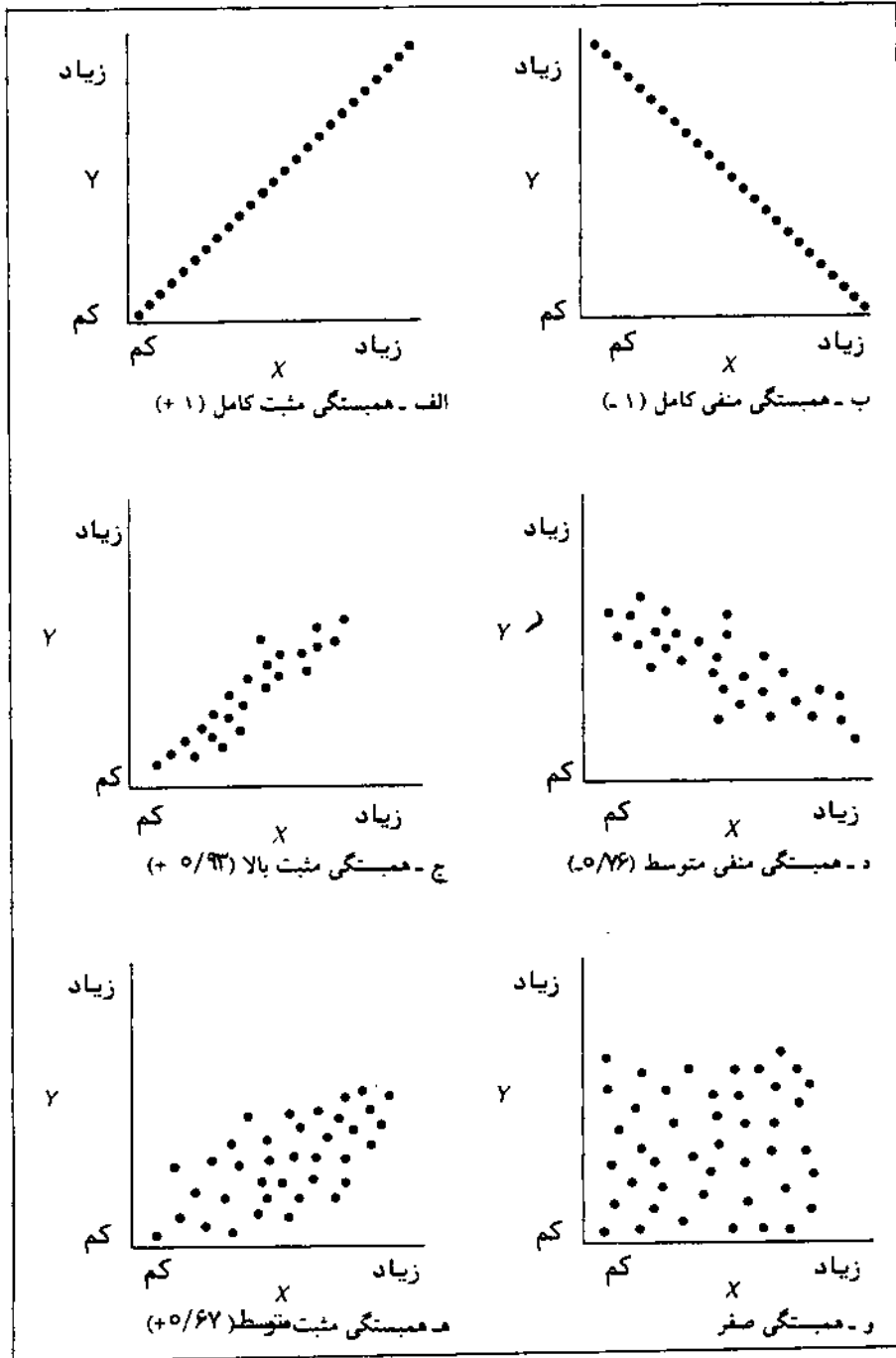
$$\bar{X} = \frac{\sum X}{N} = \frac{۲۱۰}{۱۴} = ۱۵$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{۲۳۶}{۱۴} = ۱۷$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{۵۶}{۱۴}} = ۲$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{N}} = \sqrt{\frac{۲۲۴}{۱۴}} = ۴$$

شکل ۸-۵ نمودار پراکندگی مقادیر انتخاب شده r



هنگامی که با تعداد زیادی آزمودنی سروکار داشته باشیم، فرآیند برگرداندن نمره‌های خام به نمره‌های  $Z$  کاری خسته کننده خواهد بود. این امکان وجود دارد که این مرحله را حذف کنیم و با استفاده از فرمول محاسباتی زیر، که از نظر ریاضی هم ارزش فرمول ۱۴ - ۵ است، مستقیماً با نمره‌های خام عمل کنیم:

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[ \sum X^2 - \frac{(\sum X)^2}{N} \right] \left[ \sum Y^2 - \frac{(\sum Y)^2}{N} \right]}} \quad (۵-۱۵)$$

$r$  = پیرسون  $r$

$\sum X$  = حاصل جمع نمره‌ها در توزیع  $X$

$\sum Y$  = حاصل جمع نمره‌ها در توزیع  $Y$

$\sum XY$  = مجموع حاصلضرب نمره‌های  $X$  و  $Y$  زوج شده

$\sum X^2$  = مجموع نمره‌های مربع در توزیع  $X$

$\sum Y^2$  = مجموع نمره‌های مربع در توزیع  $Y$

$N$  = تعداد نمره‌های  $X$  و  $Y$  زوج شده (آزمودنیها)

اکنون بیاید با استفاده از همان نمره‌های خام قبلی، فرمول ۱۵ - ۵ را برای محاسبه ضریب همبستگی گشتاوری پیرسون به کار ببریم. اعداد و محاسبه‌های لازم در جدول ۷-۵ داده شده است. با گذاشتن مقادیر این جدول در فرمول ۱۵ - ۵، می‌توانیم  $r$  پیرسون را محاسبه کنیم:

توجه کنید که با کاربرد این فرمول نه تنها برگردان نمره‌ها به مقادیر  $Z$  ضرورتی ندارد، بلکه همچنین محاسبه میانگینها و انحراف معیارهای دو توزیع نیز حذف شده است.

ضریب همبستگی گشتاوری به خانواده آماری میانگینها تعلق دارد. در محاسبه آن، اندازه هر یک از نمره‌ها در هر دو توزیع  $X$  و  $Y$ ، به حساب می‌آید، این ضریب، نظیر میانگین و انحراف معیار، یک آماره فاصله‌ای است که می‌توان آن را با داده‌های نسبی نیز مورد استفاده قرار داد.

مفروضه‌ای که در پس ضریب همبستگی گشتاوری وجود دارد، این است که رابطه بین دو متغیر ( $X$  و  $Y$ ) یک رابطه خطی است، یعنی یک خط مستقیم، توضیحی منطقی از رابطه یک

متغیر با متغیر دیگر به دست می‌دهد. اگر برای توضیح این رابطه به یک خط منحنی نیاز باشد، گفته می‌شود که یک رابطه منحنی الخط وجود دارد.

یک راه عملی برای دریافتن اینکه آیا رابطه بین دو متغیر خطی است یا منحنی الخط، بررسی نمودار پراکندگی داده‌ها است. شکل ۹-۵ دو نمودار را نشان می‌دهد، یکی از آنها (A) یک رابطه خطی را می‌نمایاند و دیگری (B) معرف رابطه منحنی الخط است.

اگر رابطه بین دو متغیر منحنی الخط باشد، محاسبه  $r$  پیرسون منتج به برآورد پائینی از رابطه بین دو متغیر خواهد شد که گمراه‌کننده است. در چنین حالتی باید شاخص دیگری، نظیر نسبت همبستگی ( $\eta$ ) به کار رود. بحث دربارهٔ نسبت همبستگی را می‌توان در انتهای این فصل یافت.

جدول ۷-۵ محاسبه  $r$  پیرسون با استفاده از فرمول نمره‌های خام

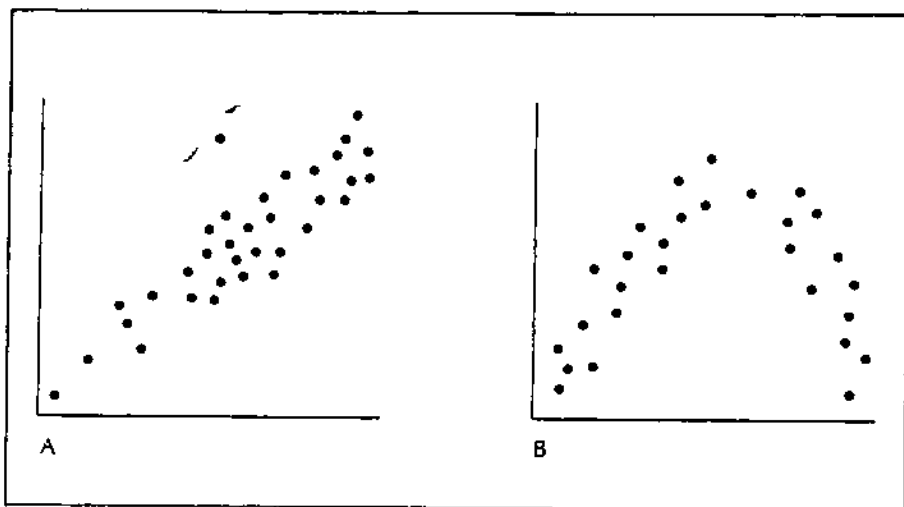
(۱)	(۲)	(۳)	(۴)	(۵)	(۶)
آزمودنیها	$X$	$Y$	$X^2$	$Y^2$	$XY$
۱	۱۸	۲۸	۳۲۴	۷۸۴	۵۰۴
۲	۱۸	۳۰	۳۲۴	۹۰۰	۵۴۰
۳	۱۷	۳۰	۲۸۹	۹۰۰	۵۱۰
۴	۱۷	۲۶	۲۸۹	۶۷۶	۴۴۲
۵	۱۶	۲۸	۲۵۶	۷۸۴	۴۴۸
۶	۱۶	۲۴	۲۵۶	۵۷۶	۳۸۴
۷	۱۵	۲۲	۲۲۵	۴۸۴	۳۳۰
۸	۱۵	۲۰	۲۲۵	۴۰۰	۳۰۰
۹	۱۴	۲۶	۱۹۶	۶۷۶	۳۶۴
۱۰	۱۴	۲۲	۱۹۶	۴۸۴	۳۰۸
۱۱	۱۳	۲۴	۱۶۹	۵۷۶	۳۱۲
۱۲	۱۳	۱۸	۱۶۹	۳۲۴	۲۳۴
۱۳	۱۲	۲۰	۱۴۴	۴۰۰	۲۴۰
۱۴	۱۲	۱۸	۱۴۴	۳۲۴	۲۱۶
$N = 14$	$\Sigma X = 210$	$\Sigma Y = 336$	$\Sigma X^2 = 3206$	$\Sigma Y^2 = 8288$	$\Sigma XY = 5132$

$$r = \frac{5132 - \frac{(210)(336)}{14}}{\sqrt{\left[3206 - \frac{(210)^2}{14}\right] \left[8288 - \frac{(336)^2}{14}\right]}} = +.82$$

// تفسیر  $r$  پیرسون

هنگامی که دو متغیر در جهت مثبت رابطه بالایی با هم داشته باشند، همبستگی بین آنها به سوی +۱ میل می‌کند. زمانی که این دو متغیر در جهت منفی با هم رابطه بالایی داشته باشند، همبستگی آنها به سوی -۱ میل می‌کند. هنگامی که رابطه کمی بین متغیرها وجود داشته باشد، همبستگی به سمت صفر میل می‌کند. بنابراین  $r$  پیرسون شاخص بامعنایی برای نمایاندن رابطه بین متغیرها فراهم می‌آورد، که علامت ضریب همبستگی جهت رابطه را نشان می‌دهد و تفاوت بین ضریب همبستگی و صفر، درجه رابطه را مشخص می‌کند.

شکل ۹-۵ همبستگی خطی (A) و منحنی الخط (B)



به هر روی، در تفسیر ضریب همبستگی باید نکات زیر را به خاطر داشت:

- ۱- همبستگی لزوماً علیت را نشان نمی‌دهد. هنگامی که در می‌یابیم دو متغیر با هم همبستگی دارند، این امر گویای آن است که جایگاه نسبی در یک متغیر با جایگاه نسبی در متغیر

دیگر همبسته است. این موضوع لزوماً به معنای آن نیست که تغییر در یک متغیر باعث تغییر در متغیر دیگر می‌شود. ما در مثال خود همبستگی  $+0.82$  را بین نمره‌های آزمون آمار توصیفی و نمره‌های آمار استنباطی یافتیم. این ضریب همبستگی به ما می‌گوید فردی که در یکی از این دو آزمون نمره بالای متوسط می‌گیرد، احتمالاً در آزمون دیگر نیز نمره بالای متوسط کسب خواهد کرد. ما نمی‌توانیم بگوییم که عملکرد عالی در یک آزمون موجب عملکرد عالی در آزمون دیگر می‌شود. نمره‌های هر دو آزمون ممکن است نتیجه دلایل دیگری، نظیر استعداد ریاضی افراد شرکت کننده در آزمون باشد.

برای روشن شدن موضوع از مثال دیگری استفاده می‌کنیم، فرض کنید که ما همبستگی مثبت بالایی بین ثروت خانوادگی و هوش کودکان آن خانواده‌ها به دست آورده‌ایم. چنین همبستگی بالایی به هیچ وجه نمایانگر یک رابطه علت و معلولی بین این دو متغیر نیست. برای این افراد، نه ثروت لزوماً هوش زیاد خلق می‌کند و نه هوش لزوماً ایجاد ثروت می‌کند.

۲- اندازه همبستگی، تا حدی، تابعی از تغییر پذیری دو توزیعی است که می‌خواهیم همبستگی آنها را تعیین کنیم. محدود کردن دامنه نمره‌هایی که قرار است با هم همبسته شوند، درجه رابطه مشاهده شده بین دو متغیر را کاهش می‌دهد. برای مثال، مشاهده شده است که موفقیت در بازی بسکتبال، با قد رابطه دارد، یعنی می‌توانیم بگوییم، هر اندازه قد کسی بلندتر باشد، احتمال بیشتری وجود دارد که این ورزش را بهتر انجام دهد. این حکم در مورد یک جمعیت بزرگ، که در آن تنوع قد وجود دارد، صادق است؛ هر چند، در یک تیم بسکتبال که همه اعضای آن بلند قد هستند، ممکن است هیچ رابطه‌ای بین قد و موفقیت نباشد، یا رابطه کمی وجود داشته باشد؛ زیرا دامنه قد در اعضای آن تیم محدود شده است.

در دانشکده‌ای که دانشجویان خود را بر مبنای دامنه گسترده نمره‌های آزمون استعداد تحصیلی<sup>(۱)</sup> می‌پذیرد، انتظار خواهیم داشت بین نمره‌های این آزمون و نمره‌های دروس دانشگاهی رابطه وجود داشته باشد. اما در دانشکده‌ای که تنها با نمره‌های بسیار بالای آزمون استعداد تحصیلی دانشجو می‌پذیرد، انتظار همبستگی بسیار کمی بین نمره‌های آزمون و نمره‌های دانشگاهی خواهیم داشت؛ زیرا در این مورد دامنه نمره‌های آزمون محدود شده است.

۳- در تفسیر ضریب همبستگی نباید آن را درصدی از همبستگی کامل محسوب کرد. از

آنجا که ضرایب همبستگی به شکل اعداد اعشاری بیان می‌شوند، افرادی که تبحری در علم آمار ندارند، گاهی ضرایب همبستگی را به عنوان درصدی از همبستگی کامل تفسیر می‌کنند. یک ضریب همبستگی  $0/8$  نمایانگر  $80$  درصد رابطه کامل بین دو متغیر نیست. این تفسیر نادرست است؛ زیرا برای مثال، یک ضریب همبستگی  $0/8$  رابطه‌ای را که دو برابر یک ضریب همبستگی  $0/4$  است، بیان نمی‌کند. یک شیوه برای تشخیص درجه‌ای که می‌توان با آن یک متغیر را از روی متغیر دیگر پیش‌بینی کرد، محاسبه شاخصی به نام ضریب تعیین<sup>(۱)</sup> است. ضریب تعیین مربع ضریب همبستگی است. احتمالاً بهترین شیوه برای معنابخشی به اندازه ضریب همبستگی ترسیم درجه پراکندگی اندازه‌های متفاوت همبستگی (همان طور که در شکل ۸-۵ نشان داده شده است) و آشنا شدن با اندازه همبستگی‌هایی است که عموماً بین متغیرهای مورد نظر مشاهده می‌شود.

۴- در تفسیر ضریب همبستگی از مطلق نگری اجتناب کنید. در تفسیر درجه همبستگی، هدف مورد نظر را به خاطر داشته باشید. برای مثال، ضریب همبستگی  $0/5+$  ممکن است برای پیش‌بینی عملکرد آتی یک گروه از افراد کافی باشد، اما امکان دارد که استفاده از همین ضریب همبستگی برای پیش‌بینی عملکرد یک فرد در آینده، کار عاقلانه‌ای نباشد. پس ضریب  $0/5+$  یک مقدار مطلق با مفهوم یکسان در هر دو حالت نیست.

### // ضریب همبستگی رتبه‌ای

ضریب همبستگی گشتاوری پیرسون یک شاخص آماری است که برای یافتن رابطه بین دو مجموعه داده فاصله‌ای که به صورت خطی توزیع شده باشند، به کار می‌رود. ما گاهی در تحقیق می‌خواهیم ضریب همبستگی بین دو مجموعه از اندازه‌ها را که به صورت رتبه‌ای منظم شده‌اند، پیدا کنیم، یعنی به جای داده‌های فاصله‌ای با داده‌های رتبه‌ای سروکار داریم.

برای مثال، فرض کنید ما می‌خواهیم همبستگی رتبه‌هایی را که دو آموزگار در زمینه خلاقیت به یک گروه از دانش‌آموزان تخصیص داده‌اند، محاسبه کنیم. شاخصی که در چنین مواردی به کار می‌رود، ضریب همبستگی رو (رتبه) اسپیرمن ( $\rho$ )<sup>(۲)</sup> است که با فرمول زیر محاسبه می‌شود.

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \quad (5-16)$$

$\rho$  = ضریب همبستگی رو اسپیرمن

$\sum D^2$  = مجموع مربعات تفاوت‌های بین رتبه‌ها

$N$  = تعداد موردها

برای روشن شدن مطلب، جدول ۵-۸ را که در آن رتبه‌بندی ۱۱ دانش‌آموز توسط دو آموزگار نشان داده شده است، در نظر بگیرید. ستونهای (۲) و (۳) جدول ۵-۸ به ترتیب نشان دهنده رتبه‌بندی آموزگار ۱ و آموزگار ۲ است ستون (۴) تفاوت بین این رتبه‌بندی‌ها را نشان می‌دهد. برای مثال، تفاوت بین رتبه‌بندی آموزگاران در مورد دانش‌آموز A، ۳-، در مورد دانش‌آموز B، ۱- و الی آخر است. مجموع مقادیر این ستون همیشه صفر است. در ستون (۵) مربع این تفاوتها داده شده است. مجموع مقادیر  $D^2$ ، ۲۶ و تعداد موردها ۱۱ است. هنگامی که این مقادیر در فرمول ۵-۱۶ جایگزین شوند، ضریب همبستگی رتبه‌ای  $+0/88$  به دست می‌آید.

جدول ۵-۸ محاسبه ضریب همبستگی بین دو مجموعه از رتبه‌ها

(۱)	(۲)	(۳)	(۴)	(۵)
دانش‌آموز	رتبه‌بندی آموزگار اول $R_1$	رتبه‌بندی آموزگار دوم $R_2$	تفاوت D	$D^2$
A	۱	۴	-۳	۹
B	۲	۳	-۱	۱
C	۳	۱	+۲	۴
D	۴	۲	+۲	۴
E	۵	۵	۰	۰
F	۶	۶	۰	۰
G	۷	۸	-۱	۱
H	۸	۹	-۱	۱
I	۹	۷	+۲	۴
J	۱۰	۱۱	-۱	۱
K	۱۱	۱۰	+۱	۱
			۰	۲۶

$$\rho = 1 - \frac{(6)(26)}{11(121 - 1)} = +0/88$$



به هنگام رتبه‌بندی افراد یا اشیاء، به منظور یافتن همبستگی بین دو مجموعه رتبه‌بندی شده، احتمال دارد که دو یا چند فرد یا شیء رتبه‌ای همسان را به خود تخصیص دهند. برای مثال، زمانی که دو نفر در رتبه سوم قرار می‌گیرند، در واقع، آنها رتبه‌های سوم و چهارم آن مجموعه هستند و از آنجا که ناممکن است که بگوییم کدام یک باید در رتبه سوم و کدامیک در رتبه چهارم قرار گیرند، ضروری است که به هر دوی آنها، متوسط این جایگاه، در این مورد  $3/5$ ، اختصاص داده شود. به نفر بعد، رتبه ۵ اختصاص خواهد یافت.

گاهی از اوقات، ما می‌خواهیم رابطه بین مجموعه‌ای از رتبه‌ها و مجموعه‌ای از اندازه‌های فاصله‌ای، نظیر دسته‌ای از نمرات امتحانی را پیدا کنیم. این امکان وجود ندارد که رتبه‌ها را از داده‌های رتبه‌ای به داده‌های فاصله‌ای ارتقاء دهیم، بنابراین، ابتدا باید نمره‌ها را به رتبه تبدیل کنیم و سپس فرمول رواسپیرمن را به کار ببریم. به عنوان مثال، جدول ۹-۵ را در نظر بگیرید.

جدول ۹-۵ محاسبه همبستگی رتبه‌ای بین یک مجموعه از داده‌های رتبه‌ای و یک مجموعه از داده‌های فاصله‌ای

(۱)	(۲)	(۳)	(۴)	(۵)	(۶)
دانش‌آموزان	رتبه‌بندی ۱	نمره‌ها	رتبه‌بندی ۲	D	D <sup>2</sup>
چک	۱	۱۹	۱	۰	۰
لیندا	۲	۱۷	۲/۵	-۱/۵	۲/۲۵
لوسی	۳	۱۸	۲	۰	۰
دیوید	۴	۱۷	۲/۵	۰/۵	۰/۲۵
تام	۵	۱۵	۳	-۱	۱
مارشا	۶	۱۴	۴	-۲	۴
جان	۷	۱۵	۳	۰	۰
دیوید	۸	۱۵	۳	۰	۰
جوآن	۹	۱۲	۵	-۱	۱
آنا	۱۰	۱۳	۴	۰	۰
جورج	۱۱	۸	۱۱	۰	۰
سر	۱۲	۵	۱۲	۰	۰
				۰	۱۵/۵

$$r = 1 - \frac{(6)(15/5)}{12(144 - 1)} = + 0/95$$

ستون (۲) جدول، پیش‌بینی یک آموزگار را در مورد رتبه‌های گروهی از دانش‌آموزان در یک امتحان، پیش از انجام آن، نشان می‌دهد. ستون (۳) نمره‌های واقعی این دانش‌آموزان را پس از امتحان نشان می‌دهد. این آموزگار برای تعیین رابطه بین رتبه‌های پیش‌بینی شده و رتبه‌های واقعی، باید نمره‌های دانش‌آموزان را به رتبه تبدیل کند. ستون (۴) رتبه‌بندی دانش‌آموزان را بر مبنای نمره‌های امتحانی آنها نشان می‌دهد. توجه داشته باشید که هم لیندا و هم دیک نمره ۱۷ را به دست آورده‌اند و بنابراین در جایگاههای سوم و چهارم قرار می‌گیرند، از این رو، به هر دوی آنها رتبه یکسان  $3/5$ ، که متوسط رتبه‌های ۳ و ۴ است، داده شده است. همین موقعیت برای تام، جان و دیوید پیش آمده است، که هر سه نفر نمره ۱۵ گرفته‌اند و در جایگاههای پنجم، ششم و هفتم سهم شده‌اند. رتبه ۶ که متوسط رتبه‌های ۵، ۶ و ۷ است، به هر یک از این سه نفر تخصیص داده شده است. طریقه یافتن  $D$  و  $D^2$  دقیقاً نظیر همان شیوه‌ای است که در جدول ۸-۵ نشان داده شد. ضریب همبستگی بین این دو مجموعه داده‌ها،  $+0/95$  است.

همبستگی رتبه‌ای اسپیرمن حالت خاصی از همبستگی گشتاوری پیرسون است و فرمول آن از فرمول ۲ پیرسون مشتق شده است. بنابراین، امکان استفاده از فرمول ۱۵-۵ برای محاسبه همبستگی بین دو مجموعه از رتبه‌ها وجود دارد. برای مثال، اگر ما فرمول ۱۵-۵ را در مورد رتبه‌های جدول ۸-۵ به کار ببریم، یک ضریب همبستگی همانند، یعنی  $+0/88$ ، به دست خواهیم آورد، اما وجود هم رتبه‌ها یک استثنا ایجاد می‌کند. با حضور هم رتبه‌ها، نظیر آنچه که در جدول ۹-۵ دیدیم، استفاده از فرمول ۲ پیرسون، دیگر یک ضریب همبستگی همانند با آنچه که از بکارگیری فرمول رتبه‌ای اسپیرمن به دست می‌آید، حاصل نخواهد کرد.

ضریب همبستگی رتبه‌ای رواسپیرمن، عضوی از خانواده آماری میانه است. این همبستگی یک آماره رتبه‌ای است که به منظور استفاده با داده‌های رتبه‌ای طراحی شده است. نظیر ضریب همبستگی گشتاوری، دامنه آن از  $-1$  تا  $+1$  تغییر می‌کند. هنگامی که هر فرد، رتبه یکسانی در هر دو متغیر داشته باشد، همبستگی رتبه‌ای  $+1$  خواهد شد و هنگامی که رتبه‌های آنان در یک متغیر، دقیقاً مخالف رتبه‌هایشان در متغیر دیگر باشد، ضریب  $-1$  خواهد شد. اگر اصلاً هیچ رابطه‌ای بین رتبه بندیها وجود نداشته باشد، ضریب همبستگی رتبه‌ای صفر خواهد شد. رواسپیرمن همسان  $T$  پیرسون تفسیر می‌شود.

## // سایر شاخصهای همبستگی

علاوه بر شاخصهای پیرسون و اسپیرمن، شاخصهای همبستگی متعدد دیگری وجود دارند که برای یافتن شدت رابطه بین انواع مختلف متغیرها مناسب هستند. بخش بعدی، برخی از این شاخصها را بدون آنکه جزئیات محاسباتی آنان را مطرح سازد، معرفی می‌کند. به دانشجویان علاقه‌مند توصیه می‌شود برای آگاهی از شیوه‌های محاسباتی به دیگر کتابهای آماری مراجعه کنند.<sup>(۱)</sup>

## // همبستگی دو رشته‌ای و همبستگی دو رشته‌ای نقطه‌ای

روشهای همبستگی دو رشته‌ای و دو رشته‌ای نقطه‌ای برای مواقعی تهیه شده‌اند که می‌خواهیم رابطه بین یک متغیر فاصله‌ای یا نسبی پیوسته و یک متغیر اسمی دو مقوله‌ای را پیدا کنیم. استفاده از همبستگی دو رشته‌ای مستلزم این مفروضه است که متغیر دو مقوله‌ای ما، پیوسته و نرمال باشد. به سخن دیگر، دو مقولگی به طور تصنعی از یک متغیر پیوسته خلق شده است. برای مثال، می‌توان افراد را از نظر خلاقیت، بر مبنای نمره‌های یک آزمون خلاقیت، به زیر متوسط و بالای متوسط طبقه‌بندی کرد. اگر نمره‌های آزمون هوش (متغیر پیوسته) را با خلاقیت، که به شکل زیر متوسط و بالای متوسط اندازه‌گیری شده است (دو مقولگی مصنوعی)، همبسته کنیم، ضریب همبستگی دو رشته‌ای، اندازه مناسبی برای تعیین رابطه آن دو خواهد بود.

همبستگی دو رشته‌ای نقطه‌ای هنگامی به کار می‌رود که دو مقولگی، امری واقعی باشد. برای مثال، هنگامی که رابطه بین جنس و نمره را در یک آزمون استدلال مطالعه می‌کنیم، به دو مقوله، مقادیر عددی ۱ یا ۰ تخصیص می‌یابد و این مقادیر با مقادیر مربوط به متغیر پیوسته همبسته می‌شوند. دیگر متغیرهای دو مقوله‌ای واقعی عبارتند از: تبعه / غیر تبعه ایالات متحده، چپ دست / راست دست، فارغ‌التحصیل دانشگاه / دانشجوی، سیگاری / غیر سیگاری و متغیرهایی نظیر آنها. همبستگی دو رشته‌ای نقطه‌ای، ساده‌سازی ریاضی ۲ پیرسون است.

۱ - برای بحث بیشتر و جزئیات محاسباتی مراجعه کنید به

G.A. Ferguson (1981). *Statistical Analysis In Psychology and Education* (5th ed.). New York: McGraw - Hill. (2) A.L. Edwards (1976). *An Introduction to Linear Regression and Correlation*. San Francisco. (3) J.D. Wynne (1982). *Learning Statistics*. New York: Macmillan.

// همبستگی چهارخانه‌ای<sup>(۱)</sup> و ضریب فی<sup>(۲)</sup>

همبستگی چهارخانه‌ای و ضریب فی برای یافتن رابطه بین متغیرها، در مواردی که هر دو متغیر دو مقوله‌ای هستند، به کار می‌روند. ماهیت دو مقولگی، شاخصی را که باید مورد استفاده قرار گیرد، تعیین می‌کند. اگر هر دو متغیر، به طور واقعی پیوسته باشند، اما به طور مصنوعی به دو مقوله تبدیل شوند، شاخص مناسب، ضریب همبستگی چهارخانه‌ای است. اگر کسی بخواهد همبستگی بین دو متغیر دو مقوله‌ای واقعی را پیدا کند، ضریب فی، شاخصی است که باید مورد استفاده قرار گیرد. برای مثال، همبستگی چهارخانه‌ای برای یافتن رابطه بین خلاقیت و هوش، هنگامی که هر دو متغیر به دو مقوله زیر متوسط و بالای متوسط طبقه‌بندی شده‌اند، به کار خواهد رفت. این شاخص همبستگی، به طور گسترده مورد استفاده قرار نمی‌گیرد. با تبدیل متغیرهای پیوسته به متغیرهای دو مقوله‌ای، اطلاعات زیادی از دست می‌رود. ضریب فی برای توصیف رابطه بین جنس یک گروه از دانش‌آموزان دبیرستان و اینکه آیا آنها کمک هزینه تحصیلی دانشکده را به دست خواهند آورد یا نه، به کار خواهد رفت. جنس، به صورت مرد - زن و کسب کمک هزینه تحصیلی، به صورت بله - خیر، به دو مقوله تبدیل می‌شوند که هر دو، مقوله‌ای واقعی هستند. مقادیر عددی (۱ و ۰) به مقوله‌ها تخصیص می‌یابند و ضریب فی محاسبه می‌شود. ضریب فی نیز ساده‌سازی ریاضی ۲ پیرسون است. جدول ۱۰ - ۵ ضرایب همبستگی را که با انواع مختلف مقیاسها به کار می‌روند، خلاصه می‌کند.

// نسبت همبستگی<sup>(۳)</sup>

خطی بودن رابطه بین دو متغیر، مفروضه‌ای است که پشتوانه ضریب همبستگی گشتاوری می‌باشد. به هر روی، گاهی رابطه بین متغیرها منحنی الخط است، یعنی این رابطه به جای خط راست، با یک خط منحنی توصیف می‌شود. در چنین مواردی، کاربرد فرمول ۲ پیرسون مناسب نیست؛ زیرا برآورد پایینی از درجه رابطه را به دست می‌دهد. هنگامی که رابطه بین X و Y منحنی الخط باشد، شاخص دیگری به نام نسبت همبستگی، اتا<sup>(۴)</sup>، مورد استفاده قرار می‌گیرد. این شاخص، شامل تصحیحی برای غیرخطی بودن است و بنابراین برآورد دقیق‌تری از مقدار

1- Tetrachoric Correlation

2- Phi coefficient

3- Correlation Ratio

4- Eta

وابستگی بین متغیرها ارائه می‌کند.

برای مثال، فرض کنید کسی می‌خواهد همبستگی بین سن و قدرت بدنی را پیدا کند. ما می‌دانیم که قدرت بدنی تا نقطه معینی با سن رابطه مثبت دارد. از آن پس، افزایش سن با کاهش قدرت بدنی همراه می‌شود. نسبت همبستگی برای استفاده در چنین مواردی شاخص مناسبی خواهد بود.

### // همبستگی سهمی و همبستگی چند متغیری

روشهای همبستگی که تاکنون مورد بحث قرار گرفتند، تنها برای استفاده با دو متغیر مناسب هستند. به هر روی، مواردی پیش می‌آیند که ما با بیش از دو متغیر سروکار داریم. همبستگی سهمی روشی است که با استفاده از آن تعیین می‌کنیم وقتی اثر متغیر سوم حذف شود، چه همبستگی‌ای بین دو متغیر دیگر باقی می‌ماند. همبستگی می‌تواند بین دو متغیر واقع شود؛ زیرا هر دوی آنها با متغیر سوم همبستگی دارند. همبستگی سهمی این متغیر سوم را کنترل می‌کند. برای مثال، فرض کنید کسی به همبستگی بین سن عقلی و مهارتهای روانی - حرکتی علاقه‌مند

### جدول ۱۰ - ۵ انواع ضرایب همبستگی و انواع مقیاسهای متناسب با آنها

نوع مقیاس	ضریب همبستگی
۱. هر دو متغیر در مقیاس فاصله‌ای یا نسبی	۱. گشاوروی پیرسون
۲. هر دو متغیر در مقیاس اسمی	۲. رتبه‌ای اسپیرمن
۳. یک متغیر در مقیاس فاصله‌ای، دیگری یک متغیر دو مقوله‌ای واقعی در مقیاس اسمی	۳. در رتبه‌ای نقطه‌ای
۴. یک متغیر در مقیاس فاصله‌ای یا نسبی، متغیر دیگر مقوله‌ای مصنوعی (۱)	۴. در رتبه‌ای
۵. دو مقولگی مصنوعی (مقیاس اسمی) مورد استفاده در هر دو متغیر، هر دو دارای توزیع پیوسته زیربنایی	۵. چهارخانه‌ای
۶. هر دو متغیر دو مقوله‌ای واقعی (مقیاس اسمی)	۶. ضریب فی

۱ - یک دو مقولگی مصنوعی، متغیر پیوسته را با استفاده از یک نقطه انقطاع، به طور اختیاری، به دو طبقه تقسیم می‌کند. برای مثال، نمره‌های آزمونی که به دو طبقه قبول و رد تقسیم شده‌اند. مثالهای دو مقولگی واقعی عبارتند از مرد - زن و زنده - مرده.

شده است. هر دوی این متغیرها به متغیر سوم، یعنی سن زمانی، وابسته‌اند. برای مثال، کودکان ۱۲ ساله، به طور کلی، دارای بلوغ عقلی بیشتری از کودکان ۸ ساله‌اند و همچنین آنها دارای مهارت‌های بدنی رشد یافته‌تری هستند. نمره‌های یک آزمون سن عقلی و یک آزمون روانی - حرکتی، با یکدیگر همبستگی خواهند داشت؛ زیرا هر دوی آنها با سن زمانی همبستگی دارند. در کار با چنین داده‌هایی، برای به دست آوردن یک اندازه‌ای از همبستگی که در آن اثر سن حذف شده است، از همبستگی سهمی استفاده می‌شود. همبستگی باقیمانده بین دو متغیر، پس از حذف همبستگی آن دو با متغیر سوم، یک همبستگی سهمی مرتبه اول نامیده می‌شود. از همبستگی سهمی می‌توان برای حذف اثر بیش از یک متغیر استفاده کرد؛ هر چند که از این روش، به دلیل تفسیر دشوار آن، اغلب استفاده نمی‌شود.

همبستگی چند متغیری نیز بیش از دو متغیر را در بردارد. این روش، محقق را قادر می‌سازد که بهترین شکل وزن دادن به دو یا چند متغیر مستقل را بیابد تا همبستگی بیشینه (ماکزیمم) را با یک متغیر وابسته به دست آورد. برای مثال، محققان برای پیش‌بینی معدل دوره دانشگاه از رتبه صدکی دبیرستان (HSPR)<sup>(۱)</sup>، نمره‌های آزمون استعداد تحصیلی (SAT) و نمره‌های آزمون شخصیت استفاده کرده‌اند. نتایج حاصله نشان داده‌اند که (۱) HSPR بهترین پیش‌بین منفرد است، (۲) ترکیب وزنی مناسبی از HSPR و SAT، از هر یک از این دو متغیر مستقل، پیش‌بین بهتری برای معدل دوره دانشگاه است، (۳) نمره‌های آزمون شخصیت چیز مفیدی به اعتبار پیش‌بینی ترکیب HSPR-SAT اضافه نمی‌کند.<sup>(۴)</sup> همبستگی چند متغیری با جزییات بیشتر در فصل ۱۱ مورد بحث قرار خواهد گرفت.

## // محاسبه گرها و برنامه‌های کامپیوتری برای آمار

بسیاری از ماشین‌حسابهای جیبی حاوی برنامه‌های از پیش آماده شده برای روشهای آماری هستند که در این فصل و فصل بعد توصیف شده‌اند. توصیف نرم‌افزارهای از پیش آماده شده برای کامپیوترها و ریز کامپیوترها در فصل ۱۵ آمده است.

1- High School Percentile Rank

۲ - یک بحث کامل درباره همبستگی چند متغیری را می‌توان در این کتاب یافت:

E. J. Pedhazur (1982), *Multiple Regression in Behavioral Research* (2d ed.), New York: Holt, Rinehart and Winston.

### /// فرا تحلیل

فرا تحلیل روشی برای ترکیب نظام‌مند داده‌های کمی حاصل از چند بررسی است که بر پرسشی یکسان متمرکز شده و از متغیرهای مشابهی استفاده کرده‌اند. گلاس، این روش را تحلیل تحلیلی تعریف کرده است.<sup>(۱)</sup> با استفاده از آماره‌های بررسیهای اصلی، که از مقایسه گروههای آزمایشی دریافت‌کننده تیمار و گروههای کنترل به دست می‌آید، کمیّت مشترکی استخراج می‌شود که آن را برای بررسی مورد نظر، یک بیان کلی نتایج، تفسیر می‌کنند.

برای هر بررسی تفاوت بین میانگین گروههای کنترل و آزمایشی را با تقسیم آن به انحراف معیار گروه کنترل، به واحدهای انحراف معیار تبدیل می‌کنند.

این نسبت در واحدهای انحراف معیار، اندازه تأثیر<sup>(۲)</sup> نامیده می‌شود.

$$ES = \frac{\bar{X}_E - \bar{X}_C}{\sigma_C} \quad (5-17)$$

ES = اندازه تأثیر برآورد شده

$\bar{X}_E$  = میانگین گروه آزمایشی

$\bar{X}_C$  = میانگین گروه کنترل

$\sigma_C$  = انحراف معیار گروه کنترل

اگر فرض کنیم که توزیع مقادیر اندازه تأثیر نرمال است، پس می‌توان هر مقدار را به عنوان یک نمره z با میانگین صفر و انحراف معیار ۱ تلقی کرد، که نقطه مرجع، یا نقطه صفر آن،  $\bar{X}_C$  است. متوسط مقادیر اندازه تأثیر همه بررسیها، تأثیر کلی برای متغیرهای آزمایشی را توصیف می‌کند.

برای مثال، اسمیت و گلاس از فراتحلیل برای بررسی این پرسش رایج که آیا روان‌درمانی تغییری ایجاد می‌کند، استفاده کردند. مرور پیشینه تحقیقاتی روشن ساخت که در این زمینه ۱۰۰۰ آزمایش وجود دارد. بررسیهایی که برای یک تحلیل کامل مناسب بود، ۸۳۳ اندازه تأثیر را

1- Glass, G V., B. McGaw, and M.L. Smith (1981). *Meta-analysis in Social Research*. Beverly Hills: Sage Publications.

2- Effect - Size

به آنان ارائه کرد. بررسیهای انتخابی شامل استراتژیهای درمانی خود<sup>(۱)</sup>، پویا<sup>(۲)</sup>، رفتاری و انسانگرایی<sup>(۳)</sup> بود که به طور آزمایشی به متغیرهای مرتبط با نتیجه نظیر عزت نفس<sup>(۴)</sup>، سازگاری<sup>(۵)</sup>، هراس - اضطراب<sup>(۶)</sup> و عملکرد درسی<sup>(۷)</sup> وابسته بودند. متوسط اندازه تأثیر ۰/۶۸ بود، یعنی میانگین پس - از - درمان برای آزمودنیهایی که درمان دریافت کردند، مساوی بود با نمره ۰/۶۸ انحراف معیار بالای میانگین آزمودنیهایی که درمانی دریافت نکرده بودند. اسمیت و گلاس نتیجه گرفتند که نتیجه روان درمانی عبارت است از کسب بهبودی در متغیر وابسته که میزان آن، پیشرفت از میانگین به صدک هفتاد و پنجم گروه کنترل است.

فرا تحلیل را نه تنها می توان در مورد بررسیهایی که میانگینها را مقایسه کرده اند، به کار برد، بلکه از آن برای بررسیهای همبستگی، نسبت و دیگر اندازهها نیز می توان استفاده کرد. گلاس، مک گا<sup>(۸)</sup> و اسمیت تبیین کاملی از روشهای فرا تحلیل و کاربردهای آن ارائه کرده اند.<sup>(۹)</sup>

### /// خلاصه

آمار توصیفی در خدمت توصیف و خلاصه کردن مشاهدات است. انتخاب روش توصیفی بسته به این است که آماره مورد نظر در خدمت چه هدفی است و از چه مقیاس اندازه گیری برای ثبت دادهها استفاده شده است.

مقیاسهای اندازه گیری ابزار کمی کردن مشاهدات و چهار نوع هستند: مقیاسهای اسمی، مشاهدات را به طبقات نامتداخل دسته بندی می کنند، مقیاسهای رتبه ای اشیاء یا طبقات اشیاء را بر مبنای جایگاه نسبی آنها مرتب می کنند، مقیاسهای فاصله ای برای اندازه گیری از فواصل مساوی استفاده می کنند و نشان می دهند یک شخص یا شیء تا چه اندازه از کیفیت خاصی برخوردار است، مقیاسهای نسبی فواصل مساوی را برای اندازه گیری به کار می برند و نقطه صفر مطلق را مبنای اندازه گیری قرار می دهند.

هنگامی که مشاهدات کمی شدند، می توان دادهها را چه خام و چه دسته بندی شده

1- Ego

2- Dynamic

3- Humanistic

4- Self- Esteem

5- Adjustment

6- Fear- Anxiety

7- School Performance

8- McGaw

9- Smith, M. L., and G. V. Glass (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752 - 760.



به صورت توزیع فراوانی مرتب کرد و به صورت نگاره‌ای با نمودار ستونی یا نمودار چندضلعی نشان داد.

اندازه‌های گرایش مرکزی - نما، میانه و میانگین - شاخص منفردی را برای نمایاندن مقدار متوسط یک مجموعه کامل از اندازه‌ها در دسترس قرار می‌دهند. نما، که یک آماره اسمی است، ناپایدارترین و ناسودمندترین اندازه در تحقیقات آموزشی است. میانه یک آماره رتبه‌ای است که رتبه نمره‌ها را در داخل یک توزیع محسوب می‌دارد، اما مقدار نمره‌های فردی را در نظر نمی‌گیرد. میانگین، که یک آماره فاصله‌ای (یا نسبی) است، پایدارترین و رایج‌ترین شاخص گرایش مرکزی می‌باشد.

شیوه دیگر توصیف مشاهدات، نشان دادن تغییر پذیری، یا پراکندگی، مقادیر یک توزیع است. دامنه، انحراف چارکی و انحراف معیار سه شاخصی هستند که برای این منظور به کار می‌روند. دامنه، فاصله بین بیشترین و کمترین مقادیر یک توزیع را نشان می‌دهد و یک آماره اسمی است.

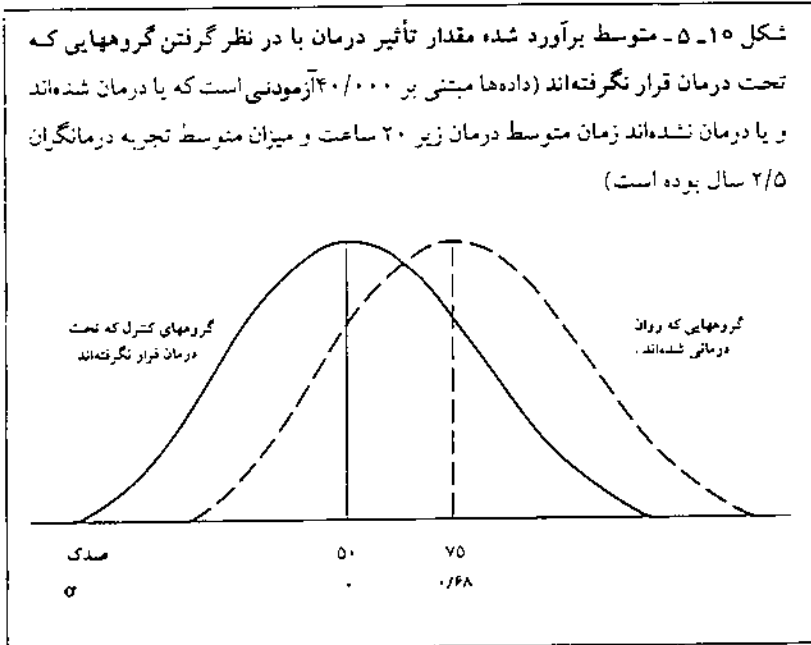
انحراف چارکی نیمه فاصله بین چارکهای بالاتر و پایین‌تر را به ما می‌دهد. انحراف چارکی یک آماره رتبه‌ای است. انحراف معیار جذر میانگین مربع مقادیر انحراف از میانگین است. انحراف معیار یک آماره فاصله‌ای (یا نسبی) و رایج‌ترین شاخص تغییر پذیری است.

نمره‌های استاندارد برای نشان دادن مکان یک نمره منفرد در یک توزیع مورد استفاده قرار می‌گیرند. رایج‌ترین نمره استاندارد، نمره  $Z$  است، که مقادیر مختلف را به واحدهای انحراف معیار تبدیل می‌کند.

با استفاده از ویژگیها و سطوح زیر منحنی نرمال، ما می‌توانیم درصد موارد پایین و بالای هر نمره  $Z$  را در یک توزیع نرمال برآورد کنیم.

روشهای همبستگی ما را قادر می‌سازد که رابطه بین دو مجموعه از اندازه‌ها را توصیف کنیم. همبستگی گشتاوری ( $r$  پیرسون) و همبستگی رتبه‌ای (رواسپیرمن) دو شاخص رایج رابطه هستند.  $r$  پیرسون با داده‌های فاصله‌ای یا نسبی مورد استفاده قرار می‌گیرد، و برای داده‌های رتبه‌ای، رواسپیرمن برای یافتن رابطه بین دو مجموعه از رتبه‌ها به کار می‌رود. جدول ۱۱ - ۵ این آماره‌ها را خلاصه می‌کند.

شکل ۱۰-۵



جدول ۱۱-۵ خلاصه آماره‌های توصیفی ارائه شده در این فصل

فاصله‌ای	رتبه‌ای	اسمی	
میانگین	میان	نما	شاخصهای گرایش مرکزی
واریانس و انحراف معیار	انحراف چارکی	دامنه	شاخصهای تغییر پذیری
نمره Z، نمره Z سایر نمره‌های استاندارد	رتبه صدکی	برجسب یا طبقه‌بندی	شاخصهای مکان
I پیرسون	رواسپیرمن	فی	شاخصهای همبستگی

// مفاهیم اصلی

انحراف چارکی	آمار توصیفی	آمار استنباطی
(دامنه نیمه چارکی)		
توزیع متقارن	توزیع فراوانی	تغییر پذیری
رابطه منحنی الخط	رابطه خطی	دامنه

ضریب همبستگی	ضریب گشتاوری پیرسون	ضریب فی
گرایش مرکزی	فراوانی تراکمی	فرا تحلیل
مقیاس فاصله‌ای	مقیاس رتبه‌ای	مقیاس اسمی
منحنی باچولگی منفی	منحنی باچولگی مثبت	مقیاس نسبی
میان	میانگین	منحنی نرمال
نمره استاندارد	نما	نسبت همبستگی (اتا)
نمودار پراکنندگی	نمره Z	نمره Z
واریانس	نمودار ستونی	نمودار چند ضلعی فراوانی
همبستگی دورشته‌ای	همبستگی چهار خانه‌ای	همبستگی
همبستگی مثبت	همبستگی رتبه‌ای اسپیرمن	همبستگی دورشته‌ای نقطه‌ای
		همبستگی منفی

### /// تمرینها

۱- در مورد هر یک از جملات زیر، نوع مقیاس اندازه‌گیری - اسمی، رتبه‌ای، فاصله‌ای، یا نسبی - را مشخص کنید:

الف - جان، آزمون ریاضی را در ۳۵ دقیقه تمام کرد، در حالی که جک، همان آزمون را در ۲۵ دقیقه به پایان رساند.

ب - جک به زبان فرانسه صحبت می‌کند، اما جان صحبت نمی‌کند.

ج - جک بلند قدرتر از جان است.

د - بلندی قد جان ۶ فوت و ۲ اینچ است.

ه - IQ جان ۱۲۰ است، در حالی که IQ جک ۱۱۰ است.

۲- یک نمودار ستونی و یک نمودار چند ضلعی برای توزیع فراوانی زیر رسم کنید:

$x$	$f$	$x$	$f$	$x$	$f$	$x$	$f$
۸۰	۱	۷۶	۶	۷۳	۲۰	۷۰	۷
۷۹	۲	۷۵	۱۵	۷۲	۱۷	۶۹	۳
۷۸	۳	۷۴	۲۲	۷۱	۹	۶۸	۱
۷۷	۱۰						

۳- با در نظر گرفتن توزیع ۱۵، ۱۴، ۱۴، ۱۳، ۱۱، ۱۰، ۱۰، ۱۰، ۸، ۵:

الف - میانگین را حساب کنید.

ب - مقدار میانه را تعیین کنید.

ج - مقدار نما را تعیین کنید.

۴- به طور خلاصه رابطه بین چولگی یک توزیع نمره‌ها و مقادیر حاصله میانگین، میانه، و

نما را شرح دهید.

۵- مناسب‌ترین اندازه‌ها - نما، میانگین یا میانه - را برای هر یک از انواع مقیاسها تعیین کنید:

الف - رتبه‌ای ب - اسمی ج - فاصله‌ای

۶- تعیین کنید که هر یک از عبارات زیر، کدام اندازه - نما، میانگین یا میانه - را تعریف می‌کند:

الف - نمره وسط

ب - متوسط حسابی

ج - نمره‌ای که بیشتر از همه تکرار شده است.

۷- نمره‌های زیر نتایج آزمون وازگان ۲۰ دانش‌آموز کلاس هفتم را نشان می‌دهد. دامنه،

انحراف معیار، و انحراف چارکی را محاسبه کنید و دربارهٔ مزایا و نقایص هر یک، به عنوان یک

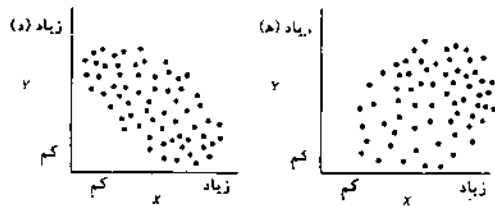
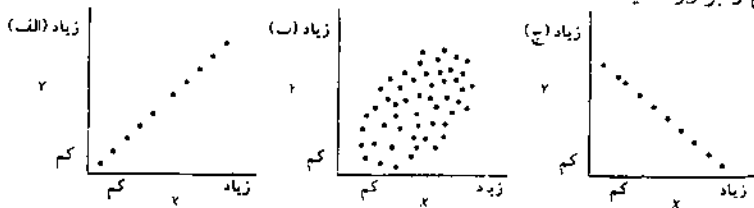
اندازه تغییر پذیری برای این نمره‌ها بحث کنید.

$x$	$f$	$fx$	$x^2$	$fx^2$
۱۶	۱	۱۶	۲۵۶	۲۵۶
۱۵	۰	۰	۲۲۵	۰
۱۴	۰	۰	۱۹۶	۰
۱۳	۰	۰	۱۶۹	۰
۱۲	۲	۲۴	۱۴۴	۲۸۸
۱۱	۰	۰	۱۲۱	۰
۱۰	۲	۲۰	۱۰۰	۲۰۰
۹	۱	۹	۸۱	۸۱
۸	۱	۸	۶۴	۶۴
۷	۱	۷	۴۹	۴۹
۶	۴	۲۴	۳۶	۱۴۴
۵	۲	۱۰	۲۵	۵۰
۴	۱	۴	۱۶	۱۶
۳	۱	۳	۹	۹
۲	۴	۸	۴	۱۶

۸- برای به حداقل رساندن اثر یک نمره انتهایی، باید انحراف چارکی را به عنوان شاخص تغییرپذیری انتخاب کرد یا انحراف معیار را؟ چرا؟  
 ۹- در یک آزمون، نمره میانگین ۴۰ و انحراف معیار ۴ است. هر یک از نمره‌های خام زیر را به یک نمره % تبدیل کنید:

الف- ۴۱ ب- ۳۰ ج- ۴۸ د- ۳۶ ه- ۴۶

۱۰- در یک توزیع نرمال چه درصدی از نمره‌ها زیر نمره ۱٪ قرار خواهند گرفت؟ چه درصدی زیر ۰٪ صفر؟ چه درصدی زیر نمره  $0.67+$ ؟  
 ۱۱- رابطه نشان داده شده به وسیله این نمودارهای پراکنندگی را توصیف کنید. سپس ضریب همبستگی را برآورد کنید.



۱۲- داده‌های زیر نمره‌های ده دانشجو در یک آزمون استدلال انتزاعی و نیز معدل آنها در رشته فلسفه است.  $\bar{x}$  پیرسون را برای این داده‌ها محاسبه کنید.

معدل	استدلال انتزاعی	دانشجو
۱/۵	۱۵	الف
۲/۵	۲۰	ب
۳	۳۰	ج
۲	۳۵	د

۳	۲۵	۵
۳/۵	۴۰	و
۴	۳۵	ز
۱	۵	ح
۲	۱۲	ط
۲/۵	۱۰	ی

۱۳- محققى به مبحث اضطراب و تأثیر آن بر عملکرد افراد در آزمونهای هوش علاقه‌مند شده است. او با کمک یک روان‌شناس بالینی اضطراب آزمودنیها را سنجیده و آنها را از ۱ تا ۲۰ رتبه بندی کرده و سپس از آزمودنیها یک آزمون هوش استاندارد شده گرفته و نمره IQ هر ۲۰ نفر را به رتبه تبدیل کرده است. این محقق برای داده‌های خود چه ضریب همبستگی‌ای را باید محاسبه کند؟ پاسخ خود را تبیین کنید.

۱۴- محققى ضریب همبستگی  $+0/6$  را بین درجه بندی مدیران مدارس در مورد نحوه لباس پوشیدن معلمان و عملکرد درسی دانش‌آموزان ۱۵۰ دبستان در ایالت خود نشان داد. او نتیجه‌گیری کرد که تشویق معلمان به آراستگی، عملکرد درسی را افزایش خواهد داد. نظر خود را درباره نتیجه‌گیری او بیان کنید.

۱۵- شاخص همبستگی مناسب را برای استفاده در حل هر یک از مسئله‌های زیر تعیین کنید:  
الف - ما می‌خواهیم ضریب همبستگی بین هوش و خلاقیت را، هنگامی که نمره‌های یک گروه از آزمودنی‌ها را در هر دو متغیر داریم، پیدا کنیم.

ب - ما می‌خواهیم ضریب همبستگی بین عملکردهای یک گروه از آزمودنیها را در دو آزمون، هنگامی که نمره‌های آزمودنیها را در یک آزمون و جایگاه آنها را به صورت بالا یا زیر میانگین در آزمون دیگر داریم، پیدا کنیم.

ج - ما می‌خواهیم همبستگی بین جنس و شرایط لازم برای یک شغل را، هنگامی که اطلاعات رد یا قبول در آزمون صلاحیت آن شغل را برای ۶۰ مرد و ۶۰ زن داریم، پیدا کنیم.

د - ما می‌خواهیم همبستگی بین پاسخهای ۱۸۰ دانشجو را به پرسشهای دو آزمون، هنگامی که پاسخهای درست و نادرست این دانشجویان را به پرسشهای هر دو آزمون در دست داریم، پیدا کنیم.

۵- ما می‌خواهیم رابطه بین هوش و انگیزه پیشرفت را، هنگامی که نمره‌های آزمون هوش ۲۰۰ آزمودنی را در آزمون هوش استانفورد - بینه و جایگاه آنان را در آزمون انگیزه پیشرفت به صورت بالا یا زیر متوسط در دست داریم، پیدا کنیم.

و - ما می‌خواهیم رابطه بین جنس و توانایی کار با ابزار را، هنگامی که نمره‌های ۶۰ پسر و ۶۰ دختر را در یک آزمون توانایی کار با ابزار داریم، پیدا کنیم.

ز - ما می‌خواهیم رابطه بین نمره‌های گروه سنجی یک گروه از دانش‌آموزان و رتبه درسی آنها را در یک کلاس دبیرستانی پیدا کنیم.

۱۶- کدام روش آماری، تحلیل تحلیلی نامیده شده است؟

### /// پاسخها

۱- الف - نسبی ب - اسمی ج - رتبه‌ای د - نسبی ه - فاصله‌ای

۲- پاسخها متفاوت است.

۳- الف - میانگین = ۱۱ ب - میانه = ۱۰/۵ ج - نما = ۱۰

۴- این سه اندازه در یک توزیع دارای چولگی مساوی نیستند. میانگین به سمت چولگی

کشیده می‌شود. بنابراین، در یک توزیع دارای چولگی مثبت، میانگین همیشه بیشتر از میانه است و نما معمولاً کمترین مقدار را دارد. در یک توزیع دارای چولگی منفی، میانگین همیشه کمتر از میانه است و نما معمولاً بیشترین مقدار را دارند.

۵- الف - میانه ب - نما ج - میانگین

۶- الف - میانه ب - میانگین ج - نما

۷- دامنه = ۱۴ - ۲ = ۱۶

$$\sigma = \sqrt{\frac{1173 - \frac{(133)^2}{20}}{20}} = \sqrt{14/4275} = 3/798$$

$$QD = \frac{Q_2 - Q_1}{2} = \frac{9/5 - 3/5}{2} = \frac{6}{2} = 3$$

۸- انحراف چارکی؛ زیرا نمره‌های انتهایی بر انحراف چارکی تأثیر نمی‌گذارند.

۹- الف- ۰/۲۵، ب- ۲/۵، ج- ۲، د- ۱، ه- ۱/۵

۱۰- ۰/۱۶، ۰/۵۰، ۰/۷۵

۱۱- الف- مثبت کامل، +۱، ب- مثبت، ۰/۷۵، ج- منفی کامل، -۱

د- منفی، -۰/۷۵، ه- بدون همبستگی، ۰

-۱۲

X	Y	XY	X <sup>2</sup>	Y <sup>2</sup>
۱۵	۱/۵	۲۲/۵	۲۲۵	۲/۲۵
۲۰	۲/۵	۵۰	۴۰۰	۶/۲۵
۳۰	۳	۹۰	۹۰۰	۹
۳۵	۲	۷۰	۱۲۲۵	۴
۲۵	۳	۷۵	۶۲۵	۹
۴۰	۳/۵	۱۴۰	۱۶۰۰	۱۲/۲۵
۳۵	۴	۱۴۰	۱۲۲۵	۱۶
۵	۱	۵	۲۵	۱
۱۲	۲	۲۴	۱۴۴	۴
۱۰	۲/۵	۲۵	۱۰۰	۶/۲۵
Σ ۲۲۷	۲۵	۶۴۱/۵	۶۴۶۹	۷۰

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[ \Sigma X^2 - \frac{(\Sigma X)^2}{N} \right] \left[ \Sigma Y^2 - \frac{(\Sigma Y)^2}{N} \right]}}$$

$$= \frac{641/5 - \frac{(227)(25)}{10}}{\sqrt{\left[ 6469 - \frac{(227)^2}{10} \right] \left[ 70 - \frac{(25)^2}{10} \right]}} = +0/74$$

۱۳- چون محقق داده‌های رتبه‌ای دارد، باید ضریب همبستگی رتبه‌ای اسپیرمن را محاسبه

کند.

۱۴- محقق هیچ توجیهی برای استنباط یک رابطه علی، صرفاً بر مبنای شواهد همبستگی

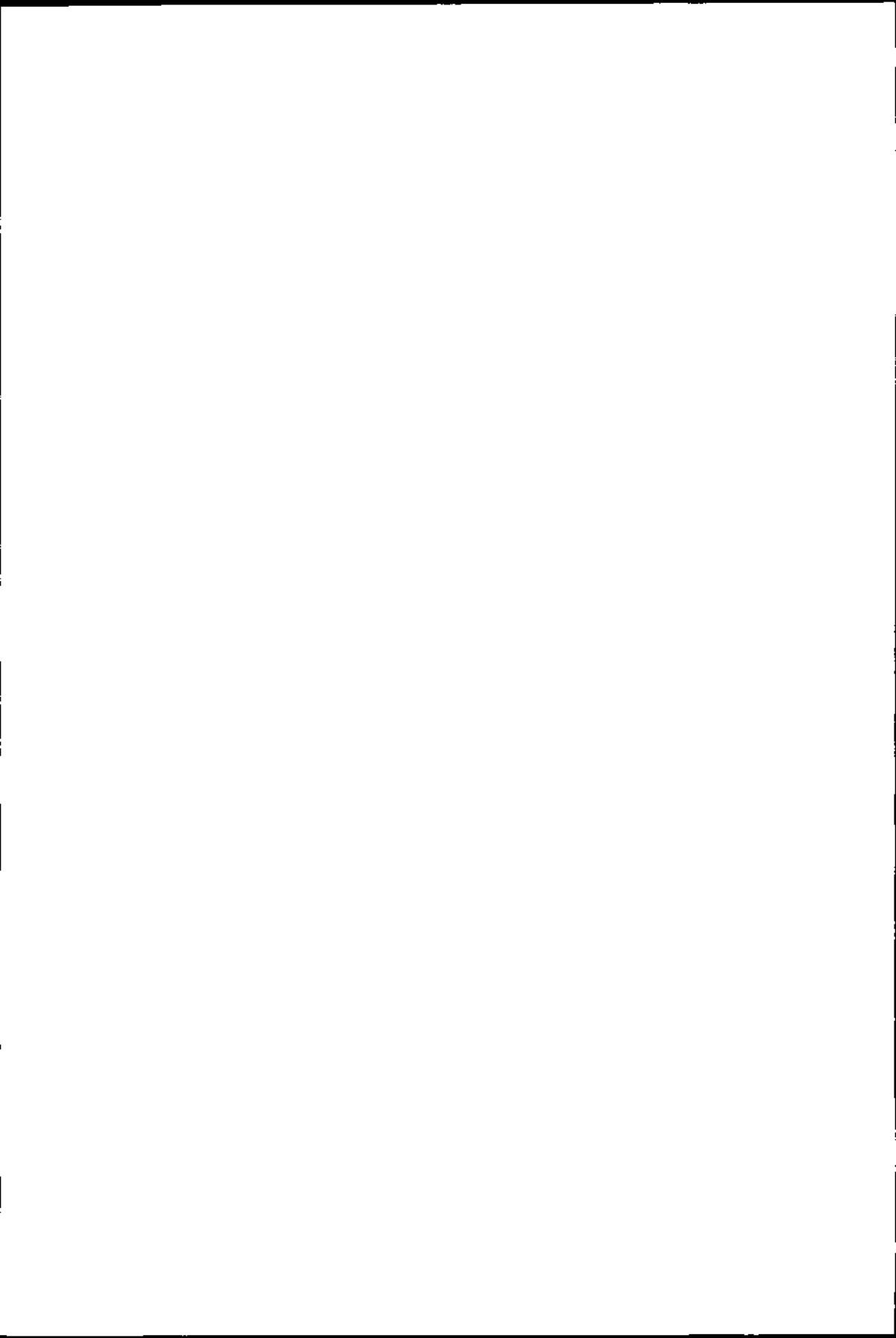


ندارد. درجه‌بندی مدیران مدارس از نحوه لباس پوشیدن معلمان و عملکرد درسی دانش‌آموزان می‌تواند تابعی از متغیرهای دیگر باشد.

۱۵- الف - ضریب گشتاوری پیرسون ب - دورشته‌ای ج - چهارخانه‌ای

د - ضریب فی ه - دورشته‌ای و - دورشته‌ای نقطه‌ای ز - رواسپیرمن

۱۶- فرا تحلیل



## فصل ششم

### نمونه گیری و آمار استنباطی

#### هدفهای آموزشی

دانشجو، بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- معنا، منطق و مراحل نمونه گیری را توصیف کند و نمونه گیری احتمالاتی را از نمونه گیری غیر احتمالاتی تشخیص دهد.
- ۲- ویژگیها، کاربردها و محدودیتهای هر یک از انواع نمونه گیری احتمالاتی و نمونه گیری غیر احتمالاتی را فهرست کند.
- ۳- معنای خطای نمونه گیری و رابطه آن را با استنباطهای آماری توضیح دهد.
- ۴- معناداری آماری را توضیح دهد.
- ۵- معنای فرضیه صفر و کاربردهای آن را در تحقیق علمی توضیح دهد.
- ۶- خطاهای نوع I و نوع II را توصیف کند.
- ۷- تفاوت بین آزمونهای معناداری جهت دار و غیر جهت دار و کاربرد مناسب هر یک از این آزمونها را توضیح دهد.
- ۸- آزمون  $t$  را برای یافتن معناداری تفاوت بین میانگینهای همبسته و غیر همبسته به کار برد.

۹- آزمون F را برای یافتن معناداری تفاوت بین گروهها در تحلیل واریانس‌های یک‌سویه و دو سویه به کار برد.

۱۰- آزمون خی‌دو را برای یافتن معناداری تفاوت بین نسبتها در طبقه‌بندیهای یک‌سویه و دو سویه به کار برد.

۱۱- استفاده از آزمون  $\chi^2$  را در مورد T پیرسون توصیف و تعیین کند که کدام ضریب همبستگی معنادار است.

۱۲- شیوه استنباط آماری مناسب برای استفاده در آزمون یک فرضیه تحقیقاتی را انتخاب کند.

۱۳- واژه‌های فنی آماری مورد استفاده در گزارش نتایج تحقیق را درک کند.

مباحث آماری که در فصل گذشته به ذکر آن پرداختیم، برای طبقه‌بندی، خلاصه‌کردن و توصیف داده‌ها مورد استفاده قرار می‌گیرد، اما در تحقیق، ما اغلب نیازمند آن هستیم که از توصیف داده‌ها فراتر رویم. بعد از انجام مشاهده‌های لازم در مورد گروه نمونه، برای آنکه یافته‌های خود را به کل جامعه آماری، که گروه نمونه از آن انتخاب شده است، تعمیم دهیم، از استقراء یا استنباط استفاده می‌کنیم. برای انجام این امر، به تکنیکهایی نیازمندیم که ما را به استنباطهای معتبر از گروه نمونه به کل جامعه آماری قادر سازد.

### /// نمونه‌گیری

یک خصیصه مهم آمار استنباطی، فرآیند از جزء به کل رسیدن است. برای مثال، ممکن است ما یک گروه ۵۰۰ نفری از دانشجویان یک دانشگاه را به منظور انجام تعمیم‌هایی درباره کل بیکره دانشجویی آن دانشگاه، مطالعه کنیم.

گروه کوچکی که مورد مشاهده قرار می‌گیرد، نمونه و گروه بزرگتر که در مورد آن تعمیم صورت می‌گیرد، جامعه آماری نامیده می‌شود. یک جامعه آماری را این گونه تعریف می‌کنند: همه اعضای یک طبقه کاملاً تعریف شده از افراد، رویدادها یا اشیاء. برای مثال، در مطالعه‌ای که نوجوانان آمریکایی جامعه آمار آن را تشکیل می‌دهند، می‌توان این جامعه آماری را به عنوان همهٔ پسران و دختران آمریکایی که در دامنه سنی ۱۲-۲۱ سال قرار دارند، تعریف کرد. یک نمونه، بخشی از یک جامعه آماری است. برای مثال، دانش‌آموزان دبیرستان واشنگتن در ایندیانا پولیس، یک نمونه از نوجوانان آمریکایی هستند. آنها بخشی از جامعه آماری نوجوانان آمریکایی‌اند؛ زیرا هم شهروند آمریکایی هستند و هم در دامنه سنی ۱۲-۲۱ سال قرار دارند.

استنباط آماری روشی است که از طریق آن پارامترها، یعنی خصیصه‌های جامعه آماری را از آماره‌ها، یعنی خصیصه‌های نمونه، برآورد می‌کنیم. چنین برآوردهایی بر قوانین احتمالات بنا شده‌اند و بیش از آنکه واقعیت مطلق باشند، بهترین برآورد ما از این واقعیتها هستند. در هر استنباطی درجه معینی از خطا نهفته است. همان گونه که بعداً خواهیم دید، آمار استنباطی، همچنین، برای آزمون فرضیه‌های مربوط به جامعه آماری به کار می‌رود.

## // منطق نمونه گیری

استدلال استقرایی، یک بخش اساسی از رویکرد علمی است. روش استقرا مستلزم انجام مشاهده‌ها و سپس استخراج نتایجی از این مشاهده‌ها است. اگر کسی بتواند همه افراد یک جامعه آماری را مورد مشاهده قرار دهد، می‌تواند با اطمینان پایه نتیجه‌گیریهای خود را درباره جامعه آماری بر این مشاهدات بنا نهد (استقرای کامل). از سوی دیگر، اگر او تنها برخی از افراد جامعه آماری را مورد مشاهده قرار دهد، نمی‌تواند چیزی بیش از استنباط این امر که این مشاهده‌ها در مورد کل جامعه آماری صدق می‌کنند، انجام دهد (استقرای ناقص). این مفهوم نمونه‌گیری است، که شامل گرفتن بخشی از جامعه آماری، انجام مشاهده‌ها بر روی این گروه کوچکتر و تعمیم یافته‌ها به جامعه مرجع، یعنی جامعه بزرگتری است که نمونه از آن گرفته شده است.

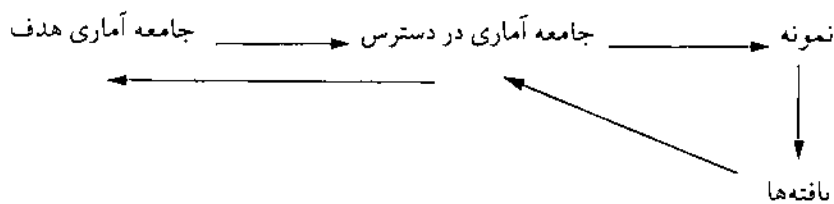
برای محقق، نمونه‌گیری یک ضرورت است. معمولاً زمان، هزینه و تلاشی که برای انجام یک تحقیق لازم است، به محقق اجازه نمی‌دهد که همه افراد ممکن یک جامعه آماری را مورد مطالعه قرار دهد. بعلاوه برای درک پدیده مورد بررسی، به طور کلی نیازی به مطالعه همه موارد ممکن نیست. نمونه‌گیری ابزاری است که به یاری محقق می‌آید و به او امکان می‌دهد که به جای مطالعه تمام جامعه آماری، تنها بخشی از آن را مورد مطالعه قرار دهد.

از آنجا که منظور از بیرون کشیدن یک نمونه از یک جامعه آماری کسب اطلاعاتی در مورد آن جامعه آماری است، کاملاً اهمیت دارد افرادی که در یک نمونه قرار می‌گیرند، برش عرضی از آن جامعه آماری را تشکیل دهند که معرف همه افراد آن باشند؛ به عبارتی برای اینکه بتوان با اطمینان، از نمونه به جامعه آماری تعمیم داد، نمونه‌ها باید معرف باشند. برای مثال، ممکن است محقق فرض کند که دانش‌آموزان دبیرستان واشنگتن معرف جوانان آمریکایی هستند. به هر ترتیب، اگر افرادی که در نمونه گنجانده شده‌اند، خصیلت‌هایی متفاوت با جامعه آماری مرجع داشته باشند، ممکن است این نمونه، معرف جامعه آماری نباشد. مکان مدرسه آنها، زمینه اقتصادی - اجتماعی آنها، موقعیت خانوادگی آنها، تجارب قبلی آنها و بسیاری دیگر از خصیلت‌های این گروه، ممکن است از آنها گروهی غیر معرف برای جوانان آمریکایی ساخته باشد. این نوع از نمونه‌ها را نمونه اریب نام می‌گذارند. یافته‌های حاصل از یک نمونه اریب، قابل تعمیم به جامعه آماری آن نیست.

### ۱/ مراحل نمونه‌گیری

اصل نخستین در نمونه‌گیری، شناسایی جامعه آماری است که در بررسی معرفی می‌شود. اگر پژوهشگری بخواهد دربارهٔ آموزگاران مدارس سنت لوتیز اطلاعاتی کسب کند، تمام کسانی که در آن مدارس تدریس می‌کنند، جامعه آماری هدف<sup>(۱)</sup> او را تشکیل می‌دهند در یک بررسی که موضوع آن نگرشها و ارزشهای جوانان آمریکایی باشد، جامعه آماری هدف، همه پسران و دختران آمریکایی در محدودهٔ سنی ۱۲-۲۱ خواهند بود؛ با این فرض که پژوهشگر، تعریف عملی جوانی را دوره سنی بین ۱۲ تا ۲۱ سال دانسته باشد.

به هر حال، از آنجا که معمولاً بررسی تمام جامعه آماری هدف امکان‌ناپذیر است، پژوهشگر باید آن بخش از جامعه آماری را که قابل دسترس است، تعیین کند. این بخش را جامعه آماری در دسترس می‌نامند و پژوهشگر نمونه‌ای را که مورد بررسی قرار می‌دهد، از این بخش انتخاب می‌کند. ماهیت جامعه آماری در دسترس از زمان و منابعی که پژوهشگر در اختیار دارد، تأثیر می‌پذیرد. در یک بررسی نگرش نوعی، برای مثال، ممکن است پژوهشگر تمام پسران و دختران کالیفرنیا را به عنوان جامعه آماری در دسترس تعیین کند یا فقط پسران و دختران سان فرانسیسکو را مورد نظر قرار دهد. پژوهشگر، از جامعه آماری در دسترس، نمونه‌ای انتخاب می‌کند که معرف آن جامعه آماری باشد. برای مثال، اگر جوانان کالیفرنیا به عنوان جامعه آماری در دسترس تعیین شده باشد، پژوهشگر از بین تمام جوانان ایالت کالیفرنیا نمونه انتخاب خواهد کرد، یا اگر جوانانی که در سان فرانسیسکو زندگی می‌کنند، به عنوان جامعه آماری در دسترس تعیین شوند، نمونه از بین این گروه خاص انتخاب خواهد شد.



یک پژوهشگر با چه اطمینانی می‌تواند یافته‌های حاصل از یک نمونه را به جامعه آماری

هدف تعمیم دهد؟ اگر نمونه انتخاب شده به راستی معرف جامعه آماری در دسترس باشد، پس برداشتن اولین گام در فرآیند تعمیم با دشواری کمی روبرو خواهد شد. اصل کلی این است: اگر یک نمونه به گونه‌ای انتخاب شده است که معرف جامعه آماری در دسترس است، یافته‌های حاصل از نمونه می‌تواند به آن جامعه آماری تعمیم یابد. برای مثال، اگر پژوهشگری یک نمونه معرف از جوانان کالیفرنیا انتخاب کرده بود، بنابراین، می‌توانست تعمیم‌هایی در مورد نگرشها و ارزشهای تمام پسران و دختران جوان کالیفرنیا انجام دهد.

به هر ترتیب، تعمیم دادن از جامعه آماری در دسترس به جامعه آماری هدف، به طور کلی، خطر<sup>(۱)</sup> بیشتری در بر دارد. در این گام، اطمینانی که پژوهشگر می‌تواند نسبت به تعمیم نتایج داشته باشد، به تشابه جامعه آماری در دسترس و جامعه آماری هدف بستگی دارد. در مثال بالا، اگر پژوهشگر به جای محدود کردن جامعه آماری در دسترس به ایالت کالیفرنیا، ایالت‌های مختلفی را در سراسر کشور به عنوان جامعه آماری در دسترس تعیین کرده بود، با اطمینان بیشتری می‌توانست نتایج حاصله را به جوانان آمریکایی تعمیم دهد. به این ترتیب تمام قسمتهای ایالات متحده در جامعه آماری جای می‌گرفت و نمونه‌گیری جامع‌تری از نگرشهای جوانان آمریکایی ممکن می‌شد.

این بیان درست است که به هنگام استفاده از مشاهده‌های یک نمونه برای برآورد خصصتهای یک جامعه آماری، محقق ناچار است که در «اعتماد» به نتایج حاصله، به «جهشی» استنباطی دست زند. احتمال درست بودن چنین استنباطهایی، تا اندازه زیادی تابع روش نمونه‌گیری است. دو شیوه اصلی نمونه‌گیری در اختیار محقق است: نمونه‌گیری احتمالاتی و نمونه‌گیری غیر احتمالاتی. نمونه‌گیری احتمالاتی نمونه‌گیری‌ای است که عناصر آن به شیوه اتفاقی برگزیده می‌شوند. ویژگی اصلی نمونه‌گیری احتمالاتی این است که هر عضو یا عنصر جامعه آماری احتمال مشخصی برای قرار گرفتن در نمونه دارد. نمونه‌گیری غیر احتمالاتی شامل روشهایی است که در آنها اعضای نمونه با شیوه‌های اتفاقی برگزیده نمی‌شوند. در این نوع نمونه‌گیری هیچ راهی وجود ندارد که بتوان احتمال قرار گرفتن یک عنصر را در نمونه مشخص کرد. موفقیت آن به دانش، تخصص و قضاوت محقق بستگی دارد. نمونه‌گیری غیر احتمالاتی هنگامی مورد استفاده قرار می‌گیرد که کاربرد نمونه‌گیری احتمالاتی امکان‌پذیر نباشد. مزایای آن، سهولت و صرفه‌جویی است.



## // نمونه‌گیری احتمالاتی

نمونه‌گیری احتمالاتی نوعی نمونه‌گیری است که در آن احتمال انتخاب هر عنصر جامعه آماری، بالاتر از صفر است.<sup>(۱)</sup> در این نوع نمونه‌گیری، امکان قرار گرفتن هر عنصر جامعه آماری در نمونه، برحسب اتفاق است و این امر از طریق انتخاب تصادفی صورت می‌گیرد. هنگامی که از نمونه‌گیری احتمالاتی استفاده می‌شود، آمار استنباطی محققان را قادر می‌سازد که تفاوت احتمالی یافته‌های مبتنی بر نمونه را با آنچه که از مطالعه تمام جامعه آماری به دست خواهد آمد، برآورد کنند. چهار نوع نمونه‌گیری احتمالاتی رایج در تحقیقات آموزشی عبارتند از: نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه‌ای، نمونه‌گیری خوشه‌ای، و نمونه‌گیری نظام‌مند (سیستماتیک).

### / نمونه‌گیری تصادفی ساده

یکی از شناخته‌شده‌ترین روشهای نمونه‌گیری، روش نمونه‌گیری تصادفی ساده است. خصلت اساسی نمونه‌گیری تصادفی ساده آن است که همه افراد جامعه آماری بخت مساوی و مستقل برای انتخاب شدن دارند، یعنی برای هر زوج  $X$  و  $Y$ ، بخت  $X$  برای انتخاب شدن مساوی بخت  $Y$  است، و انتخاب  $X$  به هیچ ترتیبی بر احتمال انتخاب  $Y$  تأثیر ندارد. مراحل نمونه‌گیری تصادفی ساده از این قرارند:

۱- تعریف جامعه آماری

۲- فهرست کردن همه اعضای جامعه آماری

۳- انتخاب نمونه با اعمال روشی که اعضای فهرست تنها بر حسب تصادف، در نمونه قرار بگیرند.

نظام‌مندترین روش برای انتخاب یک نمونه تصادفی، مراجعه به جدول اعداد تصادفی است. این جدول حاوی ستونهایی از اعداد است که برای اطمینان از ترتیب تصادفی آنها، به طور ماشینی و معمولاً با استفاده از کامپیوتر، تعیین شده‌اند. جدول ۶-۸ در ضمیمه، نمونه‌ای از جدول اعداد تصادفی است. نخستین مرحله در انتخاب یک نمونه تصادفی از یک جامعه آماری، تخصیص یک شماره شناسایی مشخص به هر عضو جامعه آماری است. سپس جدول اعداد

1- Stuart, A. (1984). *The Ideas of Sampling* (3d ed.). New York: Macmillan.

تصادفی برای انتخاب شماره شناسایی آزمودنیهایی که باید در نمونه قرار بگیرند، به کار می‌آید. اجازه دهید چگونگی استفاده از این جدول را با گرفتن نمونه‌ای از جامعه آماری دانش‌آموزان دبیرستان واشنگتن، روشن کنیم. ابتدا لازم است تمام افرادی را که در جامعه آماری قرار می‌گیرند، معین کنیم. دفتر مدیر مدرسه می‌تواند فهرستی از تمام دانش‌آموزانی که در مدرسه ثبت نام کرده‌اند، فراهم کند. سپس به منظور شناسایی هر فرد جامعه آماری شماره‌ای به هر یک تخصیص داده می‌شود. با فرض اینکه مدرسه ۸۰۰ دانش‌آموز داشته باشد، محقق می‌تواند از شماره‌های ۰۰۱، ۰۰۲، ۰۰۳، ۰۰۴، الی ۷۷۹ برای این منظور استفاده کند. سپس محقق به جدول اعداد تصادفی مراجعه می‌کند تا اعداد سه رقمی را به دست آورد، و تنها از اعدادی استفاده می‌کند که کمتر از یا مساوی با ۷۹۹ باشد. به ازای هر عدد انتخابی، فردی که در جامعه آماری با آن عدد همخوان است، در نمونه قرار می‌گیرد. محقق این فرآیند را تا آنجا ادامه می‌دهد که افراد گروه نمونه به تعداد مطلوب انتخاب شوند. مرسوم است که به هنگام استفاده از جدول اعداد تصادفی، نقطه شروع را به طور اتفاقی، انتخاب کنند. یکی از راهها آن است که چشمها را ببندند و انگشت را بر صفحه کاغذ جدول اعداد تصادفی بگذارند و از همان عدد شروع کنند.

معنای عام کلمه تصادفی<sup>(۱)</sup> عبارت است از «بدون منظور یا به طور اتفاقی». به هر ترتیب نمونه‌گیری تصادفی روشی هدفمند و مبتنی بر اسلوب معین است. بدیهی است که نمونه‌ای که به طور تصادفی انتخاب شود، از سوگیریهای پژوهشگر مبرا است. هنگامی که پژوهشگری این روش را به کار می‌برد، خود را مکلف می‌سازد که نمونه را به گونه‌ای انتخاب کند که جهت‌گیریهای او مجال بروز پیدا نکند. او تعهد می‌کند که از انتخاب آگاهانه آزمودنیهایی که فرضیه‌اش را تأیید خواهند کرد، خودداری کند. او اجازه می‌دهد که تنها تصادف تعیین کند که چه عناصری از جامعه آماری در نمونه قرار خواهند گرفت.

پژوهشگر انتظار خواهد داشت که یک نمونه تصادفی معرف جامعه آماری مرجعی باشد که نمونه از آن گرفته شده است، به هر حال، انتخاب تصادفی، بویژه در مورد نمونه‌های کوچک، تضمینی برای این نیست که نمونه، معرف جامعه آماری باشد. انتخاب تصادفی ضمانت می‌کند که هر تفاوتی بین نمونه و جامعه آماری مرجع، تنها تابعی از تصادف باشد و از جهت‌گیری محقق منتج نباشد. تفاوتی بین نمونه‌های تصادفی و جامعه آماری مرجع، نظام‌مند نیستند.

برای مثال، امکان دارد میانگین پیشرفت خواندن یک نمونه تصادفی از دانش‌آموزان کلاس ششم بیشتر از میانگین پیشرفت خواندن جامعه آماری مرجع باشد، اما همان قدر نیز محتمل است که میانگین نمونه، کمتر از میانگین جامعه آماری مرجع باشد، به سخن دیگر، در یک نمونه تصادفی، به یک میزان احتمال دارد که خطای نمونه‌گیری منفی یا مثبت باشد.

بعلاوه، نظریه پردازان علم آمار، با استدلال قیاسی نشان داده‌اند که تا چه حد می‌توان انتظار داشت که مشاهدات حاصل از یک نمونه تصادفی با آنچه که در جامعه آماری مشاهده خواهد شد، تفاوت داشته باشد. تمام روشهای آماری که در این فصل توصیف شده‌اند، بر این مقصود مبتنی هستند. شما به خاطر خواهید آورد که احتمال اینکه در یک نمونه کوچک خصلتهای مشاهده شده با خصلتهای جامعه آماری تفاوت کند، بیشتر از زمانی است که یک نمونه بزرگ به کار ببریم. هنگامی که پژوهشگر از نمونه‌گیری تصادفی استفاده می‌کند، می‌تواند با روشهای آمار استنباطی، احتمال تفاوت جامعه آماری را با نمونه خود، برآورد کند. آمار استنباطی که در این فصل آمده است، تماماً مبتنی بر نمونه‌گیری تصادفی است و تنها در مواردی به کار می‌رود که در آنها انتخاب تصادفی رعایت شده باشد.

متأسفانه انجام نمونه‌گیری تصادفی ساده نیازمند تهیه فهرستی از تمام افراد موجود در یک جامعه آماری محدود است. این امر اغلب مانعی بزرگ در راه استفاده از این روش است.

### / نمونه‌گیری طبقه‌ای (نمونه‌گیری لایه‌ای)

هنگامی که جامعه آماری شامل تعدادی زیرگروه یا طبقه باشد که احتمالاً در خصلتهای مورد مطالعه با هم فرق داشته باشند، اغلب استفاده از نوعی نمونه‌گیری احتمالاتی، که نمونه‌گیری طبقه‌ای خوانده می‌شود، مطلوب است. برای مثال، اگر کسی یک نظرخواهی برای سنجش افکار در مورد یک مسئله خاص سیاسی انجام دهد، تقسیم جامعه آماری به گروههای مختلف بر مبنای سن یا شغل خردمندانه خواهد بود. دلیل این امر آن است که انتظار می‌رود در میان گروههای سنی یا شغلی، به طور نظام‌مند، نظرات متفاوتی وجود داشته باشد. در نمونه‌گیری طبقه‌ای، ابتدا طبقات مورد نظر را تعیین و سپس تعداد مشخصی از آزمودنیها را به طور تصادفی از هر طبقه انتخاب می‌کنند. امکان دارد مبنای طبقه‌بندی جغرافیایی یا خصلتهای جامعه آماری، نظیر درآمد، شغل، جنس، سن، سال تحصیلی، دانشکده یا سطح تدریس باشد.

برای مثال، در بررسی جوانان، ممکن است پژوهشگر نه تنها به پیمایش نگرشهای جوانان نسبت به پدیده خاصی علاقه‌مند باشد، بلکه بخواهد نگرش جوانانی را که در شهرهای کوچک ساکن هستند، با آنها که در شهرهای متوسط و بزرگ زندگی می‌کنند، مقایسه کند. در چنین موردی، پژوهشگر، جامعه جوانان را بر مبنای بزرگی شهری که در آن اقامت دارند، به سه گروه تقسیم می‌کند و سپس به طور تصادفی نمونه‌های مستقلی از هر طبقه (گروه) را انتخاب می‌کند.

یک مزیت نمونه‌گیری طبقه‌ای آن است که معرف بودن نمونه را افزایش می‌دهد و محقق را قادر می‌سازد که تفاوتی را که ممکن است بین زیرگروههای یک جامعه آماری وجود داشته باشند، مطالعه کند. در این نوع نمونه‌گیری محقق می‌تواند هم از هر طبقه به تعداد مساوی و هم از هر طبقه به نسبت بزرگی طبقه در جامعه آماری انتخاب کند. روش دوم، نمونه‌گیری طبقه‌ای متناسب خوانده می‌شود، که هنگامی به کار می‌رود که ویژگیهای تمام جامعه آماری مسئله اصلی بررسی باشد. هر طبقه به نسبت دقیق فراوانی آن در جامعه آماری، در نمونه نمایانده می‌شود. برای مثال، اگر ۱۰ درصد از رأی‌دهندگان دانشجو باشند، پس ۱۰ درصد از نمونه‌ای که محقق برای نظرخواهی از رأی‌دهندگان انتخاب می‌کند، از این طبقه گرفته خواهد شد.

با این همه، در برخی از مطالعات تحقیقاتی، مسئله اصلی، تفاوتی بین طبقات مختلف است. در این حالت، محقق نمونه‌ای انتخاب می‌کند که از هر طبقه به اندازه مساوی در آن باشد. برای مثال، اگر کسی در مورد تفاوت عادات مطالعه دانشجویان دوره فوق لیسانس و لیسانس بررسی می‌کند، از هر دو گروه به تعداد مساوی در نمونه خود می‌گنجاند و سپس تفاوتی موجود بین آنان را مطالعه می‌کند. شیوه مورد استفاده برحسب ماهیت سؤال تحقیق انتخاب خواهد شد. اگر تأکید بر انواع تفاوتی بین طبقات باشد، محقق از هر طبقه به تعداد مساوی انتخاب می‌کند. اگر ویژگیهای تمام جامعه آماری مسئله اصلی باشد، نمونه‌گیری متناسب مطلوبتر است.

نمونه‌گیری طبقه‌ای هر جا که کاربرد داشته باشد، احتمالاً به ما نمونه‌ای معرفتر از نمونه‌گیری تصادفی ساده ارائه می‌کند. در نمونه‌گیری تصادفی ساده، ممکن است طبقات خاصی، بر حسب اتفاق، وزن بیشتر یا کمتر در نمونه داشته باشند. برای مثال، در نمونه تصادفی ساده از دانش‌آموزان دبیرستان، به طور نظری ممکن است (اگر چه بسیار نامحتمل) که تنها دانش‌آموزان دختر انتخاب شوند. به هر حال، اگر پسران و دختران به طور جداگانه فهرست و یک نمونه

تصادفی از هر گروه انتخاب شود، چنین امری نمی‌تواند اتفاق افتد.  
مزیت بزرگ نمونه‌گیری طبقه‌ای آن است که حضور گروه‌های تعریف شده جامعه آماری را در نمونه تضمین می‌کند.

### / نمونه‌گیری خوشه‌ای

همان‌طور که قبلاً توضیح داده شد، فهرست کردن همه افراد یک جامعه آماری موردنظر و انتخاب نمونه‌ای از میان آنها، اگر ناممکن نباشد، بسیار دشوار است. جامعه آماری دانش‌آموزان مدارس آمریکا، برای مثال، به قدری بزرگ است که کسی نمی‌تواند همه افراد آن را به قصد انتخاب نمونه‌ای از میان آنها، فهرست کند. علاوه بر آن، مطالعه نمونه‌ای که افراد آن در سراسر آمریکا پراکنده‌اند، کار بسیار پرهزینه‌ای است. در چنین حالتی، مطالعه آزمودنیها در گروه‌هایی که به طور طبیعی تشکیل می‌شوند، یا خوشه‌ها، ساده‌تر خواهد بود، یعنی پژوهشگر از فهرست مدارس، تعدادی را به طور تصادفی انتخاب خواهد کرد و سپس همه دانش‌آموزان آن مدارس را در نمونه خود می‌گنجاند. این نوع نمونه‌گیری احتمالاتی، به عنوان نمونه‌گیری خوشه‌ای شناخته شده است؛ زیرا واحد انتخابی فرد نیست، بلکه گروهی از افرادند که به طور طبیعی گرد هم آمده‌اند. این افراد تا جایی که از نظر خصلت‌های مربوط به متغیر بررسی به هم شباهت دارند، یک خوشه را تشکیل می‌دهند. اکنون فرض کنیم یک نظرخواهی همگانی در آتلانتا انجام می‌شود. احتمالاً محقق به فهرست همه افراد بالغ دسترسی نخواهد داشت، بنابراین انتخاب یک نمونه تصادفی ساده برای او ناممکن خواهد بود. روش عملی‌تر، انتخاب یک نمونه تصادفی مرکب از ۵۰ منطقه از روی نقشه شهر و سپس نظرسنجی از همه افراد بالغی است که در آن مناطق زندگی می‌کنند. هر منطقه نمایانگر یک خوشه از آزمودنیها است که در خصلت‌های ویژه‌ای که با زندگی در همسایگی همبسته است، شباهت دارند.

این یک امر اساسی است که خوشه‌هایی که واقعاً در مطالعه گنجانده می‌شوند، به طور تصادفی از جامعه آماری خوشه‌ها انتخاب شوند. اگر تنها یک خوشه منفرد به کار رفته بود - برای مثال، یک مدرسه ابتدایی در یک شهر بزرگ - پژوهشگر نمی‌توانست نتایج را به جامعه آماری تعمیم دهد. یک ضرورت دیگر آن است که زمانی که یک خوشه انتخاب می‌شود، همه افراد آن خوشه باید در نمونه گنجانده شوند. خطای نمونه‌گیری در یک نمونه خوشه‌ای، بسیار بزرگتر از خطای نمونه‌گیری تصادفی واقعی است.

## / نمونه‌گیری نظام‌مند (سیستماتیک)

شکل دیگری از نمونه‌گیری احتمالاتی را نمونه‌گیری سیستماتیک می‌نامند. این روش عبارت است از گرفتن یک نمونه با انتخاب هر مورد  $K$  ام از فهرست جامعه آماری. ابتدا پژوهشگر در مورد تعداد آزمودنیهای نمونه ( $n$ ) تصمیم می‌گیرد.

از آنجا که او تعداد کل اعضای جامعه آماری ( $N$ ) را می‌داند، به آسانی  $N$  را بر  $n$  تقسیم و فاصله نمونه‌گیری  $K$  را تعیین می‌کند. اولین فرد نمونه به طور تصادفی از میان اولین  $K$  نفر فهرست انتخاب می‌شود و سپس هر فرد  $K$  ام جامعه آماری برای نمونه انتخاب می‌شود. برای مثال، فرض کنیم که یک جامعه آماری مرکب از ۵۰۰ آزمودنی داریم و می‌خواهیم نمونه‌ای با ۵۰ عضو انتخاب کنیم، بنابراین،  $K = N/n = 500/50 = 10$ .

پژوهشگر از ابتدای فهرست شروع می‌کند؛ به طوری که مورد اول از میان ده مورد اول انتخاب شود، سپس با شمارش، موردهای ۱۰ ام فهرست را در نمونه می‌گنجانند. فرض کنیم سومین نام یا شماره فهرست، اولین کسی است که انتخاب می‌شود. سپس فاصله نمونه‌گیری  $K$  یا ۱۰، به ۳ اضافه می‌شود و نفر سیزدهم در نمونه قرار می‌گیرد، همچنین نفر بیست و سوم و به همین نحو تا آخر فهرست ادامه می‌یابد.

نمونه‌گیری سیستماتیک از این نظر با نمونه‌گیری تصادفی ساده تفاوت دارد که در آن انتخابهای مختلف، مستقل نیستند، هنگامی که مورد اول انتخاب می‌شود، تمام موردهای بعدی که باید در نمونه گنجانده شوند به طور خودکار (خود بخود) تعیین می‌شوند.

اگر فهرست اصلی جامعه آماری دارای ترتیب تصادفی بود، نمونه‌گیری سیستماتیک نمونه‌ای را حاصل می‌کرد که از نظر آماری می‌توانست یک جانشین منطقی برای نمونه تصادفی به شمار آید. به هر حال، اگر فهرست مثلاً الفبایی باشد، امکان دارد هر نفر  $K$  ام جامعه آماری از خصلتهای خاصی برخوردار باشد که بر متغیر وابسته بررسی تأثیر بگذارد و بنابراین یک نمونه اریب به‌دست دهد.

نمونه‌گیری سیستماتیک از روی یک فهرست الفبایی، احتمالاً یک نمونه معرف از گروههای مختلف ملی به بار نخواهد آورد؛ زیرا برخی از گروههای ملی تمایل دارند که در زیر حروف خاصی گردآینند و فاصله‌نمونه‌گیری می‌تواند آنها را به طور کامل حذف کند یا حداقل به میزان

کافی در نمونه نگنجانند.

باید توجه داشت که روشهای مختلف نمونه‌گیری احتمالاتی که مورد بحث قرار گرفت، نا متداخل (مانعة‌الجمع) نیستند. ترکیبهای گوناگونی ممکن است مورد استفاده قرار گیرد. برای مثال، اگر پژوهشگری یک جامعه آماری بسیار بزرگ و خیلی پراکنده را مطالعه می‌کند، می‌تواند از نمونه‌گیری خوشه‌ای استفاده کند.

در همان زمان، ممکن است پژوهشگری بخواهد نمونه را طبقه‌بندی کند تا به سؤالیهای دربارهٔ طبقات مختلف پاسخ گوید. در این مورد، پژوهشگر باید جامعه آماری را طبق معیارهای معین طبقه‌بندی کند و سپس به طور تصادفی خوشه‌هایی از آزمودنیها را از میان هر طبقه انتخاب کند.

### // نمونه‌گیری غیر احتمالاتی

در بسیاری از موقعیتهای تحقیقاتی، سرشماری عناصر جامعه آماری، که یک ضرورت اساسی برای نمونه‌گیری احتمالاتی است، اگر ناممکن نباشد، دشوار است. در چنین مواردی، محقق از نمونه‌گیری غیر احتمالاتی استفاده می‌کند. گونه‌های اصلی نمونه‌گیری غیر احتمالاتی عبارتند از نمونه‌گیری اتفاقی<sup>(۱)</sup>، نمونه‌گیری هدفمند<sup>(۲)</sup> و نمونه‌گیری سهمیه‌ای<sup>(۳)</sup>.

### / نمونه‌گیری اتفاقی

نمونه‌گیری اتفاقی، که ضعیف‌ترین شیوه نمونه‌گیری تلقی می‌شود، عبارت است از استفاده از موردهای قابل دسترسی برای انجام یک مطالعه. مصاحبه با اولین فردی که در محوطه دانشگاه با او روبرو می‌شویم، استفاده از دانشجویان کلاس خود به عنوان افراد نمونه، یا گرفتن داوطلب برای انجام مصاحبه در تحقیق پیمایشی، مثالهای مختلفی از نمونه‌گیری اتفاقی هستند. هیچ راهی (بجز تکرار مطالعه با استفاده از نمونه‌گیری احتمالاتی) برای برآورد خطایی که از طریق شیوه‌های نمونه‌گیری اتفاقی در نتایج تحقیق وارد می‌شود، وجود ندارد. اگر کسی از نمونه‌گیری اتفاقی استفاده می‌کند، در تفسیر یافته‌های خود باید بی‌نهایت محتاط باشد.

1- Accidental Sampling

2- Purposive Sampling

3- Quota Sampling

## / نمونه گیری هدفمند

در نمونه گیری هدفمند، که از آن با عنوان نمونه گیری قضاوتی نیز نام برده می شود، آن دسته از عناصر جامعه آماری که حدس زده می شود معرف آن باشند، انتخاب می شوند. بنیان این نوع نمونه گیری بر این مفروضه قرار دارد که قضاوت‌های نادرست در انتخاب عناصر جامعه آماری، یکدیگر را تعدیل می کنند.

نمونه گیری هدفمند اغلب برای پیش‌بینی نتایج انتخابات سراسری به کار می رود. در هر ایالت، نواحی کوچکی که نتایج آنها در انتخابات قبلی معرف کل ایالت بوده است، انتخاب می شوند. سپس با تمام رأی دهندگان واجد شرایط در این نواحی مصاحبه می شود و نتایج این مصاحبه‌ها برای پیش‌بینی الگوی رأی‌دهی آن ایالت مورد استفاده قرار می گیرد. کارشناسان نظرسنجی با بکارگیری همین شیوه در تمام ایالتها، نتایج سراسری را پیش‌بینی می کنند.

در نمونه گیری هدفمند سؤال حیاتی این است که برای دستیابی به یک نمونه معرف، تا چه میزان می توان به قضاوتها اتکا کرد. این مفروضه که خطای قضاوتها لزوماً یکدیگر را تعدیل خواهند کرد، همیشه معتبر نیست. بعلاوه، هیچ دلیلی برای این فرض وجود ندارد که واحدهایی که معرف جامعه آماری بوده‌اند، در یک دوره زمانی همچنان معرف جامعه آماری باقی بمانند. بنابراین، استفاده از نمونه گیری هدفمند می تواند نتایج گمراه کننده‌ای به بار آورد.

نمونه گیری هدفمند، به دلیل سهولت کار و هزینه کم، در پیمایشهای نگرش‌سنجی و نظرسنجی سودمند بوده است؛ هر چند که محقق باید از محدودیت‌های این روش آگاه باشد و از آن با نهایت احتیاط استفاده کند.

## / نمونه گیری سهمیه‌ای

نمونه گیری سهمیه‌ای انتخاب موردهای معرف از طبقات مختلف یک جامعه آماری است. سهمیه‌ها بر مبنای ویژگیهای آشکار جامعه آماری، که محقق می‌خواهد نتایج را به آن تعمیم دهد، تعیین می‌شوند. اجزای نمونه به گونه‌ای انتخاب می‌شوند که نمونه به دست آمده، با توجه به ویژگیهای انتخاب یک تقریب مینیاتوری از جامعه آماری باشد. برای مثال، اگر نتایج سرشماری نشان بدهد که ۲۵ درصد جامعه آماری یک ناحیه شهری در حومه آن زندگی می‌کنند، پس ۲۵ درصد نمونه باید از حومه شهر انتخاب شوند.



برای نمونه‌گیری سهمیه‌ای باید این مراحل را طی کنید:

۱- تعدادی از متغیرها را که کاملاً با سؤال مورد بررسی وابسته‌اند، تعیین کنید تا مبنایی برای طبقه‌بندی به دست آید. برای این کار، معمولاً از متغیرهایی نظیر جنس، سن، سطح تحصیلات و طبقه اجتماعی استفاده می‌شود.

۲- با استفاده از نتایج سرشماری یا سایر داده‌های موجود، اندازه هر بخش جامعه آماری را مشخص کنید.

۳- سهمیه هر بخش جامعه آماری را متناسب با اندازه آن بخش محاسبه کنید.

۴- موردهای معرف را از هر بخش یا طبقه جامعه آماری انتخاب کنید تا سهمیه‌ها پر شوند. نقص اصلی نمونه‌گیری سهمیه‌ای، در مرحله ۴، یعنی انتخاب افراد از هر طبقه است. محقق به آسانی تشخیص نمی‌دهد که آیا افراد انتخاب شده معرف طبقه مورد نظر هستند یا خیر. انتخاب افراد احتمالاً با توجه به قابلیت دسترسی و سهولت کار انجام می‌شود. اگر محقق برای یک بررسی پیمایشی، ۲۵ درصد از خانوارهای شهرنشین را انتخاب کند، بیشتر محتمل است که به جای خانه‌های کهنه و زهموار در رفته به خانه‌های زیبا برود، به خانه‌هایی برود که بیشتر در دسترس هستند، به خانه‌هایی برود که اعضای آن در طول روز در خانه هستند و الی آخر. چنین روشهایی خود بخود اریبی نظام‌مند در نمونه ایجاد می‌کنند؛ زیرا برخی از اجزای جامعه آماری کنار گذاشته می‌شوند. بعلاوه، هیچ مبنایی برای محاسبه خطای موجود در نمونه‌گیری سهمیه‌ای در دست نیست. با وجود این کمبودها، نمونه‌گیری سهمیه‌ای در بسیاری از پروژه‌های تحقیقاتی مورد استفاده قرار گرفته است که احتمالاً بدون استفاده از این نمونه‌گیری، انجام آنها امکانپذیر نمی‌بود. بسیاری احساس می‌کنند که سرعت گردآوری داده‌ها در این نوع نمونه‌گیری، بر نواقص آن می‌چربد. بعلاوه، سالها تجربه‌اندوزی با نمونه‌های سهمیه‌ای، شناسایی برخی از خطرات آن و نحوه اجتناب از این خطرات را ممکن کرده است.<sup>(۱)</sup>

## // اندازه نمونه

یکی از اولین سؤالاتی که برای محقق مطرح می‌شود، تعداد آزمودنیهایی است که باید در

۱- رای بحث بیشتر درباره اندازه نمونه مراجعه کنید به:

W.C. Cochran (1985), Sampling Techniques (2d ed.), New York: Wiley; L.Kish (1965), Survey Sampling, New York: Wiley; and S.Sudman (1976), Applied Sampling, New York: Academic Press.

نمونه قرار بگیرند. از نظر تکنیکی، اندازه نمونه بستگی دارد به دقتی که در برآورد پارامتر جامعه آماری در یک سطح اطمینان خاص، مورد نظر محقق است. هیچ قانون یگانه‌ای وجود ندارد که بتوان با آن اندازه نمونه را تعیین کرد. اگر پژوهشگری به طور دقیق واریانس جامعه آماری، تفاوت مورد انتظار و احتمالات مطلوب خطاهای نوع اول و دوم (به صفحه ۲۴۹ مراجعه کنید) را تعریف کند، می‌تواند اندازه نمونه لازم را از نظر جبری محاسبه کند. شماری از کتب درسی آمار، این شیوه را توصیف کرده‌اند.

بهترین پاسخ به پرسش اندازه نمونه، این است که تا آنجا که امکان دارد، نمونه بزرگتری به کار ببریم. نمونه بزرگ بسیار محتمل تر است که معرف جامعه آماری باشد. علاوه بر آن، با یک نمونه بزرگ، احتمالاً داده‌ها درست‌تر و دقیق‌تر خواهند بود. به بیان دیگر، هر چه نمونه بزرگتر باشد، خطای معیار کوچکتر می‌شود. به طور کلی، خطای معیار میانگین یک نمونه با جذر  $n$  نسبت عکس دارد. بنابراین، برای دو برابر کردن دقت یک برآورد، اندازه نمونه باید چهار برابر بشود.

سوی این حکم کلی، اندازه نمونه به همگنی جامعه آماری بستگی دارد. اگر جامعه آماری مورد مطالعه همگن باشد، یک نمونه کوچک معرف آن خواهد بود، اما با افزایش تنوع جامعه آماری، یک نمونه بزرگ ضرورت خواهد داشت. برای مثال، اگر دو جامعه آماری دانش‌آموزان کلاس اول و دانش‌آموزان دوره ابتدایی را در نظر بگیریم، برای اولی یک نمونه کوچکتر و برای دومی یک نمونه بزرگتر مورد نیاز خواهد بود.

بسیاری از مؤلفان توصیه می‌کنند که حداقل ۳۰ آزمودنی در نمونه گنجانده شود؛ زیرا این تعداد، اجازه استفاده از آماره‌های نمونه بزرگ را می‌دهد. بنابراین، در تحقیق آزمایشی باید نمونه‌ای انتخاب کنیم که به ما اجازه دهد تا در هر گروه ۳۰ نفر قرار دهیم. با این همه، در تحقیق توصیفی استفاده از نمونه بزرگ مطلوب است، بویژه هنگامی که جامعه آماری مورد نظر ناهمگن باشد.<sup>(۱)</sup>

به هر حال، باید تأکید شود که اندازه نمونه به تنهایی، دقت را تضمین نمی‌کند. معرف بودن، مهمترین جنبه در انتخاب یک نمونه است. ممکن است یک نمونه بزرگ و هنوز دارای اریبی

۱ - برای بحث بیشتر درباره روشهای نمونه‌گیری مراجعه کنید به:

باشد. چنین وضعیتی در نظر سنجی سال ۱۹۳۶ نشریه Literary Digest که شکست پرزیدنت روزولت را پیش‌بینی کرد، بخوبی روشن شده است. اگر چه نمونه این نظر سنجی شامل دو و نیم میلیون پاسخگو بود، اما این نمونه معرف رای دهندگان نبود و بنابراین به نتیجه‌گیری اشتباهی منجر شد. ارببی نمونه، حاصل انتخاب پاسخگویان این نظر سنجی از دفاتر ثبت اتومبیلها، دفاتر راهنمای تلفن و فهرست آبونمان مجلات بود. این آزمودنیها به یقین معرف کل جامعه آماری رای دهندگان در سال ۱۹۳۶ نبوده‌اند. همچنین، از آنجا که نظر سنجی به وسیله پست انجام گرفت، تفاوت‌های بین کسانی که پاسخ را ارسال کردند و آنهایی که پاسخ ندادند، در نتایج ارببی ایجاد کرد. بنابراین پژوهشگر باید آگاه باشد که اندازه نمونه، ارببی حاصله از روشهای غلط نمونه‌گیری را جبران نخواهد کرد. در انتخاب نمونه، معرف بودن باید به عنوان هدف اصلی در نظر باشد.

### // مفهوم خطای نمونه‌گیری

هنگامی که از یک نمونه، دربارهٔ جامعه آماری استنباطی صورت می‌گیرد، مقدار معینی خطا در آن استنباط وجود دارد؛ زیرا می‌توان انتظار داشت که بین نمونه‌های تصادفی از یک جامعه آماری نیز تفاوت‌هایی موجود باشد. امکان دارد که میانگین‌های نمره هوش دو نمونه تصادفی دانش‌آموزان کلاس چهارم، که از یک جامعه آماری گرفته شده‌اند، با هم متفاوت باشند. چنین تفاوت‌هایی، که خطای نمونه‌گیری نامیده می‌شوند، از این واقعیت ناشی می‌شود که محقق کل جامعه آماری را مشاهده نکرده است، بلکه یک نمونه را مبنای نتیجه‌گیری خود قرار داده است. خطای نمونه‌گیری را «تفاوت بین پارامتر یک جامعه آماری و آماره یک نمونه» تعریف می‌کنند. بنابراین، اگر کسی از میانگین تمام جامعه آماری ( $\mu$ ) و نیز میانگین یک نمونه تصادفی از همان جامعه آماری ( $X$ ) آگاه باشد، تفاوت بین این دو،  $X - \mu$ ، خطای نمونه‌گیری ( $e$ ) را می‌نمایاند. بنابراین  $e = X - \mu$ . برای مثال، اگر ما بدانیم که میانگین نمره هوش یک جامعه آماری ۱۰۰۰۰ نفری از دانش‌آموزان کلاس چهارم  $\mu = 100$  است و یک نمونه تصادفی ۲۰۰ نفری، میانگین  $X = 99$  را دارد، پس خطای نمونه‌گیری  $X - \mu = 99 - 100 = -1$  خواهد بود. از آنجا که معمولاً برای برآورد پارامتر جامعه آماری به آمارهٔ نمونه تکیه می‌کنیم، این نکته که انتظار می‌رود نمونه‌ها با جامعه آماری چقدر تفاوت داشته باشند، یک رکن اساسی در آمار

استنباطی است. به هر حال، رویکرد آمار استنباطی این است که به جای تلاش در جهت تعیین تفاوت بین آماره نمونه و پارامتر جامعه آماری (که اغلب دانسته نیست) تغییرپذیری آماره‌های نمونه‌های مختلف از یک جامعه آماری را برآورد کند. از آنجا که هر آماره نمونه یک برآورد از پارامتر همان جامعه آماری محسوب می‌شود، پس هر تفاوتی بین آماره‌های نمونه‌ها را باید به خطای نمونه‌گیری نسبت داد.

### // طبیعت قانونمند خطاهای نمونه‌گیری

با توجه به اینکه نمونه‌های تصادفی مختلف که از یک جامعه آماری انتخاب شده باشند، با یکدیگر تفاوت خواهند داشت، آیا استفاده از یک نمونه برای نتیجه‌گیری در مورد جامعه آماری، واقعاً، بهتر از گمانه زنی است؟ بله، بهتر است؛ چون خطاهای نمونه‌گیری به شیوه‌ای قانونمند و قابل پیش‌بینی عمل می‌کنند. قوانین حاکم بر خطای نمونه‌گیری از طریق منطق قیاسی به دست آمده و از طریق تجربه تأیید شده‌اند.

اگر چه ما نمی‌توانیم ماهیت و مقدار خطا در یک نمونه منفرد را پیش‌بینی کنیم، اما می‌توانیم ماهیت و مقدار خطاهای نمونه‌گیری را در کل پیش‌بینی کنیم. اجازه دهید این موضوع را با مراجعه به خطاهای نمونه‌گیری مربوط به میانگین روشن کنیم.

### / خطاهای نمونه‌گیری میانگین

هنگامی که از میانگین یک نمونه  $\bar{X}$  برای برآورد میانگین یک جامعه آماری ( $\mu$ ) استفاده می‌شود، همیشه می‌توان انتظار مقداری خطای نمونه‌گیری داشت. اگر چه، چنین برآوردی در عمل مبتنی بر میانگین یک نمونه منفرد است، اما فرض کنیم که نمونه‌های تصادفی متعددی را از یک جامعه آماری انتخاب و میانگین هر یک را محاسبه کرده‌ایم. ما درخواهیم یافت که این میانگینها با یکدیگر و همچنین با میانگین جامعه آماری (اگر آن را بدانیم) تفاوت دارند. این گوناگونی میانگینها ناشی از خطای نمونه‌گیری است که در میانگین هر نمونه تصادفی به عنوان برآورد میانگین جامعه آماری وجود دارد. خطاهای نمونه‌گیری میانگین به دقت مطالعه شده‌اند و این نتیجه به دست آمده است که آنها از قوانین مشخصی پیروی می‌کنند.

امید ریاضی میانگین خطاهای نمونه‌گیری صفر است اگر تعداد نامحدودی نمونه

تصادفی داشته باشیم که از یک جامعه آماری انتخاب شده باشند، می‌توان انتظار داشت که خطاهای مثبت، به گونه‌ای خطاهای منفی را متعادل کنند که میانگین خطاهای نمونه‌گیری صفر شود. برای مثال، اگر میانگین قد دانشجویان سال اول دانشگاه ۵ فوت و ۹ اینچ باشد و نمونه‌های تصادفی متعددی از این جامعه آماری انتخاب شوند، انتظار خواهیم داشت که برخی از نمونه‌ها، میانگین قد بیشتر از ۵ فوت و ۹ اینچ و بعضی نمونه‌ها، میانگین قد کمتر از ۵ فوت و ۹ اینچ داشته باشند. به هر روی، در بلند مدت، خطاهای نمونه‌گیری مثبت و منفی یکدیگر را تعدیل خواهند کرد. اگر ما تعداد نامحدودی نمونه تصادفی با اندازه یکسان داشتیم و میانگین هر یک از این نمونه‌ها را حساب می‌کردیم، سپس میانگین همه این میانگین‌ها را محاسبه می‌کردیم، میانگین میانگین‌ها با میانگین جامعه آماری برابر می‌شد.

از آنجا که خطاهای مثبت با خطاهای منفی برابر هستند، احتمال این که میانگین یک نمونه منفرد، برآورد بزرگتر یا برآورد کوچکتری از میانگین جامعه آماری بدهد، به یک اندازه است. بنابراین، ما می‌توانیم بگوییم که میانگین یک نمونه برآورد ناریب و معقول از میانگین جامعه آماری است.

**خطای نمونه‌گیری تابع معکوس اندازه نمونه است** هر قدر اندازه نمونه افزایش یابد، مقادیر میانگین نمونه‌های متعدد، نوسان کمتری خواهند داشت. به سخن دیگر، هر قدر اندازه نمونه افزایش یابد، خطای نمونه‌گیری مورد انتظار کاهش می‌یابد. نمونه‌های کوچکتر بیشتر از نمونه‌های بزرگ مستعد خطای نمونه‌گیری هستند. محقق انتظار خواهد داشت که میانگینهای مبتنی بر نمونه‌های ده نفری نوسان بسیار بیشتری از میانگینهای مبتنی بر نمونه‌های ۱۰۰ نفری داشته باشند. در مثالی که درباره قد دانشجویان زدیم، احتمال بسیار بیشتری وجود دارد که در یک نمونه ۴ نفری، ۳ دانشجو بلندتر از متوسط قد و ۱ دانشجو کوتاه‌تر از متوسط قد باشند، تا این که در یک نمونه ۴۰ نفری، ۳۰ دانشجو بالای متوسط قد و ۱۰ دانشجو پایین متوسط قد قرار بگیرند. هر قدر اندازه نمونه افزایش یابد، احتمال نزدیک شدن میانگین نمونه به میانگین جامعه آماری نیز افزایش می‌یابد. بین اندازه نمونه و خطای نمونه‌گیری یک رابطه ریاضی وجود دارد. ما بعداً نشان خواهیم داد که این رابطه چگونه در قالب فرمولهای استنباطی جای گرفته است.

**خطای نمونه‌گیری تابع مستقیم انحراف معیار جامعه آماری است** بین اعضای یک جامعه آماری، هر قدر پراکندگی یا تنوع بیشتری داشته باشیم، در میانگینهای نمونه نیز انتظار پراکندگی زیادتری خواهیم داشت. برای مثال، میانگین وزن یک نمونه تصادفی ۲۵ نفری، که از

جامعه آماری سوارکاران حرفه‌ای انتخاب شده باشند، از میانگین وزن نمونه تصادفی ۲۵ نفری آموزگاران مدارس، خطای نمونه‌گیری نسبتاً کمتری را نشان خواهد داد. وزنه‌های سوارکاران حرفه‌ای در یک دامنه محدود قرار می‌گیرند، اما این امر در مورد وزنه‌های آموزگاران صدق نمی‌کند. بنابراین، اگر اندازه نمونه‌ها یکسان باشد، خطای نمونه‌گیری مورد انتظار برای وزن آموزگاران بیشتر از خطای نمونه‌گیری مورد انتظار برای وزن سوارکاران حرفه‌ای خواهد بود.

خطاهای نمونه‌گیری به شکل نرمال یا نزدیک به نرمال حول امید ریاضی میانگین صفر توزیع می‌شوند. میانگینهای نمونه نزدیک به میانگین جامعه آماری اغلب بیش از میانگینهای نمونه دور از میانگین جامعه آماری به دست می‌آیند. ما هر چه از میانگین جامعه آماری دورتر و دورتر می‌شویم، می‌بینیم که میانگینهای نمونه کمتر و کمتری به دست می‌آیند. هم نظریه و هم تجربه نشان داده‌اند که میانگینهای نمونه‌های تصادفی به شکل نرمال یا نزدیک به نرمال حول میانگین جامعه آماری توزیع می‌شوند.

از آنجا که در چنین حالتی خطای نمونه‌گیری، تفاوت بین میانگین نمونه و میانگین جامعه آماری است، توزیع خطاهای نمونه‌گیری نیز از نظر شکل، نرمال یا نزدیک به نرمال است. بنابه تعریف، هر دو توزیع یکسان هستند، به استثنای اینکه توزیع میانگینهای نمونه‌ها، خود میانگینی برابر با جامعه آماری دارد، در حالی که میانگین خطاهای نمونه‌گیری صفر است.

حتی اگر جامعه‌ای که نمونه‌ها از آن انتخاب شده‌اند، به شکل نرمال توزیع نشده باشد، توزیع میانگینهای نمونه‌ها به منحنی نرمال شباهت خواهد داشت. برای مثال، در یک مدرسه ابتدایی نوعی، تعداد کودکان سنین مختلف تقریباً با هم مساوی است، بنابراین نمودار چند ضلعی سن کودکان اساساً به شکل مستطیل خواهد بود. اگر از مدرسه‌ای که تعداد دانش‌آموزان ۶ تا ۱۱ ساله آن با هم مساوی هستند، نمونه‌های تصادفی ۴۰ نفری بگیریم، درخواهیم یافت که بسیاری از نمونه‌ها دارای میانگین سنی نزدیک به میانگین جامعه آماری، یعنی  $8/5$ ، هستند. نمونه‌هایی که میانگین ۸ یا ۹ دارند، کمتر خواهند بود و نمونه‌های دارای میانگین ۷ یا ۱۰ به ندرت یافت خواهند شد.

### / خطای معیار میانگین

از آنجا که گستره و توزیع خطاهای نمونه‌گیری قابل پیش‌بینی است، ما می‌توانیم برای استنباط میانگین جامعه آماری، با یک اطمینان قابل پیش‌بینی، از میانگین نمونه استفاده کنیم. به

هر روی، هنگامی که از میانگین نمونه به عنوان برآوردی از میانگین جامعه آماری استفاده می‌کنیم، به برآوردی از مقدار خطای نمونه‌گیری که در میانگین نمونه مستتر است، نیاز داریم. برای این منظور یک وسیله مهم در دست داریم و آن خطای معیار میانگین است. قبلاً گفته شد که خطای نمونه‌گیری خود را در تغییر پذیری میانگینهای نمونه‌ها نشان می‌دهد. بنابراین، اگر انحراف معیار مجموعه میانگینهای نمونه‌های تصادفی یک جامعه آماری را محاسبه کنیم، برآوردی از مقدار خطای نمونه‌گیری به دست خواهیم آورد. قبلاً دیدیم که دو چیز بر مقدار خطای نمونه‌گیری تأثیر می‌گذارند: اندازه نمونه و انحراف معیار در جامعه آماری. هنگامی که از این دو آگاه باشیم، می‌توانیم انحراف معیار خطاهای نمونه‌گیری را پیش‌بینی کنیم. این امید ریاضی انحراف معیار خطاهای نمونه‌گیری میانگین، خطای معیار میانگین نامیده می‌شود و با نماد  $\sigma_x$  نشان داده می‌شود. با منطق قیاسی نشان داده شده است که خطای معیار میانگین مساوی است با انحراف معیار جامعه آماری ( $\sigma$ ) تقسیم بر جذر تعداد موجود در هر نمونه ( $\sqrt{n}$ ). اگر به صورت فرمول بنویسیم:

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \quad (۶-۱)$$

$\sigma_x$  = خطای معیار میانگین

$\sigma$  = انحراف معیار جامعه آماری

$n$  = تعداد در هر نمونه

در فصل ۵ دیدیم که انحراف معیار ( $\sigma$ ) شاخصی است که درجه پراکندگی در بین افراد یک جامعه آماری را نشان می‌دهد. به همان شیوه، خطای معیار میانگین ( $\sigma_x$ ) نیز شاخص پراکندگی مورد انتظار در بین میانگینهای نمونه‌هایی است که از یک جامعه آماری انتخاب شده باشند. همان‌گونه که خواهیم دید، تفسیر  $\sigma$  و  $\sigma_x$  با هم کاملاً مشابه هستند.

چون میانگینهای نمونه‌های تصادفی توزیع تقریباً نرمال دارند، می‌توانیم از مدل منحنی نرمال نیز برای استنباط میانگینهای جامعه آماری استفاده کنیم. با فرض این که امید ریاضی میانگین میانگینهای نمونه با میانگین جامعه آماری مساوی است و میانگینهای نمونه‌های تصادفی به صورت نرمال توزیع شده‌اند، می‌توانیم یک نمره  $z$  برای میانگین نمونه محاسبه

کنیم و آن را به جدول منحنی نرمال ببریم تا احتمال کسب یک میانگین نمونه را به طور اتفاقی، که تا این اندازه دور یا دورتر از میانگین جامعه آماری باشد، برآورد کنیم. نمره  $Z$  با کم کردن میانگین جامعه آماری از میانگین نمونه و تقسیم حاصل آن بر خطای معیار میانگین به دست می‌آید.

$$z = \frac{\bar{X} - \mu}{\sigma_x} \quad (۶-۲)$$

برای آنکه مطلب روشن شود، یک مسئول پذیرش دانشکده را در نظر بگیرید که از خود می‌پرسد آیا جامعه آماری داوطلبان ورود به دانشکده ما در سطح متوسط قرار دارند یا زیر متوسط؟ میانگین عمومی (سراسری) آزمون ورودی دانشگاهها<sup>(۱)</sup> ۵۰۰ و انحراف معیار آن ۱۰۰ است. او یک نمونه تصادفی ۶۴ نفری از جامعه آماری خود می‌گیرد و در می‌یابد که میانگین نمونه ۴۷۰ است. سپس از خود می‌پرسد که چقدر محتمل است که یک نمونه تصادفی ۶۴ نفری با میانگین ۴۷۰ از یک جامعه آماری با میانگین ۵۰۰ انتخاب شده باشد. مسئول پذیرش، با استفاده از فرمول ۱-۶، خطای معیار میانگین را  $۱۲/۵$  محاسبه می‌کند.

$$\begin{aligned} \sigma_x &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{۱۰۰}{\sqrt{۶۴}} \\ &= ۱۲/۵ \end{aligned}$$

او از طریق فرمول ۲-۶ نمره  $Z$  را برای میانگین نمونه محاسبه می‌کند.

$$\begin{aligned} z &= \frac{\bar{X} - \mu}{\sigma_x} \\ &= \frac{۴۷۰ - ۵۰۰}{۱۲/۵} \\ &= -۲/۴ \end{aligned}$$



بنابراین، میانگین نمونه به اندازه  $2/4$  واحد خطای معیار از میانگین جامعه آماری منحرف می‌شود. احتمال داشتن یک میانگین نمونه که به این اندازه ( $\sigma_{\bar{x}} = 2/4$ ) یا بیشتر، از میانگین جامعه آماری منحرف شود، چقدر است؟ تنها لازم است به منحنی نرمال مراجعه کند تا این انحراف ( $Z$ ) را بر حسب احتمالات بیان نماید. او با مراجعه به جدول منحنی نرمال در می‌یابد که احتمال یک  $Z = -2/4$  یا کمتر،  $0/0082$  است. این به معنای آن است که یک نمره  $Z$  با این مقدار یا کمتر از این مقدار، به طور اتفاقی، تنها ۸ بار در ۱۰۰۰ بار رخ خواهد داد. از آنجا که احتمال کسب یک میانگین نمونه که این قدر از میانگین جامعه آماری دور باشد، اندک است، او نتیجه می‌گیرد که میانگین نمونه احتمالاً از جامعه آماری با میانگین ۵۰۰ نیامده است و بنابراین میانگین جامعه آماری او، یعنی داوطلبان ورود به دانشکده او، احتمالاً از ۵۰۰ کمتر است.

### /// استراتژی آمار استنباطی

آمار استنباطی، علم تصمیم‌گیری منطقی با اطلاعات محدود است. ما از آنچه که در نمونه‌ها مشاهده می‌کنیم و از آنچه که درباره خطای نمونه‌گیری می‌دانیم، استفاده می‌کنیم تا به تصمیم‌های اشتباه پذیر، اما منطقی، در مورد جامعه آماری دست یابیم. شیوه‌های آماری که قبل از دستیابی به این تصمیم‌ها به اجرا در می‌آیند، آزمونهای معناداری نامیده می‌شوند. یک ابزار اساسی این آزمونهای آماری فرضیه صفر است.

### // فرضیه صفر

فرض کنید ۱۰۰ دانش‌آموز کلاس چهارم در یک آزمایش در مورد آموزش مفاهیم ریاضی شرکت کرده‌اند. بعلاوه، فرض کنید که فرضیه تحقیق ما این است که آموزش به روش ب بیش از روش الف، تسلط بر این مفاهیم را حاصل می‌کند. ما به طور تصادفی ۵۰ دانش‌آموز را به آموزش این مفاهیم با روش الف و ۵۰ دانش‌آموز دیگر را به آموزش با روش ب تخصیص می‌دهیم. شرایط محیطی را به گونه‌ای فراهم می‌آوریم که تنها تفاوت بین دو گروه، روش آموزش آنها باشد. در پایان آزمایش، از دانش‌آموزان امتحانی می‌گیریم که تعریف مناسبی از تسلط بر مفاهیم ریاضی مورد نظر باشد. بعد از برگزاری امتحان، در می‌یابیم که میانگین دانش‌آموزانی که با روش ب آموزش دیده‌اند، بالاتر از میانگین دانش‌آموزانی است که این مفاهیم را با روش الف یاد

گرفته‌اند. این تفاوت را چگونه تفسیر می‌کنیم؟

با فرض اینکه مراقب بوده‌ایم که برای دو گروه شرایط یادگیری، به استثنای روش آموزش، مساوی باشد، می‌توانیم علت این تفاوت را ناشی از این دو عامل بدانیم: (۱) روش آموزش این تفاوت را موجب شده است، یا (۲) این تفاوت بر حسب تصادف ایجاد شده است. حتی اگر آزمودنیها به طور تصادفی به تیمارها تخصیص یافته باشند، امکان دارد که بر حسب اتفاق، گروه ب شامل دانش‌آموزانی باشد که باهوش‌تر هستند، انگیزه بیشتری برای یادگیری دارند، یا به دلایل دیگری مفاهیم ریاضی را بهتر از دانش‌آموزان گروه الف یاد گرفته‌اند، فارغ از اینکه با چه روشی آموزش دیده‌اند.

بنابراین تفاوت بین گروهها می‌توانست یا نتیجه یک رابطه بین متغیرها باشد (روش آموزش و تسلط بر مفاهیم) یا تنها نتیجه یک اتفاق باشد (خطای نمونه‌گیری). ما چگونه بدانیم که کدامیک از این دو تبیین درست است؟ در نهایت ما قادر به دانستن این امر نیستیم. بنابراین آنچه که ما انجام می‌دهیم، برآورد این امر است که چقدر احتمال دارد که تفاوت حاصله ناشی از یک تصادف باشد و تعیین می‌کنیم که در نتیجه این برآورد کدام تبیین را می‌پذیریم.

تبیین تصادفی بودن، تفاوت را به عنوان فرضیه صفر می‌شناسند. همان‌طور که از فصل ۴ به خاطر خواهید آورد، فرضیه صفر حکمی است در مورد اینکه هیچ رابطه واقعی بین متغیرها وجود ندارد و هر رابطه مشاهده شده‌ای تنها تابعی از یک تصادف است. در مثال ما، فرضیه صفر می‌گوید که هیچ رابطه‌ای بین روش آموزش و تسلط بر مفاهیم ریاضی وجود ندارد.

راه دیگر بیان فرضیه صفر در مثال ما اظهار این امر است که میانگین دانش‌آموزانی که با روش الف آموزش دیده‌اند، برابر با میانگین دانش‌آموزانی است که با روش ب یاد گرفته‌اند. اگر فرضیه صفر را به شکل فرمول بنویسیم و نماد  $\mu$  را برای میانگین جامعه آماری بکار ببریم، این حکم به صورت زیر در می‌آید.

$$H_0: \mu_A = \mu_B$$

$H_0$  = فرضیه صفر

$\mu_A$  = میانگین دانش‌آموزان کلاس چهارم که با روش الف آموزش دیده‌اند

$\mu_B$  = میانگین دانش‌آموزان کلاس چهارم که با روش ب آموزش دیده‌اند

پژوهشگران در تفسیر تفاوت مشاهده شده بین گروه‌ها، باید بین تبیین تصادفی بودن این تفاوت (فرضیه صفر) و تبیینی که می‌گوید بین متغیرها رابطه‌ای وجود دارد (فرضیه تحقیق)، یکی را انتخاب کنند. پژوهشگر باید این انتخاب را بدون آگاهی از واقعیت نهایی مربوط به جامعه‌های آماری مورد نظر به عمل آورد. پژوهشگر بر مبنای اطلاعات ناکامل تصمیم‌گیری می‌کند و بنابراین در معرض ارتکاب اشتباه احتمالی است.

### // اشتباه نوع اول و اشتباه نوع دوم

محقق می‌تواند فرضیه صفر را هم به عنوان فرضیه صفر بپذیرد و هم آن را رد کند. هر یک از این دو تصمیم ممکن است درست یا غلط باشد. اگر فرضیه صفر درست باشد، محقق در صورت پذیرفتن آن در مسیر صحیحی گام برداشته و در صورت رد کردن آن قدم در راهی اشتباه گذاشته است. رد کردن یک فرضیه صفر درست، اشتباه نوع اول نامیده می‌شود.

اگر فرضیه صفر غلط باشد، محقق در صورت پذیرفتن آن در اشتباه است و اگر آن را رد کند، کار درستی انجام داده است. پذیرفتن یک فرضیه صفر غلط، اشتباه نوع دوم نامیده می‌شود. چهار حالت ممکن در جدول ۲ - ۶ خلاصه شده است.

اکنون بیاید برخی از نتایج حاصله از دو نوع اشتباه در مثال را مورد توجه قرار دهیم.

### / اشتباه نوع اول

محقق اعلام می‌کند که بین روش آموزش و تسلط بر مفاهیم ریاضی رابطه وجود دارد. او روش ب را به عنوان روش بهتر توصیه می‌کند. مدارس کتابها و سایر موادی را که روش الف را متجلی می‌کنند، به کناری می‌نهند و موادی را که بر مبنای روش ب تهیه شده‌اند، خریداری می‌کنند. برای آشنایی معلمان با شیوه آموزش روش ب آموزشهای حین خدمت برگزار می‌شود. بعد از هزینه این همه پول و زمان، مدارس پیشرفتی در یادگیری مفاهیم ریاضی مشاهده نمی‌کنند. آزمایشهای بعدی نیز نتایجی را که در تحقیق اصلی مشاهده شده است، منعکس نمی‌کنند. اگر چه درستی و نادرستی فرضیه صفر هنوز نامشخص است، اما شواهدی که فرضیه صفر را تأیید می‌کنند، غلبه دارند. محقق اصلی بر آشفته و تحقیر می‌شود.

## جدول ۱ - ۶ ارائه شماتیک اشتباه نوع اول و اشتباه نوع دوم

		وضعیت واقعی در جامعه آماری	
		فرضیه صفر درست است	فرضیه صفر نادرست است
تصمیم‌گیری محقق پس از انجام آزمون معناداری	فرضیه صفر را رد می‌کند	اشتباه نوع اول	درست
	فرضیه صفر را می‌پذیرد	درست	اشتباه نوع دوم

## / اشتباه نوع دوم

محقق نتیجه می‌گیرد که تفاوت بین دو گروه می‌تواند تنها به طور اتفاقی پیش آمده باشد و فرضیه صفر احتمالاً درست است. او اعلام می‌کند برای نتیجه‌گیری در مورد این که یک روش بهتر از روش دیگر است، شواهد کافی در دست نیست.

محققان بعدی به این نتیجه می‌رسند که روش ب از روش الف بهتر است و مدارسی که روش آموزش خود را از الف به ب تغییر داده‌اند، گزارشهای گیرایی از دستاوردهای خود در زمینه یادگیری دانش‌آموزان ارائه می‌کنند. اگر چه واقعیت نهایی نامشخص باقی می‌ماند، اما کوهی از شواهد، فرضیه تحقیق را تأیید می‌کنند. محقق اصلی بر آشفته می‌شود (اما احتمالاً تحقیق نمی‌شود).

اشتباهات نوع اول عموماً به تغییراتی منجر می‌شوند که تضمینی برای آنها وجود ندارد. اشتباهات نوع دوم، معمولاً هنگامی که تضمینی برای تغییر وجود دارد، به حفظ وضع موجود منتهی می‌شوند. پیامدهای یک اشتباه نوع اول، به طور کلی، جدی‌تر از پیامدهای یک اشتباه نوع دوم پنداشته می‌شوند، اگرچه استثناهایی نیز بر این امر وجود دارد.

## /// سطح معنی داری

به خاطر داشته باشید که نتیجه‌گیریهای علمی بیش از آنکه احکامی از واقعیت محض باشند، حکمهایی هستند که از احتمال بالایی برای درست بودن برخوردارند. این احتمال چقدر باید بالا باشد، تا محقق بخواهد اعلام کند که بین متغیرها رابطه‌ای وجود دارد؟ به سخن دیگر، فرضیه

صفر چقدر باید نا محتمل باشد تا کسی آن را رد کند؟ پیامدهای نسبی اشتباه نوع اول و اشتباه نوع دوم را می‌سنجند و قبل از انجام آزمایشهای خود، تصمیم می‌گیرند که شواهد تا چه حد باید قوی باشند تا آنها فرضیه صفر را رد کنند. این سطح از پیش تعیین شده که در آن فرضیه صفر رد خواهد شد، سطح معنی‌داری نامیده می‌شود. احتمال یک اشتباه نوع اول مستقیماً در کنترل محقق است؛ زیرا سطح معنی‌داری را او مطابق با نوع اشتباهی که می‌خواهد خود را در مقابل آن حفاظت کند، تعیین می‌کند.

البته، می‌توان همیشه فرضیه صفر را پذیرفت و از خطای نوع اول اجتناب کرد یا همیشه فرضیه صفر را رد کرد و از خطای نوع دوم اجتناب کرد. هیچیک از این دو روش سازنده نیست. اگر پیامدهای یک اشتباه نوع اول بسیار مهم باشد، اما یک اشتباه نوع دوم پیامد کمی داشته باشد، محقق تنها ممکن است در صورتی که احتمال برآورد شده اتفاقی بودن رابطه مشاهده شده یک در هزار یا کمتر باشد، خطر یک اشتباه نوع اول را بپذیرد. این کار را آزمون فرضیه در سطح معنی‌داری  $0/001$  می‌نامند، و این سطح یک سطح کاملاً محافظه‌کارانه محسوب می‌شود. در چنین مواردی محقق بسیار مراقب است که به هنگامی که هیچ رابطه‌ای وجود ندارد، کشف رابطه‌ای را اعلام نکند. به هر حال، این تصمیم‌گیری به معنای پذیرش احتمال بالایی از یک اشتباه نوع دوم است، یعنی اعلام این که هیچ رابطه‌ای وجود ندارد، در حالی که به واقع رابطه وجود دارد.

اگر به نظر برسد که پیامدهای یک اشتباه نوع اول جدی نخواهد بود، امکان دارد محقق، در صورتی که احتمال اتفاقی بودن رابطه مشاهده شده یک در ده یا کمتر باشد، تصمیم بگیرد که وجود رابطه را اعلام کند. این را آزمون فرضیه در سطح معنی‌داری  $0/1$  می‌نامند. در اینجا محقق نوعی دوراندیشی اعتدالی در برابر اشتباه نوع اول اعمال می‌کند، در حالی که در برابر اشتباه نوع دوم خطر بزرگی متقبل نمی‌شود.

سطح معنی‌داری احتمال یک اشتباه نوع اول است که محقق مایل است با رد فرضیه صفر خطر آن را بپذیرد. اگر محققی سطح معنی‌داری را  $0/01$  قرار دهد، به این معنا است که تنها اگر احتمال برآورد شده در مورد پدیده اتفاقی بودن رابطه مشاهده شده یک درصد باشد، فرضیه صفر را رد خواهد کرد. اگر او سطح معنی‌داری را  $0/0001$  قرار دهد، فرضیه صفر را تنها در صورتی رد خواهد کرد که احتمال برآورده شده تابعی از اتفاقی بودن رابطه مشاهده شده، یک در

هزار یا کمتر باشد. دو سطح معنی داری که در حیطه تعلیم و تربیت بیشترین مورد استفاده را دارند، سطوح ۰/۰۵ و ۰/۰۱ می باشند.

محققان به طور سنتی، سطح معنی داری را بعد از سنجش اهمیت اشتباه نوع اول و اشتباه نوع دوم، اما قبل از شروع آزمایش تعیین می کنند. اگر داده هایی که بعد از تکمیل آزمایش به دست آمده اند، نشان دهند که احتمال درست بودن فرضیه صفر مساوی با یا کمتر از احتمال از پیش تعیین شده پذیرفتنی است، فرضیه صفر رد و اعلام می شود که نتایج از نظر آماری معنی دار (معتبر) هستند. اگر این احتمال بزرگتر از احتمال از پیش تعیین شده پذیرفتنی باشد، نتایج را نامعتبر توصیف می کنند، یعنی فرضیه صفر به قوت خود باقی می ماند.

منظور از واژه «معنی دار»<sup>(۱)</sup> در آمار این است: «کمتر از یک احتمال از پیش تعیین شده، محتمل است که تابعی از اتفاق باشد». نتایج پژوهشها می توانند از نظر آماری معنی دار (معتبر) باشند، بدون اینکه به طور ذاتی با معنی یا مهم باشند.

### // آزمونهای جهت دار و فاقد جهت<sup>(۲)</sup>

به طور معمول، در آزمون فرضیه صفر، جهت تفاوت برای ما اهمیت زیادی ندارد. در مقابل علاقه مندیم که دوری یا نزدیکی احتمالی آماره نمونه را از پارامتر جامعه آماری تشخیص دهیم. در مثال قبل درباره مسئول پذیرش دانشکده، دلمشغولی اصلی ما تعیین این مسئله بود که آیا تفاوت بین میانگین نمونه و میانگین جامعه آماری، از آن مقداری که می توانست بر حسب اتفاق روی دهد، افزونتر است یا خیر. آزمونی که در آنجا انجام شد، آزمون فاقد جهت نامیده می شود؛ زیرا محقق به وجود تفاوت در هر دو جهت غلاقه مند است. محقق تنها اعلام می کند که تفاوت وجود خواهد داشت.

به شکل ۱-۶ توجه کنید که ناحیه رد به طور مساوی بین دو دنباله توزیع تقسیم شده است. بنابراین، اگر یک میانگین نمونه به دست آوریم که خیلی بزرگتر یا خیلی کوچکتر از مقدار مفروض باشد، فرضیه صفر رد می شود. جهت تفاوت مهم نیست. اگر محقق بر مبنای تجربه، تحقیقات قبلی، یا نظریه، قادر به بیان جهت تفاوتهای احتمالی باشد، آنگاه آزمون جهت دار انجام خواهد داد. یک فرضیه جهت دار تعیین خواهد کرد که پارامتر بزرگتر از، یا کوچکتر از، مقدار مفروض است. بنابراین، در آزمونهای جهت دار ناحیه رد در یکی از دو دنباله توزیع قرار

1- Significant

2- Directional and Nondirectional Tests

می‌گیرد. در یک منحنی نرمال، این ناحیه نقطه‌ای مساوی با  $z = 1/645$  است. معنای این گفته آن است که ما در آزمونهای جهت‌دار، برخلاف آزمونهای فاقد جهت، ۵ درصد را بین دو طرف منحنی تقسیم نمی‌کنیم، بلکه به جای این کار، تمام ۵ درصد خطای اتفاقی را در یک طرف منحنی قرار می‌دهیم. آشکار است که آزمون جهت‌دار در مقایسه با آزمون فاقد جهت، برای رد فرضیه صفر در یک سطح مفروض به مقدار  $z$  کوچکتری نیاز دارد ( $z = 1/645$  را با  $z = 1/96$  مقایسه کنید). پس یک آزمون جهت‌دار، رد فرضیه صفر را سهل‌تر می‌سازد و بنابراین این احتمال را که فرضیه صفر رد شود، افزایش می‌دهد.

در واژگان آمار اغلب از آزمون فاقد جهت به عنوان آزمون معنی داری دو دامنه و از آزمون جهت‌دار به عنوان آزمون یک دامنه نام برده می‌شود. تصمیم برای استفاده از هر یک از آزمونهای یک دامنه یا دو دامنه باید در ابتدای مطالعه، قبل از انجام هر گونه آزمون آماری، آغاز شود. محقق منتظر نمی‌ماند تا داده‌ها بررسی بشوند و سپس آزمون یک دامنه یا دو دامنه را انتخاب کند. شیوه‌های متعددی برای آزمون فرضیه صفر وجود دارند. رایجترین این شیوه‌ها، آزمون  $t$ ، تحلیل واریانس، و آزمون  $\chi^2$  دو می‌باشند.

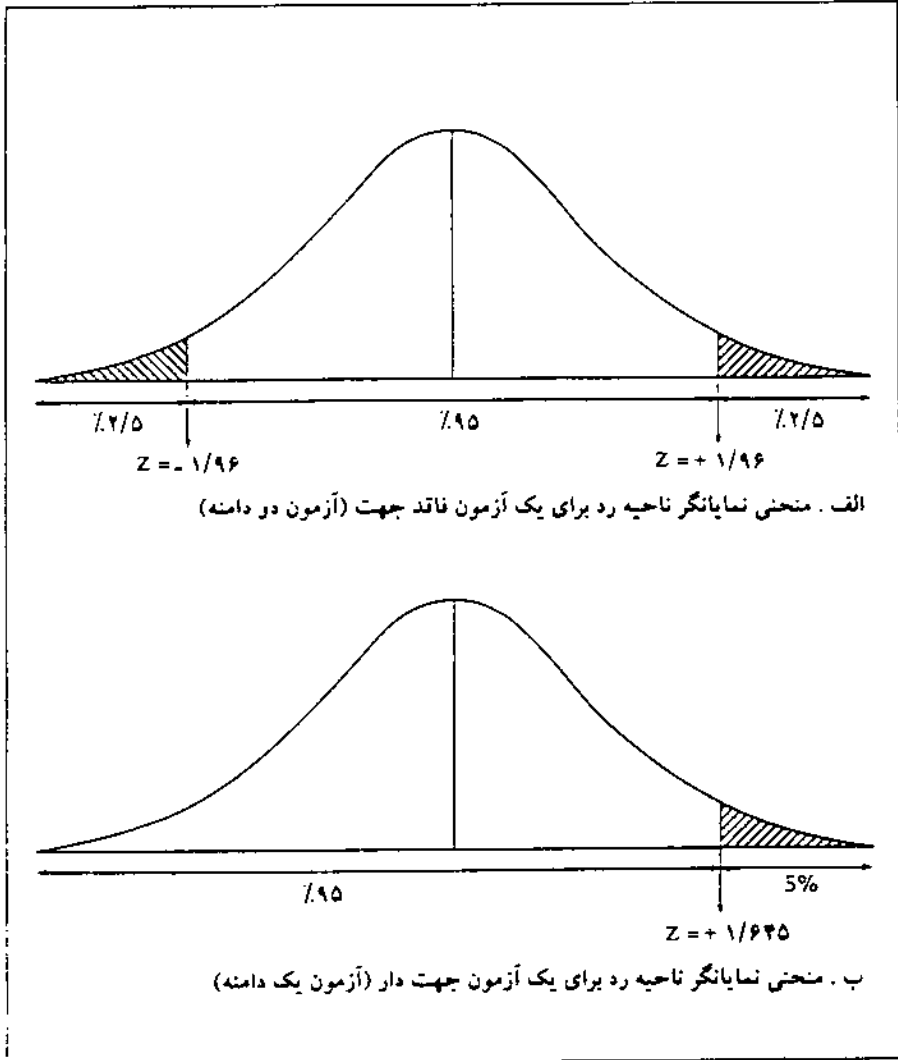
### /// آزمون $t$

ما نشان دادیم که برای مقایسه میانگین نمونه با میانگین جامعه آماری می‌توان منحنی احتمالات نرمال را به کار برد. بدین منظور و برای آنکه در یابیم آیا میانگین نمونه معرف میانگین جامعه آماری هست یا خیر، از نمره  $z$  استفاده می‌کنیم. در مثال قبل، مسئول پذیرش دانشکده خطای معیار میانگین را برای توزیع نمونه پیدا کرد، سپس فرمول  $(\bar{X} - \mu) / \sigma_x$  را به کار برد. در استفاده از این روش تناسب موجود در منحنی احتمالات نرمال نهفته است.

به هر روی، از نظر ریاضی نشان داده شده است که منحنی نرمال تنها هنگامی برای آزمون فرضیه مناسب است که انحراف معیار جامعه آماری مشخص باشد؛ همان طور که در مورد نمره‌های آزمون ورودی دانشگاهها مشخص بود. در بسیاری از موقعیتهای تحقیقاتی انحراف معیار جامعه آماری روشن نیست و باید با فرمول زیر برآورد شود.

$$S = \sqrt{\frac{\sum x^2}{n-1}} \quad (6-3)$$

شکل ۱-۶ منحنی‌های نمایانگر ناحیه‌های رد آزمونهای جهت‌دار و فاقد جهت





برآورد انحراف معیار جامعه آماری  $S$

مجموع مربع نمره‌های انحراف از میانگین یا  $\sum x^2 = \sum (X - \bar{X})^2$

تعداد نمونه  $n$

هنگامی که در محاسبه خطای معیار میانگین، این برآورد ( $S$ ) جانشین انحراف معیار جامعه آماری ( $\sigma$ ) می‌شوند، مرسوم است که فرمول ۱ - ۶ به صورت زیر بیان شود.

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \longrightarrow s_{x.} = \frac{S}{\sqrt{n}}$$

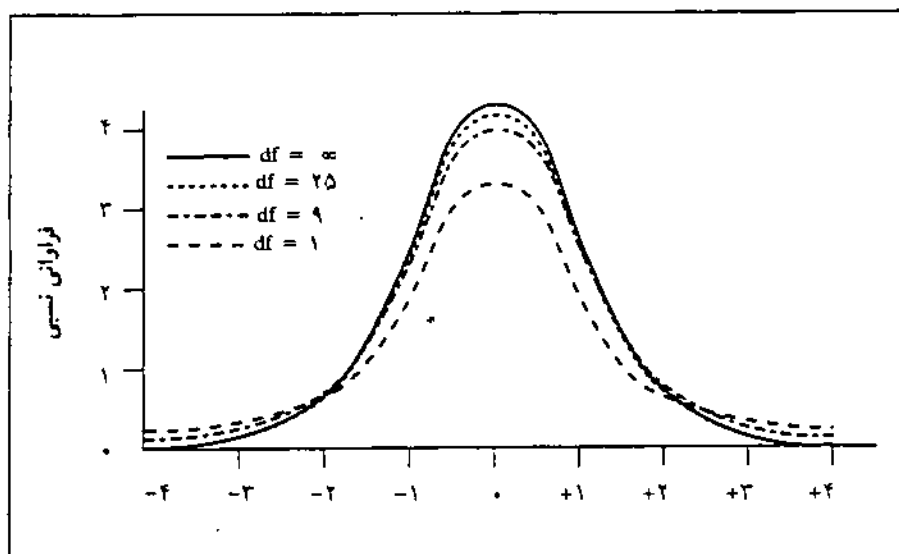
وقتی از  $s_{x.}$  به جای  $\sigma_{x.}$  استفاده شود، هر نمونه محدود، توزیع احتمالات خاص خود را خواهد داشت. این توزیعها به عنوان منحنی ۱ شناخته می‌شوند. هر قدر اندازه نمونه افزایش می‌یابد، این توزیعها بیشتر و بیشتر به منحنی نرمال شباهت پیدا می‌کنند. مجموعه‌ای از توزیعها به نام توزیعهای ۱ برای آزمون فرضیه‌های مربوط به میانگین جامعه آماری، با استفاده از نمونه‌های کوچک تهیه شده‌اند. هر قدر اندازه نمونه کوچکتر شود، توزیع ۱ نیز به طور فزاینده‌ای با توزیع ۲ متفاوت می‌شود. در تحقیقاتی که ما انجام می‌دهیم، نیازی به دانستن چگونگی محاسبه توزیع ۱ نیست؛ زیرا ضروری‌ترین نتایج این محاسبات در جدول ۲ - A در ضمیمه آمده است. منحنی ۱ کندتر از منحنی نرمال به خط پایه نزدیک می‌شود. در شکل ۲ - ۶، تعدادی از منحنی‌های ۱، به همراه منحنی نرمال - که با خط تیره و با عنوان  $df = \infty$  مشخص شده - نشان داده شده است.

منحنی‌های ۱ برحسب درجه‌های آزادی آنان، که علامت اختصاری آن  $df$  است، نامگذاری می‌شوند. پیش از بحث بیشتر درباره ویژگیهای منحنی‌های ۱، اجازه دهید توجه خود را به مفهوم درجه‌های آزادی معطوف کنیم.

### / درجه آزادی

عدد درجه آزادی به عددی اطلاق می‌شود که مستقل از هر تغییری حول یک پارامتر ثابت است. برای روشن شدن مفهوم کلی درجه آزادی، تصور کنید که آموزگاری از یک دانش‌آموز می‌خواهد که پنج عدد را که به ذهنش خطور می‌کند، بیان کند. دانش‌آموز آزاد خواهد بود که هر

شکل ۲-۶ منحنی های t برای درجه های آزادی مختلف



پنج عددی را که خود انتخاب می‌کند، نام ببرد. ما خواهیم گفت که این دانش آموز، پنج درجه آزادی دارد. اکنون فرض کنید که آموزگار به دانش آموز می‌گوید که پنج عدد را نام ببرد، اما اطمینان حاصل کند که میانگین این پنج عدد برابر ۲۰ است. اکنون این دانش آموز آزاد است که برای چهار عدد اول، هر عددی را که انتخاب می‌کند، نام ببرد، اما برای عدد آخر، باید عددی را نام ببرد که مجموع پنج عدد را به ۱۰۰ می‌رساند تا به میانگین ۲۰ دست یابد. اگر دانش آموز برای چهار عدد اول ۱۰، ۱۶، ۲۰ و ۳۵ را نام ببرد، پس عدد پنجم او باید ۱۹ باشد. این دانش آموز می‌تواند پنج عدد را نام ببرد و تنها یک محدودیت دارد. بنابراین درجه آزادی او، پنج منهای یک، یعنی چهار است ما می‌توانیم این را به صورت فرمول نشان دهیم.

$$df = n - 1$$

$$= 5 - 1$$

$$= 4$$

در نظر بگیرید که آموزگار از دانش آموز خواسته بود که از هفت عدد نام ببرد، به گونه‌ای که میانگین سه عدد اول ۱۰ و میانگین هفت عدد ۱.۲ باشد. در اینجا او هفت عدد و دو محدودیت دارد، بنابراین درجه آزادی عبارت است از:

$$\begin{aligned}df &= n - 2 \\ &= 7 - 2 \\ &= 5\end{aligned}$$

هنگامی که انحراف معیار نامعلوم جامعه آماری از انحراف معیار نمونه برآورد می‌شود،  $s_x = \frac{S}{\sqrt{n}}$ ، یک درجه آزادی از دست می‌رود؛ زیرا آماره (های) نمونه از طریق انحراف از میانگین نمونه، که مجموع آن همیشه باید صفر باشد، به دست می‌آیند. بنابراین، همه بجز یکی از انحرافهای از میانگین، آزادی تغییر دارند؛ به عبارتی  $df = N - 1$ .

### / آزمون t برای نمونه‌های مستقل

پژوهشگران اغلب دو نمونه را به طور تصادفی از یک جامعه آماری بیرون می‌کشند و تیمار آزمایشی ویژه‌ای را به هر یک تخصیص می‌دهند. این دو گروه، بعد از قرار گرفتن در معرض تیمارها، برای یافتن اثر تیمارها، با توجه به برخی از خصیصه‌ها با هم مقایسه می‌شوند. ممکن است بعد از تیمارها تفاوتی بین دو گروه مشاهده شود، اما امکان دارد این تفاوت از نظر آماری نامعتبر<sup>(۱)</sup> باشد؛ یعنی به شانس نسبت داده شود.

در این حالت، شاخصی که برای یافتن معنی‌داری تفاوت بین میانگین‌های دو نمونه مورد استفاده قرار می‌گیرد، آزمون t برای گروههای مستقل نامیده می‌شود. این نمونه‌ها را از آن رو مستقل می‌نامند که بدون جور کردن<sup>(۲)</sup> یا رابطه دیگری بین دو گروه، به طور مستقل از یک جامعه آماری انتخاب شده‌اند.

اجازه دهید از یک مثال استفاده کنیم. فرض کنید پژوهشگری می‌خواهد در یابد که آیا تنشها بر عملکرد حل مسئله تأثیر دارند یا خیر. اولین مرحله، تخصیص تصادفی دانش‌آموزان یک کلاس به دو گروه است.

از آنجا که افراد دو گروه به طور تصادفی انتخاب شده و به گروهها تخصیص یافته‌اند، میانگین عملکرد دو گروه در تکلیف حل مسئله نباید به صورت معنی‌داری قبل از ارائه تیمار فرق داشته باشد. به هر ترتیب، اگر تنش واقعاً با حل مسئله مرتبط باشد، بعد از ارائه تیمار باید میانگین عملکرد دو گروه به طور معنی‌داری متفاوت باشد. جدول ۲ - ۶، نمره‌های حل مسئله

پس از ارائه تیمار (X)، نمره‌های انحراف از میانگین (x)، و مربع نمره‌های انحراف از میانگین ( $x^2$ ) افراد دو گروه را که یکی از آنها در شرایط تنش و دیگری در شرایط آرام (بدون تنش) کار کرده‌اند، نشان می‌دهد. میانگین نمره عملکرد آزمودنی‌های گروه تنش ۱۰ و میانگین عملکرد گروه دیگر (گروه بدون تنش) ۱۴ است.

روشن است که بین این دو میانگین تفاوت وجود دارد. ضروری است تعیین کنیم که این تفاوت می‌توانست به آسانی بر حسب اتفاق بوقوع بپیوندد یا خیر. برای این کار ما باید برآورد کنیم که چه مقدار تفاوت بین گروهها تنها بر حسب اتفاق با یک فرضیه صفر درست، مورد انتظار خواهد بود. یک روش مناسب برای انجام این کار خطای معیار تفاوت بین دو میانگین ( $S_{x_1 - x_2}$ ) است. فرمول خطای معیار تفاوت بین دو میانگین در مورد نمونه‌های مستقل عبارت است از:

$$S_{x_1 - x_2} = \sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (۶-۴)$$

$S_{x_1 - x_2}$  = خطای معیار تفاوت بین دو میانگین

$n_1$  = تعداد موردها در گروه ۱

$n_2$  = تعداد موردها در گروه ۲

$\sum x_1^2$  = مجموع مربع نمره انحراف از میانگین در گروه ۱

$\sum x_2^2$  = مجموع مربع نمره انحراف از میانگین در گروه ۲

گاهی به خطای معیار تفاوت بین دو میانگین، واژه خطا برای آزمون ۱ نیز می‌گویند. در مثال

ما این مقدار را چنین محاسبه می‌کنند:

$$\begin{aligned} S_{x_1 - x_2} &= \sqrt{\frac{۱۰۶ + ۴۴}{۱۵ + ۱۵ - 2} \left[ \frac{1}{15} + \frac{1}{15} \right]} \\ &= \sqrt{\frac{۱۵۰}{28} \left[ \frac{2}{15} \right]} \\ &= \sqrt{۰/۷۱۴} \\ &= ۰/۸۴ \end{aligned}$$

جدول ۲-۶ محاسبه مقدار  $t$  برای دو میانگین نمونه

گروه ۱			گروه ۲		
گروه بدون تنش			(گروه با تنش)		
$X_1$	$x_1$	$\bar{x}_1$	$X_2$	$x_2$	$\bar{x}_2$
۱۸	+۴	۱۶	۱۳	+۳	۹
۱۷	+۳	۹	۱۲	+۲	۴
۱۶	+۲	۴	۱۲	+۲	۴
۱۶	+۲	۴	۱۱	+۱	۱
۱۶	+۲	۴	۱۱	+۱	۱
۱۵	+۱	۱	۱۱	+۱	۱
۱۵	+۱	۱	۱۰	۰	۰
۱۵	+۱	۱	۱۰	۰	۰
۱۴	۰	۰	۱۰	۰	۰
۱۴	۰	۰	۱۰	۰	۰
۱۳	-۱	۱	۹	-۱	۱
۱۲	-۲	۴	۹	-۱	۱
۱۱	-۳	۹	۸	-۲	۴
۱۰	-۴	۱۶	۷	-۳	۹
۸	-۶	۳۶	۷	-۳	۹
$\Sigma \bar{x}_1 = ۱۰۶$			$\Sigma \bar{x}_2 = ۴۴$		

این محاسبه، تفاوت بین دو میانگین را، که اگر فرضیه صفر درست باشد صرفاً به طور اتفاقی مورد انتظار خواهد بود، به ما می‌دهد. به سخن دیگر، عدد  $۰/۸۴$  تفاوتی است که ما بین میانگینهای نمره‌های عملکرد دو گروه انتظار خواهیم داشت؛ اگر آن دو گروه به طور تصادفی از یک جامعه آماری بیرون کشیده شوند و در معرض تیمارهای متفاوت قرار نگیرند. در چنین حالتی، اگر تعداد نامحدودی نمونه داشته باشیم، انتظار خواهیم داشت که در ۶۸ درصد محاسبات تفاوت بین این گروههای تصادفی تفاوت کمتر از  $۰/۸۴$  و در ۳۲ درصد دیگر آنها

تفاوت بیشتر از ۰/۸۴ را مشاهده کنیم (دلیل اینکه چرا کار برد فرمول خطای معیار تفاوت بین میانگینها تفاوت برآورد شده مناسبی را به دست می دهد که ناشی از اتفاق خواهد بود، خارج از حوصله این متن است).<sup>(۱)</sup>

در مثال ما، با یک فرضیه صفر درست، باید انتظار تفاوت ۰/۸۴ را بر حسب اتفاق داشته باشیم. دیدیم که تفاوت بین دو میانگین ۴ است. آیا تفاوت مشاهده شده به اندازه کافی بزرگتر از تفاوت مورد انتظار هست که ما را به رد فرضیه صفر قادر سازد؟ برای پاسخ به این سؤال، ابتدا نسبت دو عدد را به دست می آوریم. این نسبت را نسبت  $t$  می نامند. فرمول آن به صورت زیر است:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{x_1 - x_2}} \quad (5-6)$$

نسبت  $t = t$

تفاوت مشاهده شده بین دو میانگین  $\bar{X}_1 - \bar{X}_2$

خطای معیار تفاوت بین دو میانگین (تفاوت مورد انتظار بین دو  $S_{x_1 - x_2}$ )

میانگین هنگامی که فرضیه صفر درست است)

ما می توانیم فرمول نسبت  $t$  را با گنجاندن فرمول خطای معیار تفاوت بین دو میانگین در آن به صورت کاملتری بنویسیم:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[ \frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} \right] \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad (6-6)$$

در مثال ما مقدار نسبت  $t$  عبارت است از:

$$\frac{14 - 10}{0.84} = 4.76$$

۱ - برای اطلاعات بیشتر در مورد منطق این شیوه مراجعه کنید به:

تفاوت مشاهده شده ما  $4/76$  بار بزرگتر از تفاوت مورد انتظار با یک فرضیه صفر درست است. آیا این مقدار به اندازه کافی بزرگ هست که فرضیه صفر در سطح  $0/05$  رد شود؟ برای پاسخ به این سؤال ما تنها نیاز داریم که درجه آزادی را محاسبه و به جدول ۱ مراجعه کنیم. درجه آزادی برای آزمون ۱ مستقل عبارت است از تعداد موردها در گروه اول، بعلاوه تعداد موردها در گروه دوم، منهای ۲.

$$df = n_1 + n_2 - 2$$

در مثال ما درجه آزادی  $28 = 15 + 15 - 2$  است. اکنون می‌توانیم از جدول  $A - 2$  ضمیمه، برای تعیین معنی‌داری نتایج استفاده کنیم. در این جدول ستون اول درجه آزادی نام نهاده شده است. محقق با جایگذاری درجه آزادی که در مطالعات خود به دست آورده است، ردیف مناسب را پیدا می‌کند. در مثال ما، به ردیف مربوط به درجه آزادی ۲۸ مراجعه خواهیم کرد. ستونهای بعدی مقادیر ۱ همبسته با احتمالات معین برای آزمونهای جهت‌دار و فاقد جهت را نشان می‌دهند. از آنجا که در مثال ما، متغیر مستقل می‌تواند بر عملکرد حل مسئله، هم در جهت مثبت و هم در جهت منفی، تأثیر بگذارد، ضروری است که یک آزمون فاقد جهت انجام دهیم. در ردیف مربوط به درجه آزادی ۲۸، در ستونی که عنوان  $0/1$  برای آزمون فاقد جهت دارد، عدد  $1/701$  را پیدا می‌کنیم. این عدد به ما می‌گوید که با فرضیه صفر درست و درجه آزادی ۲۸، یک نسبت  $1/701$  یا بیشتر یا  $1/701$  - یا کمتر، در هر ده بار یک بار، برحسب اتفاق به وقوع خواهد پیوست. عدد  $2/048$  در ستونی که عنوان  $0/05$  دارد، گویای آن است که طبق فرضیه صفر درست و با درجه آزادی ۲۸، یک نسبت  $2/048 \pm$  یا بیشتر در پنج درصد مواقع بر حسب اتفاق به وقوع خواهد پیوست.

نسبت مشاهده شده ما  $4/76$  بزرگتر از  $2/048$  است. معنای این امر آن است که تفاوت بین گروههای ما بزرگتر از مقدار لازم برای رد فرضیه صفر در سطح معنی‌داری  $0/05$  است. احتمال برآورد شده درست بودن فرضیه صفر کمتر از ۵ درصد است ( $P < 0/05$ ). اگر چه ما به یقین نمی‌دانیم که متغیرها، تنش و حل مسئله، با هم رابطه دارند، اما شواهدی که طبق معیارهای از قبل مقرر شده به دست آمده‌اند، به اندازه کافی معتبر هستند که ما را قادر سازند نتیجه بگیریم که رابطه مشاهده شده تنها یک رویداد اتفاقی نیست. اگر نسبت ۱ مشاهده شده کمتر از  $2/048$

می‌بود، نتیجه می‌گرفتیم که شواهد به اندازه کافی معتبر نیستند تا ما را به اعلام وجود رابطه‌ای بین متغیرها راهبر شوند و به سخن دیگر، فرضیه صفر را می‌پذیرفتیم.

توجه داشته باشید که ارقام  $t$ ، از چپ به راست، مقادیر لازم  $t$  را برای رد فرضیه در سطوح بالاتر معنی‌داری نشان می‌دهند. برای  $df = 28$  مقدار  $df = 2/736$  یا بیشتر به رد فرضیه صفر در سطح  $0/01$  رهنمون خواهد شد. مقدار  $3/674$  یا بیشتر به رد فرضیه صفر در سطح  $0/01$  منجر خواهد شد. بنابراین مقدار  $t$  ما، یعنی  $4/76$ ، نه تنها در سطح  $0/05$  ( $P < 0/05$ ) معنی‌دار است، بلکه همچنین در سطح  $0/01$  ( $P < 0/01$ ) و در سطح  $0/001$  ( $P < 0/001$ ) نیز معنی‌دار می‌باشد.

### / منطقی آزمون $t$

اگر نسبت  $t$  را در نظر بگیریم، صورت آن تفاوت واقعی مشاهده شده بین دو گروه است. مخرج کسر  $S_x - x_p$  برآوردی است از این امر که انتظار می‌رود این دو گروه بر حسب اتفاق چقدر با هم تفاوت داشته باشند، یعنی تفاوتی را نشان می‌دهد که انتظار می‌رود بین دو گروه، که با روش تصادفی از یک جامعه آماری مشترک انتخاب شده‌اند، وجود داشته باشد. این مخرج مبتنی است بر: (۱) تعدادی که در نمونه‌ها هستند،  $n_1 + n_2$  (هر چه این تعداد بیشتر باشد تفاوت‌های تصادفی کمتری بین میانگین‌های نمونه انتظار می‌رود)، (۲) تغییر در داخل گروه‌ها،  $S_p$  و  $S_1$  (هر چه تغییر در داخل گروه‌ها بیشتر باشد، تفاوت تصادفی بیشتری در بین گروه‌ها انتظار می‌رود). از آنجا که مخرج کسر اندازه‌ای است از این که چه مقدار تفاوت آشکار بر حسب اتفاق می‌توان انتظار داشت، آن را واژه خطای آزمون  $t$  می‌نامند.

اگر نسبت تفاوت مشاهده شده (صورت کسر) بر واژه خطا (مخرج کسر) تقسیم شود و مقدار حاصله برابر مقدار معین در جدول مقادیر  $t$ ، یا بیشتر از آن باشد، می‌توان فرضیه صفر را در سطح معنی‌داری نشان داده شده رد کرد.

### / آزمون $t$ برای گروه‌های نامستقل

تاکنون بحث ما حول مقایسه میانگین‌های به دست آمده از دو گروه نمونه مستقل تمرکز یافته بود. در یک نمونه مستقل، هر نفر به طور تصادفی از جامعه آماری انتخاب می‌شود و ترکیب یک



گروه، اثری بر گروه دیگر ندارد. به هر حال، در برخی از مواقع ممکن است محقق بخواهد که در بعضی از صفات، که برای هدف تحقیق اهمیت دارند، آزمودنیهای دو گروه را با هم جور کند، یا بخواهد میانگینهایی را که از یک گروه تحت دو شرایط آزمایشی متفاوت به دست آمده است، با هم مقایسه کند. در چنین حالتی دیگر گروهها مستقل نیستند؛ زیرا ترکیب یک گروه با ترکیب گروه دیگر مرتبط است. در چنین حالتی ما انتظار خواهیم داشت نمره‌های متغیر وابسته همبستگی داشته باشند. بنابراین باید آزمون  $t$  برای میانگین‌های نامستقل یا همبسته به کار برده شود؛ این آزمون به نام آزمون  $t$  همبسته یا وابسته نیز شناخته می‌شود. اندازه‌ای که با آزمون  $t$  نامستقل تحلیل می‌شود، تفاوت بین نمره‌های جفت شده است.

بیاید مثالی را در نظر بگیریم. فرض کنید ما می‌خواهیم بدانیم که آیا گذراندن یک درس روش تحقیق، بر نگرش دانشجویان نسبت به تحقیق تأثیر می‌گذارد؟ برای بررسی این مسئله ما یک کلاس روش تحقیق را انتخاب می‌کنیم و نگرش دانشجویان را در روز اول و روز آخر کلاس نسبت به امر تحقیق می‌سنجیم. فرض کنید چنین داده‌هایی را جمع‌آوری کرده‌ایم و نتایج آنرا در جدول ۶-۳ آورده‌ایم. ستونهای ۲ و ۳ نمره‌های هر دانشجو را در اولین و دومین آزمون نشان می‌دهند. ستون ۴ تفاوت بین نمره‌های اول و دوم هر دانشجو را ارائه می‌کند. مجموع این تفاوت‌ها به  $+۳۰$  می‌رسد. میانگین این تفاوتها،  $+۲$ ، با تقسیم  $(\sum D)$  بر  $n$  تعداد مشاهده‌های جفت شده یا ۱۵، به دست می‌آید. ستون ۵ مربع تفاوتها را نشان می‌دهد.

فرمول آزمون  $t$  نامستقل عبارت است از:

$$t = \frac{\bar{D}}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N(N-1)}}} \quad (6-7)$$

$t$  = مقدار  $t$  برای میانگینهای نامستقل (همبسته)

$D$  = تفاوت بین نمره‌های جفت شده

$\bar{D}$  = میانگین تفاوتها

$\sum D^2$  = مجموع مربع تفاوتها

$N$  = تعداد زوجها

جدول ۳-۶ نمره‌های قبل - و - بعد ۱۵ دانش آموز در کلاس مقدماتی تحقیق

شماره آزمودنی	پیش آزمون	پس آزمون	D	D <sup>۲</sup>
۱	۱۰	۱۲	+۲	+۴
۲	۹	۱۳	+۴	+۱۶
۳	۸	۱۲	+۴	+۱۶
۴	۱۱	۹	-۲	+۴
۵	۱۰	۸	-۲	+۴
۶	۷	۹	+۲	+۴
۷	۱۰	۱۲	+۲	+۴
۸	۹	۱۱	+۲	+۴
۹	۸	۱۰	+۲	+۴
۱۰	۶	۱۰	+۴	+۱۶
۱۱	۱۰	۱۲	+۲	+۴
۱۲	۷	۱۳	+۶	+۳۶
۱۳	۱۰	۶	-۴	+۱۶
۱۴	۹	۱۳	+۴	+۱۶
۱۵	۱۰	۱۴	+۴	+۱۶
			$\Sigma D = +۳۰$	$\Sigma D^2 = +۱۶۴$

با جایگزینی مقادیر جدول ۳-۶ در فرمول ۷-۶ به دست می‌آوریم:

$$t = \frac{\frac{30}{15}}{\sqrt{\frac{164 - \frac{(30)^2}{15}}{15(15-1)}}} = \frac{2}{\sqrt{\frac{164 - 60}{210}}} = \frac{2}{\sqrt{\frac{104}{210}}} = \frac{2}{\sqrt{0.4952}} = \frac{2}{0.704} = 2/84$$

نسبت ۱ می‌گوید که تفاوت مشاهده شده ۲/۸۴ بار بزرگتر از تفاوتی است که با یک فرضیه صفر درست مورد انتظار خواهد بود. اکنون برای تعیین معنی‌داری آماری نسبت مشاهده شده

باید به جدول مقادیر  $t$  (جدول ۲-A ضمیمه) مراجعه کنیم.

عدد درجه آزادی برای آزمون  $t$  نامستقل برابر است با  $1 - N$  که  $N$  تعداد زوجهای مشاهده است. ما در مثال خود درجه آزادی  $14 = 15 - 1$  را داریم. در جدول مقادیر  $t$  به هنگام انجام یک آزمون فاقد جهت می‌بینیم که با درجه آزادی ۱۴، برای آن که  $t$  در سطح  $0.05$  معنی‌دار باشد، به عدد  $2/145$  و برای آن که در سطح  $0.01$  معنی‌دار باشد، به عدد  $2/977$  نیاز است. مقدار به دست آمده  $t$ ،  $2/84$ ، افزونتر از مقدار داده شده در جدول برای سطح  $0.05$  است، اما به مقدار داده شده برای سطح  $0.01$  نمی‌رسد. معنای این امر آن است که تفاوت بین دو میانگین در سطح  $0.01$  معنی‌دار نیست. اگر ما سطح معنی‌داری را  $0.05$  تعیین کرده بودیم، می‌توانستیم نتیجه بگیریم که نگرش دانشجویان در مورد تحقیق تغییر کرده است.

### ۱ / آزمون $t$ برای ضریب همبستگی $r$ پیرسون

کاربرد مهم دیگر آزمون  $t$ ، آزمون فرضیه‌های مربوط به همبستگی جامعه آماری ( $p$ ) است. قبل از آن که محقق بتواند از داده‌های نمونه، در مورد همبستگی جامعه آماری نتیجه‌گیری کند، به آزمون معنی‌داری رابطه بین متغیرها نیاز دارد. در چنین مواردی رایجترین فرضیه صفر می‌گوید که همبستگی جامعه آماری صفر و همبستگی مشاهده‌شده در نمونه ( $r$ ) تابعی از اتفاق است. برای مثال، یک محقق می‌تواند یک نمونه ۲۵ نفری از دانشجویان سال اول بگیرد، آزمونهای واژگان و موقعیت‌یابی فضایی را با آنها اجرا کند، و عدد  $0.2$  را برای  $r$  پیرسون بین آن دو آزمون به دست آورد. گام بعدی، تصمیم‌گیری در مورد این امر است که آیا این ضریب همبستگی می‌توانست یک نتیجه اتفاقی در جامعه آماری باشد که همبستگی واقعی ( $p$ ) در آن صفر است؟ توزیع  $t$  مدل مناسبی برای آزمون فرضیه صفر است که می‌گوید ضریب همبستگی جامعه آماری صفر است. آماره  $t$  را می‌توان با استفاده از فرمول زیر محاسبه کرد:

$$t = r \sqrt{\frac{N-2}{1-r^2}} \quad (6-8)$$

$t$  = مقدار  $t$

$r$  = ضریب همبستگی بدست آمده

$N$  = تعداد مشاهده‌های زوج شده

با کاربرد فرمول ۸ - ۶ در مورد مثال بالا، در می‌یابیم که مقدار  $t$ ،  $۱/۰۲$  است:

$$t = ۰/۲۰ \sqrt{\frac{۲۷ - ۲}{۱ - (۰/۲۰)^2}} = ۱/۰۲$$

مقدار  $t$  محاسبه شده با مقادیر جدول  $t$  در درجه آزادی  $N - ۲$  مقایسه می‌شوند. اگر مقدار مشاهده شده در سطح معنی‌داری مورد نظر با مقدار جدول مساوی یا از آن بیشتر باشد، یک رابطه معنی‌دار وجود دارد. جدول ۲ - A ضمیمه نشان می‌دهد که وقتی یک آزمون فاقد جهت انجام می‌دهیم، برای معنی‌داری در سطح  $۰/۰۵$  با درجه آزادی (df)  $۲۵$ ، به عدد  $t$ ،  $۲/۰۶$  نیاز داریم. چون  $t$  محاسبه شده، یعنی  $۱/۰۲$  کمتر از مقدار جدول است، محقق باید نتیجه بگیرد که برای بیان وجود رابطه بین دو متغیر شواهد کافی در دست ندارد.

جداولی تهیه شده‌اند که مقدار بحرانی لازم ضریب همبستگی  $r$  را برای رد فرضیه صفر در سطوح معنی‌داری مختلف نشان می‌دهند (به جدول ۵ - A ضمیمه مراجعه کنید). یک  $t$  معنی‌دار آن است که مقدار آن مساوی یا بزرگتر از مقدار جدول، با درجه آزادی  $N - ۲$ ، باشد ( $N$  تعداد زوجهای نمره است). جدول ۵ - A نشان می‌دهد که هنگامی که یک آزمون دو دامنه انجام می‌دهیم، برای رد فرضیه صفر، با  $df = ۲۵$ ، به  $r$  پیرسون بیشتر از  $۰/۳۸۰۹$  + یا کمتر از  $۰/۳۸۰۱$  - نیاز داریم.<sup>(۱)</sup>

بنابراین، همبستگی  $۰/۲$  به دست آمده در بررسی مربوط به دانشجویان سال اول معنی‌دار نیست. اگر نمونه نسبتاً بزرگی داشته باشیم، ممکن است یک ضریب همبستگی از نظر مقدار کم، اما از نظر آماری معنی‌دار باشد. از آنجا که مقدار همبستگی میزان رابطه بین متغیرها را نشان می‌دهد، یک همبستگی کم همیشه نمایانگر یک رابطه ضعیف است؛ حتی هنگامی که همبستگی از نظر آماری معنی‌دار باشد. برای مثال جکسون و لاهادرن با یک نمونه ۱۴۴ نفری از دختران کلاس ششم، همبستگی  $۰/۲۵$  + را بین پاسخ دانش‌آموزان به پرسشنامه رضایت از مدرسه و پیش‌بینی آموزگاران آنها در مورد نحوه پاسخگویی دانش‌آموزان به این پرسشنامه به دست آوردند.<sup>(۲)</sup> از

۱ - از یک آزمون یک دامنه با دو دامنه می‌توان استفاده کرد. اگر محقق تنها به یک ضریب مثبت یا منفی نظر داشته باشد، از آزمون یک دامنه استفاده می‌کند. به هر روی، اگر هر دو ضریب مثبت یا منفی مورد نظر باشند، آزمون دو دامنه مورد استفاده قرار می‌گیرد.

2 - Jackson, P.W., and H.M. Lahaderne (1967). Scholastic success and attitude toward school in a population of sixth graders. *Journal of Educational Psychology*, 58, 15 - 18.

آنجا که اندازه نمونه بزرگ بود، همبستگی اگر چه پایین، اما از نظر آماری در سطح ۰/۰۱ معنی دار بود. یافته‌ها نشان می‌دهند که در جامعه آماری‌ای که نمونه معرف آن است، همبستگی احتمالاً صفر نمی‌شود. هر چند، همبستگی کم خود گویای آن است که اگر چه آموزگاران می‌توانند رضایت دانش‌آموزان را در سطحی بالاتر از تصادف و اتفاق پیش‌بینی کنند، اما پیش‌بینی آموزگاران تنها رابطه ضعیفی با رضایت واقعی دانش‌آموزان دارد.

همچنین، آزمون  $t$  می‌تواند برای آزمون فرضیه‌هایی درباره همبستگیهای به غیر از صفر جامعه آماری به کار رود. همچنین، از آن می‌توان برای آزمون این فرضیه که همبستگی مشاهده شده در دو نمونه می‌توانسته است از یک جامعه آماری نشئت گرفته باشد، استفاده کرد. از آنجا که این کتاب، یک کتاب مقدماتی است، از گنجاندن چنین آزمونهایی در آن خودداری کردیم. توصیف سودمندی از این آزمونها را می‌توان در کتاب گلاس و هاپکینز<sup>(۱)</sup> و بسیاری دیگر از کتابهای آماری یافت.

### /// تحلیل واریانس

در تحلیل واریانس (ANOVA) نظیر آزمون  $t$  نسبت تفاوت مشاهده شده به واژه خطا، برای آزمون فرضیه، مورد استفاده قرار می‌گیرد. این نسبت که نسبت  $F$  نامیده می‌شود، از واریانس ( $\sigma^2$ ) میانگینهای گروهی به عنوان اندازه‌ای از تفاوت‌های مشاهده شده در بین گروهها، استفاده می‌کند. معنای این امر آن است که ANOVA روشی متنوع‌تر از آزمون  $t$  است. از آزمون  $t$  تنها برای آزمون تفاوت بین دو میانگین می‌توان استفاده کرد. برخی از آمار دانان هرگز از آزمون  $t$  استفاده نمی‌کنند؛ زیرا از ANOVA می‌توان در هر موقعیتی که آزمون  $t$  به کار می‌آید، استفاده کرد و بیش از آن می‌توان کارهای بسیاری صورت داد که از آزمون  $t$  بر نمی‌آید.

منطق اساسی ANOVA آن است که در یک آزمایش، واریانس کل همه آزمودنیها را می‌توان به دو منبع تقسیم کرد. واریانس بین گروهی و واریانس درون گروهی.

واریانس بین گروهی در نسبت  $F$  بخشی از صورت کسر است. واریانس درون گروهی قسمتی از واژه خطا یا مخرج کسر است؛ همان گونه که در آزمون  $t$  نیز هست. هر چقدر واریانس

1- Glass, Gene V., and K.D. Hopkins (1984). *Statistical Methods in Education and Psychology* (2d ed.). Englewood Cliffs, NJ: Prentice - Hall.

بین گروهی افزایش می‌یابد،  $F$  نیز زیاد می‌شود. هر قدر واریانس درون‌گروهی افزایش می‌یابد، از نسبت  $F$  کاسته می‌شود. تعداد آزمودنیها بر نسبت  $F$  تأثیر می‌گذارد. این تعداد هر چه بزرگتر باشد، نسبت  $F$  بزرگتر می‌شود. هنگامی که صورت و مخرج کسر مساوی هستند، تفاوت‌های بین میانگینهای گروهی بزرگتر از مقداری نیست که تنها بر حسب اتفاق انتظار می‌رود. اگر صورت کسر بزرگتر از مخرج آن باشد، محقق به جدول مقادیر  $F$  (در ضمیمه) مراجعه می‌کند تا تعیین کند که آیا نسبت  $F$  به اندازه کافی بزرگ هست که او را به رد فرضیه صفر در یک سطح از پیش تعیین شده، قادر سازد.

### // محاسبه نسبت $F$ (تحلیل واریانس ساده)

فرض کنید ما سه شرایط آزمایشی تنش زیاد، تنش متوسط و بدون تنش داریم و می‌خواهیم عملکرد سه گروه از افراد را که به طور تصادفی به این شرایط سه گانه تخصیص یافته‌اند، در یک عمل ساده حل مسئله مقایسه کنیم. فرض کنید داده‌هایی که در جدول ۴-۶ ارائه شده، خلاصه مشاهدات ما از عملکرد این سه گروه است و اکنون می‌خواهیم فرضیه صفر را، مبنی بر این که تفاوت معنی‌داری میان این مشاهدات نیست، در سطح ۰/۰۱، آزمون کنیم.

می‌توان دید که میانگینها با یکدیگر و با میانگین کل ۳۰ آزمودنی ( $X$ ، میانگین کل) تفاوت دارند. آیا تفاوت‌های بین میانگینها به اندازه کافی بزرگ هستند که از نظر آماری معنی‌دار باشند یا احتمال دارد که این تفاوتها به طور اتفاقی به وقوع پیوسته باشند؟ برای پاسخ به این سؤال نسبت  $F$  را محاسبه می‌کنیم.

اولین گام یافتن مجموع مربع انحراف نمره‌های فردی از میانگین کل است. این شاخص مجموع مربعهای کل نامیده می‌شود و تمام اثرات تیمار و خطای نمونه‌گیری را منعکس می‌کند.

$$SS_T = \sum x_i^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{N} \quad (۶-۹)$$

در مثال ما این مقدار عبارت است از:

$$SS_T = ۶۸۳۰ - \frac{(۴۴۴)^2}{۳۰} = ۲۵۸/۸$$

جدول ۴-۶ اندازه‌های حاصله از سه نمونه تصادفی مبتنی بر عملکرد آنها در شرایط تنش زیاد، تنش متوسط و بدون تنش

گروه ۱ تنش زیاد		گروه ۲ تنش متوسط		گروه ۳ بدون تنش	
$X_1$	$X_1^2$	$X_2$	$X_2^2$	$X_3$	$X_3^2$
۱۹	۳۶۱	۲۲	۴۸۴	۱۵	۲۲۵
۱۸	۳۲۴	۲۰	۴۰۰	۱۴	۱۹۶
۱۷	۲۸۹	۱۹	۳۶۱	۱۴	۱۹۶
۱۶	۲۵۶	۱۸	۳۲۴	۱۳	۱۶۹
۱۵	۲۲۵	۱۷	۲۸۹	۱۳	۱۶۹
۱۵	۲۲۵	۱۶	۲۵۶	۱۲	۱۴۴
۱۴	۱۹۶	۱۶	۲۵۶	۱۲	۱۴۴
۱۳	۱۶۹	۱۵	۲۲۵	۱۱	۱۲۱
۱۲	۱۴۴	۱۴	۱۹۶	۱۱	۱۲۱
۱۱	۱۲۱	۱۲	۱۴۴	۱۰	۱۰۰
$\Sigma X_1 = 150$	$\Sigma X_1^2 = 2310$	$\Sigma X_2 = 169$	$\Sigma X_2^2 = 2920$	$\Sigma X_3 = 125$	$\Sigma X_3^2 = 1585$
$\bar{X}_1 = 15/0$		$\bar{X}_2 = 16/9$		$\bar{X}_3 = 12/5$	$\bar{X} = 14/8$
$\Sigma X_1 = 444$		$\Sigma X_2 = 6830$			

سپس مجموع کل مربعاتی انحراف میانگینهای گروهی از میانگین کل را پیدا می‌کنیم. این شاخص، مجموع مربعاتی بین گروهی نامیده می‌شود (اگر بیش از دو گروه داشته باشیم، از نظر دستوری درست‌تر است بگوییم مجموع مربعاتی بین گروهها. به هر حال، استفاده از اصطلاح مجموع مربعاتی بین گروهی یک سنت دیر پا است، و برای این که با سایر کتابهای درسی آمار در تضاد نباشیم، این اصطلاح را بکار می‌بریم). این شاخص با استفاده از فرمول زیر به دست می‌آید.

$$SS_b = \Sigma x_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \dots - \frac{(\Sigma X)^2}{N} \quad (6-10)$$

در مسئله ما این مقدار عبارت است از:

$$SS_b = \sum x_b^2 = \frac{(150)^2}{10} + \frac{(169)^2}{10} + \frac{(125)^2}{10} - \frac{(444)^2}{30} = 97/4$$

سپس مجموع کل مربعات انحراف نمره‌های فردی از میانگین گروه خود را پیدا می‌کنیم. این شاخص، مجموع مربعات درون گروهی نامیده می‌شود و با کاربرد فرمول نمره‌های خام برای مجموع مربع انحراف از میانگین هر گروه و بعد جمع همه گروه‌ها به دست می‌آید.

$$(SS_w = SS_1 + SS_2 + SS_3)$$

$$SS_w = \sum x_w^2 = \sum X_1 - \frac{(\sum X_1)^2}{n_1} + \sum X_2 - \frac{(\sum X_2)^2}{n_2} + \dots \quad (6-11)$$

در مسئله ما این مقدار عبارت است از:

$$SS_w = \sum x_w^2 = 2310 - \frac{(150)^2}{10} + 2935 - \frac{(169)^2}{10} + 1585 - \frac{(125)^2}{10} = 161/4$$

همچنین می‌توانستیم مجموع مربعات درون گروهی را با منهای کردن مجموع مربعات بین گروهی از مجموع مربعات کل به دست آوریم، یعنی:

$$SS_w = SS_t - SS_b \quad \text{یا} \quad \sum x_w^2 = \sum x_t^2 - \sum x_b^2 \quad (6-12)$$

در مورد مثال ما

$$SS_w = 258/8 - 97/4 = 161/4$$

### F / آزمون معنی‌داری

جدول ۵-۶ نتایج محاسباتی را که تاکنون انجام داده‌ایم، همراه با نتایج محاسبات بعدی، در خود جای داده است. ستون اول این جدول سه منبع واریانس را فهرست کرده است: واریانس بین گروهی، واریانس درون گروهی و واریانس کل. ستون دوم مجموع مربع‌ها را شامل است، که ما قبلاً آنها را محاسبه کردیم. ستون سوم این جدول عدد درجه آزادی همبسته با هر منبع واریانس را فهرست کرده است. عدد درجه آزادی برای واریانس بین گروهی برابر است با



(۱ - G) که G تعداد گروهها است. در مثال ما این مقدار عبارت است از  $3-1=2$ . درجه‌های آزادی برای واریانس درون گروهی برابر است با  $\dots + n_2 - 2 + n_1 - 1$  (یا  $(n - G)$ )، تعداد کل نمره‌ها  $[n]$  منهای تعداد گروهها  $[G]$ . در مثال ما این مقدار عبارت است از  $27 = 30 - 3 = 27$  یا  $27 = 10 - 1 + 10 - 1 + 10 - 1$ . عدد درجه آزادی برای واریانس کل برابر است با  $N - 1$ ، در مثال ما،  $29 = 30 - 1$ . این عدد آخر را می‌توانستیم با جمع کردن درجه‌های آزادی بین‌گروهی و درون‌گروهی به دست آوریم.

سپس، گام بعدی محاسبه میانگین مربع بین‌گروهی و درون‌گروهی است. این مقادیر با تقسیم مجموع مربعهای بین‌گروهی و درون‌گروهی به درجه‌های آزادی آنها به دست می‌آید. مقادیر حاصله میانگین مربع‌ها هستند. در مثال ما میانگین مربع بین‌گروهی عبارت است از  $\frac{97}{4} = 48/7$  و میانگین مربع درون‌گروهی عبارت است از  $\frac{161}{4} = 5/98$ . میانگین مربع درون‌گروهی واژه خطا برای نسبت F است. سرانجام، با کاربرد فرمول زیر روش محاسبه تحلیل واریانس، نسبت F، به پایان می‌رسد:

جدول ۵-۶ خلاصه تحلیل واریانس سه گروه

(۱)	(۲)	(۳)	(۴)	(۵)	(۶)
منبع	SS	df	MS	F	سطح
واریانس					معنی‌داری
بین‌گروهی	۹۷/۴	۲	۴۸/۷	۸/۱۴	۰/۰۱
درون‌گروهی	۱۶۱/۴	۲۷	۵/۹۸		
کل	۲۵۸/۸	۲۹			

$$F = \frac{MS_b}{MS_w} = \frac{SS_b / df_b}{SS_w / df_w} \quad (۶-۱۳)$$

در مثال ما این مقدار عبارت است از:

$$F = \frac{48/70}{5/98} = 8/14$$

اکنون برای تعیین معنی داری نسبت  $F$  حاصله از نظر آماری، به جدول ۳-۳ ضمیمه مراجعه می‌کنیم. ستونی را که در راس آن درجه آزادی بین‌گروهی (صورت کسر) آزمایش ما قرار دارد، پیدا می‌کنیم و در این ستون رو به پایین می‌رویم تا به ردیفی که مدخل آن منطبق با عدد درجه آزادی درون‌گروهی (مخرج کسر است) است، برسیم. در اینجا دو مقدار پیدا خواهیم کرد: یک عدد با حروف ریز و یک عدد با حروف درشت.

اگر نسبت  $F$  ما برابر یا بزرگتر از مقداری باشد که با حروف ریز نوشته شده، در سطح  $0/05$  معنی دار است. اگر نسبت  $F$  به دست آمده برابر یا بزرگتر از مقداری باشد که با حروف درشت نوشته شده، در سطح  $0/01$  نیز معنی دار است. در مثال ما، با درجه‌های آزادی ۲ و ۲۷، برای رد فرضیه صفر در سطح  $0/05$  به یک نسبت  $F$   $3/35$  و در سطح  $0/01$  به یک نسبت  $F$   $5/49$  نیاز داریم. از آنجا که نسبت  $F$  به دست آمده بزرگتر از این دو مقدار است، در سطح  $0/01$  معنی دار است و فرضیه صفر در این سطح رد می‌شود.

روش تحلیل واریانس بر این مفروضه مبتنی است که اگر گروههایی که قرار است با هم مقایسه شوند، از یک جامعه آماری انتخاب شده‌اند، میانگین مربعهای بین‌گروهی نباید بیش از مقداری که به طور اتفاقی انتظار خواهیم داشت، با میانگین مربعهای درون‌گروهی تفاوت داشته باشند.

بنابراین، ما با یک فرضیه صفر درست، انتظار خواهیم داشت که نسبت  $F$  تقریباً مساوی ۱ باشد. از سوی دیگر، اگر فرضیه صفر نادرست باشد، تفاوت بین میانگین گروهها باید بزرگتر از آن چیزی باشد که برحسب اتفاق انتظار داریم، به گونه‌ای که میانگین مربع بین‌گروهی افزونتر از میانگین مربع درون‌گروهی باشد. در چنین مواردی، نسبت  $F$ ، یعنی میانگین مربع بین‌گروهی تقسیم بر میانگین مربع درون‌گروهی، بزرگتر از ۱ خواهد بود. سپس به جدول مقادیر  $F$  مراجعه می‌کنیم (جدول ۳-۳) تا ببینیم این نسبت برای داده‌های ما به اندازه کافی بزرگتر از ۱ هست تا بتوانیم فرضیه صفر را در سطح مورد نظر رد کنیم.

همچنانکه تفاوت بین این میانگین مربعها افزایش می‌یابد، نسبت  $F$  نیز افزایش می‌یابد و احتمال درستی فرضیه صفر کاهش می‌یابد.

هنگامی که فرضیه صفر به عنوان نتیجه این روش تحلیل واریانس رد می‌شود، نمی‌توانیم چیزی بیش از این بگوییم که اندازه‌هایی که از این گروهها به دست آورده‌ایم، با هم تفاوت دارند و

این تفاوتها بزرگتر از مقداری است که تنها به طور اتفاقی انتظار خواهیم داشت. یک نسبت  $F$  معنی‌دار لزوماً به این مفهوم نیست که تمام گروهها به طور معنی‌داری با تمام گروههای دیگر تفاوت دارند.  $F$  معنی‌دار ممکن است نتیجه تفاوت موجود بین یک گروه با بقیه گروهها باشد. برای نمونه امکان دارد در مسئله ما گروه ۳ به طور معنی‌داری با گروه ۱ و گروه ۲ تفاوت داشته باشد، اما گروه ۱ و ۲ با هم تفاوت معنی‌داری نداشته باشند. آزمونهای آماری بسیاری وجود دارند که می‌توان آنها را برای یافتن مکان معنی‌داری تفاوتها به کار برد؛ آزمونهایی که به وسیله تاکی<sup>(۱)</sup> و شفه<sup>(۲)</sup> تهیه شده‌اند، در این زمینه خیلی سودمند هستند.<sup>(۳)</sup>

در مثال ما، سه گروه را به طور تصادفی از یک جامعه آماری برگزیدیم و بنابراین می‌توانیم فرض کنیم که قبل از آن که تیمارهای آزمایشی را در مورد آنها اعمال کنیم، با هم تفاوتی فراتر از آنچه که معلول تصادف است، ندارند. معنی‌داری نسبت  $F$  نشان می‌دهد که تفاوت یافت شده بین این گروهها، بعد از تیمار، فراتر از آن است که حاصل اتفاق باشد. ما این تفاوت را به تیمار آزمایشی نسبت می‌دهیم و نتیجه می‌گیریم که سطح تنش بر عملکرد افراد در تکالیف ساده حل مسئله تأثیر می‌گذارد.

این تمام آن چیزی است که ما در تفسیر این نسبت  $F$  می‌توانیم بگوییم. اگر به تحلیل آماری بیشتری نیاز داشته باشیم، می‌توانیم از آزمون‌های تاکی، شفه، یا سایر آزمونها برای تعیین معنی‌داری بین گروههای مشخص یا ترکیبهای گروهها استفاده کنیم. این روشها (تکنیکها) به ما می‌گویند که چگونه شرایط تنش خاصی بر عملکرد تأثیر می‌گذارند و می‌توانند به سوالاتی نظیر «آیا در شرایط تنش متوسط و زیاد، تنش متوسط و بدون تنش، تنش زیاد و بدون تنش، در نمره‌های عملکرد تفاوتی وجود دارد؟»، پاسخ گویند.

## // تحلیل واریانس چند عاملی

امکان دارد بخواهیم اثر ترکیبی سطح تنش و نیاز پیشرفت در عملکرد یک تکلیف مسئله گشایی را بررسی کنیم. برای بررسی این مسئله هم سطح تنش و هم نیاز پیشرفت را تغییر خواهیم داد. طرح یک آزمایش که اثرهای ترکیبی دو یا چند متغیر مستقل را بررسی می‌کند، یک طرح عاملی خوانده می‌شود و نتایج آن به وسیله تحلیل واریانس چند عاملی بررسی می‌شود.

اجازه دهید فرض کنیم که این آزمایش را با استفاده از پنج آزمودنی در هر گروه انجام داده‌ایم و داده‌های جدول ۶-۶ خلاصه‌ای از مشاهدات ما را در مورد عملکرد آزمودنیها منعکس می‌کند. کاربرد تحلیل واریانس چند عاملی ما را قادر می‌سازد که این موارد را پیدا کنیم: (۱) آیا تفاوت معنی‌داری بین عملکرد آزمودنیهایی که در شرایط تنش زیاد قرار دارند و آنها که در شرایط تنش کم هستند، وجود دارد، (۲) آیا تفاوت معنی‌داری بین عملکرد آزمودنی‌ها با نیاز پیشرفت زیاد و آزمودنیها با نیاز پیشرفت کم، وجود دارد و (۳) آیا دو متغیر، تنش و نیاز پیشرفت، بر عملکرد آزمودنیها اثر ترکیبی دارند؟ اثرهایی که با تحلیل‌های اول و دوم بررسی شدند، اثرهای اصلی نامیده می‌شوند، در حالی که سومی اثر تعاملی خوانده می‌شود. حاصل نهایی این تحلیل‌ها سه نسبت F خواهد بود، دو تای آنها معنی‌داری دو اثر اصلی و سومی معنی‌داری اثر تعاملی را نشان می‌دهد.

جدول ۶-۶ اندازه‌های تکلیف مسئله‌گشایی آزمودنیهایی با نیاز پیشرفت کم و زیاد تحت شرایط تنش زیاد و کم

		تنش			
		زیاد	کم		
زیاد		۲۰	۲۳	گروه ۱ $\bar{X} = 19$	$\sum X_{r1} = 200$ $\bar{X}_{r1} = 20$
		۲۰	۲۲		
		۱۹	۲۱		
		۱۹	۲۰		
		۱۷	۱۹		
	$\sum X_{105}$	$\sum X_{105}$			
نیاز پیشرفت کم		۲۲	۱۸	گروه ۲ $\bar{X} = 20$	$\sum X_{r2} = 175$ $\bar{X}_{r2} = 17/5$
		۲۱	۱۶		
		۲۰	۱۵		
		۱۹	۱۴		
		۱۸	۱۲		
	$\sum X_{100}$	$\sum X_{105}$			
$\sum X_{c1} = 195$		$\sum X_{c2} = 180$		$\sum X = 275$ کل	
$\bar{X}_{c1} = 19/5$		$\bar{X}_{c2} = 18$		$\bar{X}_{(میانگین کل)} = 18/75$	

محاسبه این نسبت‌های F شامل گامهای زیر است:

- ۱- مجموع مربعات کل، مجموع مربعات بین گروهی و مجموع مربعات درون گروهی را با استفاده از همان روشها و فرمولهایی که در تحلیل واریانس ساده به کار رفت، پیدا کنید. این مقادیر، که از داده‌های جدول ۶-۶ استخراج شده‌اند، عبارتند از:

$$SS_t = 7181 - \frac{(375)^2}{20} = 149/75$$

$$SS_b = \frac{(95)^2}{5} + \frac{(105)^2}{5} + \frac{(100)^2}{5} + \frac{(75)^2}{5} - \frac{(375)^2}{20} = 103/75$$

$$SS_w = 149/75 - 103/75 = 46$$

- ۲- مجموع مربعات بین گروهی را به سه مجموع مربعات جداگانه بشکنید (تبدیل کنید): (الف) مجموع مربعات بین ستونها، (ب) مجموع مربعات بین ردیفها و (ج) مجموع مربعات برای تعامل بین ستونها و ردیفها. الف - مجموع مربعات بین ستونها معرف مجموع توان دوم انحرافهای ناشی از تفاوت بین میانگینهای ستونی و میانگین کل است. این مجموع را با استفاده از فرمول زیر پیدا می‌کنند:

$$SS_{bc} = \frac{(\sum X_{c1})^2}{n_{c1}} + \frac{(\sum X_{c2})^2}{n_{c2}} + \dots - \frac{(\sum X)^2}{N} \quad (6-14)$$

- با استفاده از این فرمول، مجموع مربعات بین ستونها برای داده‌های جدول ۶-۶ عبارت است از:

$$SS_{bc} = \frac{(195)^2}{10} + \frac{(180)^2}{10} - \frac{(375)^2}{20} = 11/25$$

- ب - مجموع مربعات بین ردیفها عبارت است از مجموع توان دوم انحرافهای ناشی از تفاوت بین میانگینهای ردیفی و میانگین کل. این مجموع را با کاربرد فرمول زیر به دست می‌آورند:

$$SS_{br} = \frac{(\sum X_{r1})^2}{n_{r1}} + \frac{(\sum X_{r2})^2}{n_{r2}} + \dots - \frac{(\sum X)^2}{N} \quad (6-15)$$

برای داده‌های جدول ۶-۶، این مقدار عبارت است از:

$$SS_{br} = \frac{(200)^2}{10} + \frac{(175)^2}{10} - \frac{(375)^2}{20} = 31/25$$

ج - تعامل مجموع مربعات بخشی از انحراف بین میانگینهای گروهی و میانگینهای کلی است که نه از تفاوت‌های ردیفی و نه از تفاوت‌های ستونی ناشی می‌شود.

به سخن دیگر، این مجموع، تفاوت بین کل مجموع مربعات بین گروهها و مجموع مربع‌های بین ردیفها است، یعنی:

$$SS_{int} = SS_b - (SS_{bc} + SS_{br}) \quad (6-16)$$

اگر به زبان کلمات بیان کنیم، مجموع مربعات تعاملی برابر است با مجموع مربع‌های بین گروهی منهای مجموع مربع‌های بین ستونی و مجموع مربعات بین ردیفی برای داده‌های جدول ۶-۶، مقدار این تعامل عبارت است از:

$$SS_{int} = 103/75 - (11/25 + 31/25) = 61/25$$

۳- عدد درجه آزادی همبسته با هر منبع تغییر را تعیین کنید. درجه‌های آزادی را به صورت زیر پیدا می‌کنیم:

df برای مجموع مربعات بین ستونها = ۱ - C

df برای مجموع مربعات بین ردیفها = ۱ - R

df برای تعامل = (R - ۱) (C - ۱)

df برای مجموع مربعات بین گروهها = ۱ - G

df برای مجموع مربعات درون گروهها = N - G

df برای مجموع مربعات کل = ۱ - N

C = تعداد ستونها

R = تعداد ردیفها

G = تعداد گروها

N = تعداد آزمودنی‌ها در همه گروها

۴- مقادیر میانگین مربعها را با تقسیم هر مجموع مربع بر عدد درجه آزادی همبسته آن پیدا کنید.

۵- نسبتهای F را برای اثرهای اصلی و تعاملی با تقسیم میانگین مربعهای بین‌گروهی بر میانگین مربع درون‌گروهی برای هر یک از سه جزء محاسبه کنید.

۶- نتایج محاسباتی که بر مبنای داده‌های جدول ۶-۶ انجام گرفته، به طور خلاصه در جدول

۶-۷ آمده است. سه نسبت F در این جدول فهرست شده است. برای یافتن معنی‌داری هر یک از

این مقادیر، مانند قبل، به جدول مقادیر F مراجعه می‌کنیم. در مراجعه به این جدول از عدد

درجه آزادی همبسته با هر نسبت F (df برای صورت کسر) و عدد درجه آزادی همبسته با

میانگین مربع درون‌گروهی (df برای مخرج کسر) استفاده می‌کنیم. برای مثال، نسبت F بین

ستونهای ما ۳/۹۱۳ است. با مراجعه به جدول F، در می‌یابیم که با درجه‌های آزادی ۱ و ۱۶،

یک نسبت F ۴/۴۹ یا بیشتر برای معنی‌داری در سطح ۰/۰۵، مورد نیاز است. چون نسبت F ما

کمتر از مقداری است که در جدول نشان داده شده است، پس معنی‌دار نیست.

جدول ۶-۷ خلاصه یک تحلیل واریانس چند عاملی ۲ × ۲

منبع واریانس	SS	df	MS	F	سطح معنی‌داری
بین ستونها (تشریح)	۱۱/۲۵	۱	۱۱/۲۵	۳/۹۱۳	-
بین ردیفها (نیاز پیشرفت)	۳۱/۲۵	۱	۳۱/۲۵	۱۰/۸۶۹	۰/۰۱
ستونها در ردیفها (تعامل)	۶۱/۲۵	۱	۶۱/۲۵	۲۱/۳۰۴	۰/۰۱
بین گروها	۱۰۳/۷۵	۳	۳۴/۵۸۳		
درون گروها	۴۶	۱۶	۲/۸۷۵		
کل	۱۴۹/۷۵	۱۹			

نسبت  $F$  برای بین ردیفها، با درجه‌های آزادی ۱ و ۱۶، برای معنی‌دار بودن باید به  $۴/۴۹$  (در سطح  $۰/۰۵$ ) یا  $۸/۵۳$  (در سطح  $۰/۰۱$ ) برسد. چون مقدار  $F$  به دست آمده  $۱۰/۸۶۹$ ، از هر دوی این مقادیر بیشتر است، در سطح  $۰/۰۱$  معنی‌دار است. برای تعامل بین ستونها و ردیفها، با درجه‌های آزادی ۱ و ۱۶ یک نسبت  $F$   $۴/۴۹$  (در سطح  $۰/۰۵$ ) یا  $۸/۵۳$  (در سطح  $۰/۰۱$ ) مورد نیاز است. مقدار  $F$  به دست آمده  $۲۱/۳۰۴$ ، از هر دوی این مقادیر بیشتر است، بنابراین در سطح  $۰/۰۱$  معنی‌دار است.

### تفسیر نسبتهای $F$

اولین نسبت  $F$  در جدول ۷ - ۶ (بین ستونها) معنی‌دار نیست و نشان می‌دهد که شرایط تنش‌زا در تأثیرشان بر عملکرد آزمودنیها در آزمایش، به طور معنی‌دار با هم تفاوت ندارند. این تجزیه و تحلیل مقایسه‌ عملکرد ترکیبی گروههای ۱ و ۲ با عملکرد ترکیبی گروههای ۳ و ۴ است. ما می‌توانستیم با استفاده از روش آزمون ۱ به همین نتیجه برسیم.

نسبت  $F$  دوم (بین ردیفها)، که در سطح  $۰/۰۱$  معنی‌دار است، بر مبنای مقایسه عملکرد آزمودنیهای گروههای ۱ و ۳ با آزمودنیهای گروههای ۲ و ۴ بنا نهاده شده است. ما از معنی‌داری این نسبت  $F$  می‌توانیم استنباط کنیم که تفاوت بین عملکرد آزمودنیهای با نیاز پیشرفت زیاد و آزمودنیهای با نیاز پیشرفت کم فوآتر از یک تفاوت اتفاقی است. با بررسی داده‌های جدول ۷ - ۶ در می‌یابیم که گروههای با نیاز پیشرفت زیاد میانگین ترکیبی ۲۰ را در مقایسه با میانگین  $۱۷/۵$  گروههای با نیاز پیشرفت کم کسب کرده‌اند. از آنجا که برای این تفاوت یک نسبت  $F$  معنادار داریم، نتیجه می‌گیریم که در شرایطی مشابه شرایط آزمایش ما، از آزمودنیهایی که نیاز پیشرفت زیاد داشته باشند، سطح بالاتری از عملکرد را می‌توان انتظار داشت.

نسبت  $F$  سوم اثر تعاملی بین دو متغیر، سطح تنش و نیاز پیشرفت را نشان می‌دهد. در این مورد معنی‌داری نسبت  $F$ ، یعنی اثر سطح تنش بر عملکرد یک کار مسئله‌گشایی به درجه نیاز پیشرفت بستگی دارد.

مقایسه نتایج مشاهده شده با نتایج مورد انتظار در صورت وجود تعامل بین دو متغیر مستقل، این پدیده را با وضوح بیشتری نشان می‌دهد.



اکنون اجازه دهید میانگین چهار گروه را در صورتی که هیچ تعاملی وجود نمی داشت، محاسبه کنیم. میانگین تمام آزمودنیها  $۱۸/۷۵$  است. میانگین برای ده آزمودنی که در شرایط تنش زیاد قرار دارند،  $۱۹/۵$  است که  $۰/۷۵$  بیشتر از میانگین کل است، در حالی که میانگین ده آزمودنی که در شرایط تنش کم قرار دارند،  $۰/۷۵$  کمتر است. میانگین ده آزمودنی با نیاز پیشرفت زیاد،  $۲۰$  است که  $۱/۲۵$  بیشتر از میانگین تمام آزمودنیها است، در حالی که میانگین ده آزمودنی که نیاز پیشرفت کم دارند،  $۱/۲۵$  کمتر است.

ما می توانیم برای هر گروه میانگینی را محاسبه کنیم که اگر هیچ تعاملی وجود نمی داشت، انتظار می رفت آن گروه چنین میانگینی را دارا باشد. ما این کار را با افزودن تفاوت مربوط به ستونی که گروه در آن قرار دارد و تفاوت مربوط به ردیفی که گروه در آن است، به میانگین کل انجام می دهیم.

اگر تعاملی وجود نمی داشت، انتظار می داشتیم که میانگین گروه ۱ چقدر باشد؟ با میانگین کل،  $۱۸/۷۵$ ، شروع می کنیم،  $۰/۷۵$  را به آن اضافه می کنیم؛ زیرا آزمودنیها در شرایط تنش زیاد قرار داشتند و سپس  $۱/۲۵$  را به آن می افزاییم، زیرا آنها نیاز پیشرفت زیاد داشتند. بعد از این افزایشها یک میانگین  $۲۰/۷۵$  را به دست می آوریم.

با پیروی از همین روش برای هر چهار گروه، مقادیر مورد انتظار زیر را به دست خواهیم آورد:

مقدار مورد انتظار	=	تفاوت مربوط به نیاز پیشرفت	+	تفاوت مربوط به تنش	+	میانگین رویهم
$۲۰/۷۵$		$+ ۱/۲۵$		$+ ۰/۷۵$		گروه ۱ $۱۸/۷۵$
$۱۸/۲۵$		$- ۱/۲۵$		$+ ۰/۷۵$		گروه ۲ $۱۸/۷۵$
$۱۹/۲۵$		$+ ۱/۲۵$		$- ۰/۷۵$		گروه ۳ $۱۸/۷۵$
$۱۶/۷۵$		$- ۱/۲۵$		$- ۰/۷۵$		گروه ۴ $۱۸/۷۵$

اکنون میانگینهای گروهی واقعی را با این میانگینهای گروهی مورد انتظار مقایسه می کنیم:

		واقعی				مورد انتظار			
		زیاد	کم			زیاد	کم		
زیاد	گروه ۱	$X=19$	$X=21$	$X=20$	آسان	گروه ۱	$X=20/75$	$X=19/25$	$X=20$
	گروه ۲	$X=20$	$X=15$			گروه ۲	$X=18/25$	$X=16/75$	
کم	گروه ۳	$X=20$	$X=15$	$X=17/5$	دشوار	گروه ۳	$X=18/25$	$X=16/75$	$X=17/5$
	گروه ۴	$X=20$	$X=15$			گروه ۴	$X=18/25$	$X=16/75$	
		$\bar{X}=19/5$	$\bar{X}=18$	$\bar{X}=18/75$		$\bar{X}=19/5$	$\bar{X}=18$	$\bar{X}=18/75$	

(توجه کنید که ما می‌توانستیم از تفاوت‌های بین مقادیر مورد انتظار و مقادیر مشاهده‌شده برای محاسبه مجموع مربعات برای تعامل، مستقیماً استفاده کنیم. میانگین واقعی هر گروه با میانگین مورد انتظار آن  $1/75$  تفاوت دارد. این مقدار را به دو می‌رسانیم و در تعداد موردها ضرب می‌کنیم تا  $20 \times 1/75 = 61/25$  را به دست آوریم.)

می‌بینیم که گروه ۱ بدتر از آن که ما انتظار داشتیم، عمل کردند و می‌دانیم که آنها در شرایط تنش زیاد هستند و نیاز پیشرفت زیاد دارند. گروه ۲، که نیاز پیشرفت کم داشتند و در شرایط تنش زیاد بودند، بهتر از انتظار ما بودند. اگر گروه‌های تحت تنش کم را در نظر بگیریم، در می‌یابیم که گروه ۳ با نیاز پیشرفت زیاد بهتر از انتظار عمل کرده، در حالی که گروه ۴ با نیاز پیشرفت کم، کمتر از انتظار عمل کرده است.

از آنجا که آزمون F نشان داد که تعامل معنی‌دار است، نتیجه می‌گیریم که تنش زیاد هنگامی که با نیاز پیشرفت کم ترکیب شود، نمره‌های بهتری ایجاد می‌کند تا اینکه با نیاز پیشرفت زیاد ترکیب شود، در حالی که تنش کم زمانی که با نیاز پیشرفت زیاد همراه می‌شود، نسبت به ترکیب آن با نیاز پیشرفت کم نمره‌های بهتری را موجب می‌شود.

در تحقیقات آموزشی استفاده از تحلیل چند عاملی از ارزش زیادی برخوردار است؛ زیرا بسیاری از سؤالاتی که دست اندرکاران آموزش و پرورش، تحقیق در مورد آنها را ضروری می‌بینند، به طور ذاتی دارای طبیعت پیچیده‌ای هستند. این روشها ما را قادر می‌سازد که اثرات ترکیبی دو یا چند متغیر مستقل را در رابطه با یک متغیر وابسته تحلیل کنیم. برای مثال، یک مقایسه ساده میانگینهای متغیر وابسته دو گروه از دانش‌آموزان که با روشهای متفاوتی آموزش دیده‌اند، ممکن است نتایج نامعتبر حاصل کند. اما اگر هوش را به عنوان یک متغیر مستقل

اندازه گیری شده در آزمایش دخالت دهیم، ممکن است دریابیم که یک روش با دانش آموزان کم هوش تر، نتیجه بهتری می دهد، در حالی که روش دیگر برای دانش آموزان باهوش تر مؤثرتر است. تحلیل واریانس چند عاملی، مانند مثال ما، به دو متغیر مستقل محدود نیست. امکان داخل کردن هر تعداد از متغیرهای مستقل در این روش وجود دارد. کتابهای آمار متعددی، از جمله کتاب ادواردز، محاسبه و تفسیر این روشها را توصیف کرده اند.<sup>(۱)</sup>

### /// آزمون معنی داری خی دو<sup>(۲)</sup>

گاهی از اوقات نیازمند آن هستیم که معنی داری تفاوت میان نسبتهای آزمودنیها، اشیاء، رویدادها و نظایر آنها را که در طبقه بندیهای مختلفی جای می گیرند، پیدا کنیم. آزمون آماری که در چنین مواردی به کار گرفته می شود، آزمون خی دو ( $\chi^2$ ) نامیده می شود.

در آزمون خی دو دو مجموعه از فراوانیها با هم مقایسه می شوند: فراوانی مشاهده شده و فراوانی مورد انتظار. فراوانی مشاهده شده، همان گونه که از نام آن آشکار است، فراوانی واقعی است که در نتیجه مشاهده به دست آمده است. فراوانی مورد انتظار یک فراوانی نظری است که برای مقایسه به کار می رود.

این فرضیه را در نظر بگیرید که نسبت دانش آموزان پسر به دختر در درس آمار، با نسبت کل دانش آموزان پسر و دختر در مدرسه متفاوت است. اگر بدانیم که ۴۰ درصد از کل نام نویسی در مدرسه، مربوط به پسران است و ۳۰۰ دانش آموز در درس آمار نام نوشته اند، فراوانیهای مورد انتظار عبارت خواهند بود از:

$$\left. \begin{array}{l} ۱۲۰ \text{ دانش آموزان پسر} \\ ۱۸۰ \text{ دانش آموزان دختر} \end{array} \right\} ۳۰۰$$

اکنون فرض کنید که فراوانیهای مشاهده شده ما چنین باشد.

$$\left. \begin{array}{l} ۱۴۰ \text{ دانش آموزان پسر} \\ ۱۶۰ \text{ دانش آموزان دختر} \end{array} \right\} ۳۰۰$$

1- Edwards, A. L (1984). *Experimental Designs in Psychological Research* (5 th ed.; ch. 11 and 12). New York: Harper & Row.

2- Chi Square

می‌خواهیم تعیین کنیم آیا تفاوت بین فراوانیهای مورد انتظار و مشاهده شده، از نظر آماری معنی دار است؟ برای تعیین این مسئله فرمول خبی دو را به کار می‌بریم، که به صورت زیر است:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \quad (۶-۱۷)$$

$\chi^2$  = مقدار خبی دو

$f_o$  = فراوانی مشاهده شده

$f_e$  = فراوانی مورد انتظار

اگر داده‌ها را در این فرمول بگذاریم، خواهیم داشت:

$$\chi^2 = \frac{(۱۴۰ - ۱۲۰)^2}{۱۲۰} + \frac{(۱۶۰ - ۱۸۰)^2}{۱۸۰} = ۵/۵۵$$

برای تعیین اینکه این مقدار خبی دو معنی دار است، به جدول مقادیر  $\chi^2$  در ضمیمه مراجعه می‌کنیم (جدول ۴ - A). اولین ستون این جدول عدد درجه آزادی مربوط به مسئله خبی دو را نشان می‌دهد. ستونهای دیگر، مقادیر مورد نیاز برای سطوح مختلف معنی داری را ارائه می‌کنند. عدد درجه آزادی، همان گونه که قبلاً بحث کردیم، مبتنی است بر تعداد مشاهده‌های مستقل از تغییر، به هنگامی که محدودیتهای خاصی در مورد داده‌ها اعمال شده است. هنگامی که تعداد ثابتی مشاهده داشته باشیم که تنها به دو طبقه تقسیم شده‌اند، به محض اینکه تعداد موجود در یک طبقه مشخص شد، دیگری نیز ثابت می‌شود. بنابراین، زمانی که دانش‌آموزان پسر را ۱۴۰ نفر تعیین کردیم، تعداد دانش‌آموزان دختر در کلی ۳۰۰ نفر، باید ۱۶۰ نفر باشد. در این مثال تنها یک درجه آزادی وجود دارد. در مسائلی نظیر این، عدد درجه آزادی برابر است با  $K-۱$  وقتی  $K$ ، تعداد طبقه‌های مورد استفاده برای طبقه‌بندی باشد. با مراجعه به جدول  $\chi^2$  در می‌یابیم که مقدار مشاهده شده ۵/۵۵، در سطح ۰/۰۵ (و ۰/۰۲) از نظر آماری معنی دار است.

با تفسیر این نتیجه اکنون می‌توانیم بگوییم که در مدرسه ما، با سطح اطمینان ۰/۰۵، نسبت پسرنانی که در درس آمار شرکت می‌کنند، از دختران بیشتر است. سطح معنی داری ۰/۰۵ به این معنا است که در هر ۱۰۰ بار مشاهده، چنین تفاوتی بین نسبت‌های دانش‌آموزان پسر و دختر، کمتر از ۵ بار آن می‌تواند نتیجه اتفاق باشد. بنابراین داده‌ها، فرضیه تحقیق ما را مبنی بر اینکه

نسبت دانش‌آموزان پسری که در درس آمار نام‌نویسی می‌کنند، بیشتر از دانش‌آموزان دختر است، تأیید می‌کنند.

استفاده از آزمون  $\chi^2$  دو تنها به موقعیتهایی که دارای دو طبقه‌بندی باشند، محدود نمی‌شود، بلکه همچنین می‌توان آن را در آزمون فرضیه صفر، مبنی بر عدم تفاوت معنی‌دار بین آزمودنی‌هایی که در طبقه‌بندی‌های مختلفی قرار دارند، به کار برد. برای مثال، فرض کنید از یک نمونه ۱۲۰ نفری دانشجویان دوره لیسانس سؤال کرده‌ایم آیا ترجیح می‌دهند در خوابگاه زندگی کنند، یا در خانه شخصی، و یا برای آنها فرقی نمی‌کند. نتایج این پرسش در جدول ۸-۶ آمده است. اگر هیچ تفاوتی بین سه طبقه از پاسخها وجود نداشته باشد، در هر طبقه ۴۰ پاسخ خواهیم داشت. این مقادیر، فراوانیهای مورد انتظار ما خواهند بود، همان‌گونه که در جدول ۹-۶ نشان داده شده است.

مقایسه‌ای بین دو مجموعه فراوانی ارائه شده در جداول ۸-۶ و ۹-۶ نشان می‌دهد که بین داده‌های مورد انتظار و مشاهده شده تفاوتی وجود دارد. برای تعیین اینکه آیا این تفاوتها معنی‌دار هستند، آزمون  $\chi^2$  دو را به کار می‌بریم. با استفاده از فرمول ۱۷-۶ مقدار  $\chi^2$  برای این داده‌ها عبارت خواهد بود از:

$$\chi^2 = \frac{(40 - 40)^2}{40} + \frac{(50 - 40)^2}{40} + \frac{(30 - 40)^2}{40} = 5$$

درجه آزادی برابر است با تعداد طبقه‌ها منهای یک ( $K - 1$ ) یا در این مورد،  $2 = 3 - 1$ . با مراجعه به جدول  $\chi^2$  می‌بینیم که با درجه آزادی ۲، برای معنی‌داری در سطح ۰/۰۵ به  $\chi^2$  دو ۵/۹۹۱ یا بیشتر نیاز است. به هر ترتیب مقدار  $\chi^2$  کمتر از آن مقدار است و بنابراین از نظر آماری معنی‌دار نیست. معنای این امر آن است که تفاوت مشاهده شده بین طبقه‌های ما به آسانی می‌توانست در اثر اتفاق به وقوع بپیوندد. در نتیجه فرضیه صفر مبنی بر عدم تفاوت معنی‌دار بین فراوانیهای سه طبقه‌بندی ما نمی‌تواند رد شود. به سخن دیگر، اگر نسبتهای رجحان برای این سه طبقه‌بندی در کل جامعه آماری دانشجویان دوره لیسانس مساوی باشند، انتظار خواهیم داشت که تفاوت بین نمونه‌ها، به همان بزرگی که در مورد نمونه ما وجود داشت، بیش از ۵ بار در هر ۱۰۰ بار نمونه‌گیری، در نتیجه اتفاق روی بدهد.

جدول ۸ - ۶ فراوانیهای مشاهده شده در پاسخهای ۱۲۰ دانشجوی دربارهٔ رجحان آنها در مورد مکان زندگی

کل	فرقی نمی‌کند	خانه شخصی	خوابگاه	آزمودنی
۱۲۰	۳۰	۵۰	۴۰	دانشجویان دوره لیسانس

جدول ۹-۶ فراوانیهای موردانتظار پاسخهای ۱۲۰ دانشجوی دربارهٔ رجحان آنها در مورد مکان زندگی

کل	فرقی نمی‌کند	خانه شخصی	خوابگاه	آزمودنی
۱۲۰	۴۰	۴۰	۴۰	دانشجویان دوره لیسانس

### // آزمون ناوابستگی خی دو

تا اینجا، تنها مثالهایی را در نظر گرفتیم که در آنها مشاهددها در جهت یک بعد منفرد طبقه‌بندی شده‌اند. به هر حال، گاهی می‌خواهیم از بیش از یک بعد برای طبقه‌بندی استفاده کنیم. برای مثال، فرض کنید بعد دیگری به مسئله قبلی اضافه می‌کنیم و هم از دانشجویان دوره لیسانس و هم از دانشجویان دوره فوق لیسانس می‌خواهیم که رجحانهای خود را در مورد مکان زندگی بیان کنند. فرض کنید فراوانیهای جدول ۱۰ - ۶ نتیجه این پرسش است.

در این مورد فرضیه صفر ما این است که رجحان برای مکان زندگی در مورد دانشجویان فوق لیسانس همان است که برای دانشجویان لیسانس بود، یعنی متغیرهای پایگاه اجتماعی دانشجویان با رجحان در مورد مکان زندگی نامرتبط یا مستقل هستند. در این آزمون ناوابستگی خی دو، فرضیه صفر همیشه می‌گوید که متغیرها در جامعه آماری مستقل یا ناوابسته هستند.

مشاهدات ما نشان می‌دهد که ۳۰ درصد از کل دانشجویان، خوابگاه و ۴۵ درصد خانه شخصی را ترجیح می‌دهند و ۲۵ درصد نیز بین این دو فرقی نمی‌گذارند. اگر فرضیه صفر درست باشد، ما انتظار خواهیم داشت که بین دانشجویان لیسانس و فوق لیسانس نسبتهای مساوی پیدا

کنیم، همان گونه که در جدول ۱۱ - ۶ نشان داده شده است. ما می‌توانیم فراوانی طبقات مورد انتظار را با ضرب کردن فراوانی ردیف همبسته یا طبقه در فراوانی ستون همبسته با همان طبقه و سپس تقسیم حاصلضرب بر کل ( $E = f_r \cdot f_c / N$ ) محاسبه کنیم.

برای مثال، فراوانی مورد انتظار پاسخهای دانشجویان دوره لیسانس که می‌خواهند در خوابگاه زندگی کنند، عبارت است از  $36 = 120 \times 60 / 200$ ، برای آنها که ترجیح می‌دهند در خانه شخصی زندگی کنند، عبارت است از  $54 = 120 \times 90 / 200$  و برای دانشجویان دوره فوق لیسانس که می‌خواهند در خوابگاه زندگی کنند، عبارت است از  $24 = 80 \times 60 / 200$ . با استفاده از این روش، فراوانی مورد انتظار برای هر طبقه را پیدا می‌کنیم.

جدول ۱۰ - ۶ فراوانیهای مشاهده شده در پاسخهای ۲۰۰ دانشجوی دوره‌های لیسانس و فوق لیسانس دربارهٔ رجحان آنها در مورد مکان زندگی

کل	فرقی نمی‌کند	خانه شخصی	خوابگاه	آزمودنی
۱۲۰	۳۰	۵۰	۴۰	دانشجویان دوره لیسانس
۸۰	۲۰	۴۰	۲۰	دانشجویان دوره فوق لیسانس
۲۰۰	۵۰	۹۰	۶۰	کل

جدول ۱۱ - ۶ فراوانیهای مورد انتظار پاسخهای ۲۰۰ دانشجوی دوره‌های لیسانس و فوق لیسانس دربارهٔ رجحان آنها در مورد مکان زندگی

کل	فرقی نمی‌کند	خانه شخصی	خوابگاه	آزمودنی
۱۲۰	۳۰	۵۴	۳۶	دانشجویان دوره لیسانس
۸۰	۲۰	۳۶	۲۴	دانشجویان دوره فوق لیسانس
۲۰۰	۵۰	۹۰	۶۰	کل

توجه کنید که در جدول ۶-۱۱، اعداد کل ردیف و ستون، دقیقاً همان اعدادی هستند که در جدول ۶-۱۰ نیز آمده‌اند. اکنون ما در پی یافتن این هستیم که آیا فراوانیهای مشاهده شده به حد کفایت، با فراوانیهای مورد انتظار تفاوت دارند که ما را به رد احتمال بروز اتفاقی چنین تفاوتی قادر سازند. با کاربرد فرمولی که به دست آورده‌ایم، این امر را آزمون می‌کنیم.

$$\chi^2 = \frac{(۴۰ - ۳۶)^2}{۳۶} + \frac{(۵۰ - ۵۴)^2}{۵۴} + \frac{(۳۰ - ۳۰)^2}{۳۰} + \frac{(۲۰ - ۲۴)^2}{۲۴} + \frac{(۴۰ - ۳۶)^2}{۳۶} + \frac{(۲۰ - ۲۰)^2}{۲۰}$$

$$\chi^2 = ۱/۸۵۱۸$$

برای یافتن عدد درجه آزادی در مورد یک جدول دو طرفه از فرمول زیر استفاده می‌شود:

$$df = (C - ۱)(R - ۱) \quad (۶ - ۱۸)$$

عدد درجه آزادی = df

C = تعداد ستونها

R = تعداد ردیف‌ها

با کاربرد این فرمول عدد درجه آزادی برای مسئله مورد نظر ما، عبارت خواهد بود از:

$$df = (۳ - ۱)(۲ - ۱) = ۲$$

با مراجعه به جدول ۴ - A می‌بینیم که مقدار  $\chi^2$  لازم برای معنی‌داری در سطح ۰/۰۵ با دو درجه آزادی ۵/۹۹۱ است، اما مقدار  $\chi^2$  به دست آمده ما، یعنی ۱/۸۵۸۱، کمتر از مقدار جدول است و بنابراین معنی‌دار نیست. این امر به معنای آن است که تفاوت بین فراوانیهای مورد انتظار و مشاهده شده فراتر از آن مقداری نیست که به طور اتفاقی انتظار می‌رود. به سخن دیگر، ما در جامعه آماری که از آن نمونه‌گیری کردیم، شواهد معتبری در مورد رابطه بین متغیر «پایگاه اجتماعی دانشجوی» و متغیر «رجحان مکان زندگی» در دست نداریم. فرضیه صفر را نمی‌توان رد کرد.

/ مفروضه‌های  $\chi^2$  دو

استفاده از  $\chi^2$  دو آن قدر آسان است که شاید فراموش کنیم برای تفسیر معتبر آن، باید



مفروضه‌هایی را مدنظر قرار دهیم:

۱- مشاهده‌ها باید مستقل باشند، یعنی در هر نمونه، آزمودنیها باید به طور تصادفی و مستقلاً انتخاب شوند.

۲- طبقه‌بندیها باید نامتداخل باشند، هر مشاهده‌ای می‌تواند در یکی و تنها در یکی از طبقه‌بندیهای جدول ما قرار بگیرد.

۳- مشاهده‌ها برحسب فراوانی اندازه‌گیری می‌شوند.

### /// خلاصه

پژوهشگران آرزو دارند که با مطالعه گروههایی از افراد، که از یک جامعه آماری انتخاب شده‌اند، به تعمیم‌هایی در مورد آن جامعه آماری دست پیدا کنند. این تعمیم‌ها تنها هنگامی بی‌نقص خواهند بود که گروههای انتخابی مورد استفاده در این مطالعات (نمونه‌ها) معرف گروههای بزرگتری که این گروههای انتخابی از میان آنها برگزیده شده‌اند (جامعه آماری)، باشند. مابین دو نوع اصلی نمونه‌گیری تمایز قایل شده‌اند: نمونه‌گیری احتمالاتی و نمونه‌گیری غیر احتمالاتی. مشخصه نمونه‌گیری احتمالاتی انتخاب تصادفی اعضای جامعه آماری است. در نمونه‌گیری غیراحتمالاتی، قضاوت محقق جای انتخاب تصادفی را می‌گیرد. نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه‌ای، نمونه‌گیری خوشه‌ای و نمونه‌گیری سیستماتیک (نظام‌مند) اشکال نمونه‌گیری احتمالاتی هستند. در نمونه‌گیری تصادفی ساده، همه اعضای جامعه آماری برای قرار گرفتن در نمونه، بخت مساوی دارند. در نمونه‌گیری طبقه‌ای، نمونه‌های مستقل، از زیرگروهها یا طبقات جامعه آماری انتخاب می‌شوند. در نمونه‌گیری خوشه‌ای، گروههایی که به طور طبیعی در جامعه آماری تشکیل شده‌اند، انتخاب می‌شوند. سپس تمام افرادی که در خوشه‌های انتخابی قرار دارند، به عنوان نمونه مورد استفاده قرار می‌گیرند. سرانجام، در نمونه‌گیری سیستماتیک، موردهای Kام از فهرست جامعه آماری، به عنوان اعضای نمونه انتخاب می‌شوند. اشکال نمونه‌گیری غیر احتمالاتی عبارتند از نمونه‌گیری اتفاقی، نمونه‌گیری هدفمند و نمونه‌گیری سهمیه‌ای. در نمونه‌گیری اتفاقی، موردهای قابل دسترسی به عنوان نمونه مورد استفاده قرار می‌گیرند. در نمونه‌گیری هدفمند، موردهای نوعی جامعه آماری مورد نظر، بنا بر قضاوت محقق، نمونه را تشکیل می‌دهند. سرانجام، در نمونه‌گیری سهمیه‌ای، به طبقات

مختلف جامعه آماری سهمیه تخصیص می‌یابد، سپس در درون هر طبقه موردهایی که معرف آن طبقه تشخیص داده شوند، به عنوان نمونه انتخاب می‌شوند.

آمار استنباطی ابزاری به دست پژوهشگران می‌دهد تا به وسیله آن بتوانند میزان اطمینان را در پاسخگویی به این پرسش که «اگر پدیده‌ای در نمونه مشاهده شد، آیا در جامعه آماری مربوط به آن نمونه نیز مشاهده خواهد شد»، برآورد کنند. به سخن دیگر، آمار استنباطی ما را قادر می‌سازد تا مقدار اعتبار مشاهدات خود را برآورد کنیم.

یک راهبرد اصلی در آمار استنباطی، محاسبه مقدار تفاوت مابین مشاهداتی است که احتمالاً تنها به طور اتفاقی پدید خواهند آمد. نتیجه این محاسبه را معمولاً واژه خطا می‌نامند. تفاوت موجود بین مشاهده‌ها با واژه خطا مقایسه می‌شود. اگر تفاوت‌های موجود مشابه تفاوت‌هایی باشند که می‌توانست به طور اتفاقی پدید آید، محقق نخواهد توانست این احتمال را که تفاوت‌های موجود صرفاً تابعی از اتفاق هستند، رد کند. اگر تفاوت‌های موجود بزرگتر از واژه خطا باشند؛ محقق برای آن که تعیین کند که نسبت مشاهده به خطا، آن قدر بزرگ هست که وقوع اتفاقی پدیده را در سطح معنی‌داری معین رد کند، به جداول آماری مراجعه می‌کند.

شاخصهایی که بیشترین موارد استفاده را در آمار استنباطی دارند، عبارتند از: آزمون  $t$ ، تحلیل واریانس، و آزمون  $\chi^2$ . از آزمون  $t$  برای تعیین این که آیا تفاوت بین دو میانگین نمونه از نظر آماری معنی‌دار است یا نه، استفاده می‌شود. دو نوع  $t$  وجود دارد: (۱) آزمون  $t$  برای گروه‌های مستقل، هنگامی مورد استفاده قرار می‌گیرد که هر دو نمونه به طور مستقل از یک جامعه آماری انتخاب شده باشند، و (۲) آزمون  $t$  برای گروه‌های نامستقل، هنگامی به کار می‌رود که آزمودنی‌های دو نمونه با هم زوج شده باشند یا آزمودنی‌های یک نمونه دو بار اندازه‌گیری شده باشند. تحلیل واریانس برای مقایسه میانگین‌های دو یا چند نمونه و برای آزمون فرضیه صفر که می‌گوید تفاوت معنی‌داری بین میانگین‌های حاصل از این نمونه‌ها موجود نیست، مورد استفاده قرار می‌گیرد. تحلیل واریانس چند عاملی به ما این توانایی را می‌دهد که اثر بیش از یک متغیر مستقل و اثر تعاملی چنین متغیرهایی را آزمون کنیم.

آماره  $\chi^2$  دو شاخصی است که برای یافتن معنی‌داری تفاوت بین نسبت‌هایی از آزمودنی‌ها، اشیاء، رویدادها و غیره که در طبقه‌بندی‌های مختلفی می‌گنجند، از طریق مقایسه فراوانی‌های مشاهده شده و مورد انتظار به کار می‌رود.

	// مفاهیم اصلی
آزمون ۱ برای نمونه‌های مستقل	آزمون ۱ برای نمونه‌های مستقل
آزمون دو دامنه	آزمون جهت‌دار (یک دامنه)
آزمون معنی‌داری	آزمون فاقد جهت
آزمون یک دامنه	آزمون F
آماره	آزمون z
پارامتر	اثر اصلی
تحلیل واریانس چند عاملی	تحلیل واریانس (ANOVA)
جامعه آماری	تعامل
جامعه آماری هدف	جامعه آماری قابل دسترسی
خطای معیار میانگین	جدول اعداد تصادفی
خطای نوع I و II	خطای نمونه‌گیری
درجه آزادی	خی‌دو
طرح عاملی	سطح معنی‌داری
فراوانی مورد انتظار	فراوانی مشاهده شده
نمونه اریب	فرضیه صفر
نمونه‌گیری احتمالاتی	نمونه‌گیری اتفاقی
نمونه‌گیری سهمیه‌ای	نمونه‌گیری خوشه‌ای
نمونه‌گیری طبقه‌ای	نمونه‌گیری سیستماتیک
نمونه‌گیری هدفمند	نمونه‌گیری غیراحتمالاتی

### /// تمرینها

۱- آیا دقت یک نمونه در زمینه نمایاندن ویژگیهای جامعه آماری مربوط به خود، همیشه با اندازه نمونه افزایش می‌یابد؟ شرح دهید.

۲- از شما خواسته شده است که تعیین کنید آیا آموزگاران ناحیه آموزشی مرکزی، از مفهوم

«تحصیل در تمام طول سال» پشتیبانی می‌کنند یا خیر. از آنجا که این بخش نسبتاً بزرگ است، از

شما درخواست می‌شود که تنها با ۵۰۰ آموزگار تماس بگیرید. تعدادی را که شما از هر یک از سطوح زیر برای گرفتن یک نمونه تصادفی طبقه‌ای متناسب انتخاب خواهید کرد، تعیین کنید.

تعداد کل	سطح
۳۵۰۰	دوره دبستان
۲۱۰۰	دوره راهنمایی
۱۴۰۰	دوره دبیرستان
۷۰۰۰	کل

۳- از شما خواسته می‌شود یک نظرسنجی در دانشگاهی با جامعه آماری ۱۵۰۰۰ دانشجویان انجام دهید. شما برای گرفتن یک نمونه معرف از این دانشجویان برای انجام پیمایش خود چگونه عمل خواهیم کرد؟

۴- هیأت سردبیری یک مجله که در سراسر کشور منتشر می‌شود و یک میلیون نفر مشترک دارد، می‌خواهد بداند کدام یک از جنبه‌های مجله مورد علاقه خوانندگان است و کدام یک نیست؟ هیأت سردبیری به این نتیجه می‌رسد که مصاحبه فردی بهترین روش برای کسب اطلاعات است. به دلایل اجرایی و اقتصادی تنها با ۵۰۰ نفر در پنج شهر مصاحبه خواهد شد. در چنین وضعیتی، مشخص کنید:

الف - جامعه آماری هدف

ب - جامعه آماری قابل دسترسی

ج - نمونه

۵- کدام یک از نمونه‌گیرهای زیر، نمونه‌گیری احتمالاتی است؟ کدام یک نمونه‌گیری غیراحتمالاتی است؟

الف - نمونه تصادفی

ب - نمونه اتفاقی

ج - نمونه خوشه‌ای

د - نمونه طبقه‌ای

ه- نمونه هدفمند

و- نمونه سهمیه‌ای

ز- نمونه سیستماتیک

۶- پژوهشگران می‌خواهند درباره این مسئله که «آیا موطلایی‌ها از زندگی خود لذت بیشتری می‌برند؟» مطالعه کنند.

الف- در این مسئله فرضیه صفر چیست؟

ب- در این مورد خطای نوع I چه خواهد بود؟

ج- در این مورد خطای نوع II چه خواهد بود؟

د- اگر در بررسی این مسئله یک محقق از سطح معنی‌داری  $0/05$  و محقق دیگر از سطح معنی‌داری  $0/01$  استفاده کند، کدام یک بیشتر محتمل است مرتکب خطای نوع اول شود؟

ه- اگر در بررسی این مسئله یک محقق از سطح معنی‌داری  $0/05$  و محقق دیگر از سطح معنی‌داری  $0/01$  استفاده کند، کدام یک بیشتر محتمل است مرتکب خطای نوع دوم شود؟

۷- آمار استنباطی محقق را قادر می‌سازد:

الف- به نتایج بدون اشتباه برسد.

ب- با اطلاعات ناکامل به نتایج منطقی برسد.

ج- به آنچه که صرفاً حدس و گمان است، ظاهر واقعیت بدهد.

۸- برای آن که یک خطای نوع I واقع شود، کدام دو شرط ضرورت دارد؟

۹- کدام یک از عبارات زیر نقش فرضیه صفر را در تحقیق توصیف می‌کند؟

الف- فرضیه صفر ما را قادر می‌سازد هنگامی که هیچ رابطه واقعی بین متغیرها موجود نیست، احتمال وقوع یک رویداد را به طور اتفاقی، تعیین کنیم.

ب- فرضیه صفر ما را قادر می‌سازد وجود یک رابطه واقعی بین متغیرها را ثابت کنیم.

ج- فرضیه صفر ما را قادر می‌سازد عدم یک رابطه واقعی بین متغیرها را ثابت کنیم.

۱۰- یک خطای نوع II هنگامی اتفاق می‌افتد که محقق:

الف- یک فرضیه صفر نادرست را رد می‌کند.

ب- یک فرضیه صفر درست را رد می‌کند.

ج- قبلاً یک خطای نوع I را مرتکب شده است.

د - یک فرضیه صفر نادرست را حفظ می‌کند.

ه - یک فرضیه صفر درست را حفظ می‌کند.

۱۱ - اصطلاح سطح معنی‌داری به کدام یک از عبارات زیر اطلاق می‌شود:

الف - احتمال وقوع یک رویداد تنها بر حسب اتفاق، که بعد از تحلیل داده‌های یک آزمایش محاسبه می‌شود.

ب - احتمال یک خطای نوع I که محقق می‌خواهد آن را بپذیرد.

ج - احتمال واقعی یک خطای نوع II

د - احتمال یک خطای نوع II که محقق می‌خواهد آن را بپذیرد.

۱۲ - یک محقق سطح معنی‌داری را برای استفاده در یک آزمایش چگونه تعیین می‌کند؟

۱۳ - یک صاحب کارخانه دخانیات، محققی را استخدام کرده است تا نرخ بروز سرطان ریه را در بین سیگاریها و غیر سیگاریها مقایسه کند. با در نظر گرفتن نتایج تحقیقات قبلی درباره این مسئله، صاحب کارخانه محقق را ترغیب خواهد کرد تا از کدام یک از خطاهای زیر اجتناب کند:

الف - خطای نوع I

ب - خطای نوع II

۱۴ - وقتی که نتایج یک مطالعه از نظر آماری معنی‌دار نباشد، این امر بر چه چیزی دلالت دارد؟

۱۵ - نمونه‌گیری طبقه‌ای را با نمونه‌گیری سهمیه‌ای مقایسه کنید.

۱۶ - مقدار  $\chi^2$  دو را برای مجموعه داده‌های زیر پیدا کنید. اگر مقدار  $\chi^2$  دو در سطح  $0.05$  معنی‌دار است، آن را نشان دهید.

۲	۸	۱۰
۶	۴	۱۰
۸	۱۲	

۱۷ - ما پاسخهای دو گروه از دانشجویان (رشته تعلیم و تربیت و رشته‌های دیگر) را به یک پرسش مقیاس نگرش نوع لیکرت داریم. مقدار  $\chi^2$  دو را برای این مسئله پیدا کنید و نشان دهید آیا مقدار  $\chi^2$  دو حاصله از نظر آماری معنی‌دار است یا نه.

کاملاً مخالف	مخالف	نامشخص	موافق	کاملاً موافق	
۶	۹	۱۰	۸	۷	دانشجویان رشته تعلیم و تربیت
۷	۸	۷	۱۰	۸	دانشجویان رشته‌های دیگر

۱۸- کدام یک از شیوه‌های آماری زیر برای آزمون فرضیه‌های الف تا ز مناسب هستند؟

آزمون t برای میانگین‌های مستقل

آزمون F برای میانگین‌های نامستقل

تحلیل واریانس عاملی

خی دو

الف - نسبت دانشجویان دوره دکترا در کلاس، که نمره الف می‌گیرند، از نسبت دانشجویان دوره فوق لیسانس که نمره الف می‌گیرند، بزرگتر خواهد بود.

ب - نمره میانگین آن گروه از دانشجویان که در ساعت ۱ بعدازظهر امتحان نهایی می‌دهند، از نمره میانگین دانشجویانی که در ساعت ۷ بعدازظهر امتحان می‌دهند، به طور معنی‌داری بیشتر خواهد بود.

ج - برای دانشجویان دختر، شرکت در کلاس عصر، مزیتی در جهت کسب نمره بهتر است، در حالی که برای دانشجویان پسر مانعی برای کسب نمره بهتر محسوب می‌شود.

د - اگر همان امتحان آماری را که در اولین روز کلاس از دانشجویان گرفته‌اند، در روز آخر کلاس نیز تکرار کنند، دانشجویان پیشرفت تحصیلی معنی‌داری را نشان خواهند داد.

ه - بین پذیرفته شدن یا رد شدن در کلاس و اظهار رضایت یا عدم رضایت از دوره درسی رابطه وجود دارد.

و - هر تفاوتی بین میانگین نمره‌های کلاس ترم پاییز و کلاس ترم تابستانی یک استاد را می‌توان بسادگی برحسب اتفاق دانست.

ز - بین تجربه تدریس و تجربه تحقیق و موفقیت در یک دوره درسی تحقیق، اثر تعاملی معنی‌داری وجود دارد.

۱۹- تفاوت یک آزمون جهت‌دار با یک آزمون فاقد جهت در این است که در آزمون جهت‌دار

محقق به تغییراتی علاقه‌مند است که:

الف - تنها در جهت مثبت صورت می‌گیرد.

ب - تنها در جهت منفی صورت می‌گیرد.

ج - در هر دو جهت مثبت و منفی صورت می‌گیرد.

د - یا در جهت مثبت یا در جهت منفی صورت می‌گیرد.

۲۰- دو گروه که به طور تصادفی انتخاب شده‌اند، در یک آزمایش شرکت کردند که در آن آزمایش، گروه I تیمار را دریافت کرده و گروه II تیمار را دریافت نکرده است. فرضیه محقق این است که عملکرد متوسط گروه I بالاتر از عملکرد متوسط گروه II خواهد بود. آزمون 1 را در مورد اطلاعات زیر به کار ببرید و بگویید که آیا فرضیه محقق می‌تواند تایید شود یا خیر.

$\sum x^2$	N	X	
۳۸۲/۰۲	۳۰	۴۵/۳۲	گروه I
۲۶۴/۳۲	۳۰	۴۱/۷۸	گروه II

۲۱- محقق می‌خواهد این فرضیه را که همبستگی بین متغیر A و متغیر B به طور معنی‌داری بزرگتر از صفر است، آزمون کند. او با استفاده از ۲۲ آزمودنی  $r = ۰/۲۱$  را به دست آورده است. برای آن که دریابید آیا فرضیه مزبور را می‌توان در سطح  $۰/۰۵$  رد کرد (یک دامنه)، از جدول A-۵ استفاده کنید.

۲۲- جدول زیر داده‌های یک طرح آزمایشی  $۲ \times ۳$  است. برای آزمون معنی‌داری اثرهای اصلی و اثر تعاملی، تحلیل واریانس چند عاملی را به کار ببرید.

ستونها

		A	B	C
ردیف‌ها	A	۲۵ و ۲۳ و ۲۰ و ۱۷ و ۱۵	۲۲ و ۲۰ و ۱۸ و ۱۶ و ۱۴	۲۰ و ۱۸ و ۱۶ و ۱۴ و ۱۲
	B	۱۶ و ۱۴ و ۱۲ و ۱۰ و ۸	۱۸ و ۱۶ و ۱۴ و ۱۲ و ۱۰	۱۹ و ۱۸ و ۱۶ و ۱۴ و ۱۳



جدولی تهیه کنید که در آن مجموع مربعات (SS)، درجه آزادی (df)، میانگین مربعات (MS) و مقادیر F نشان داده شده است. سپس به پرسشهای زیر پاسخ دهید:

الف - کدام مقادیر F و در چه سطحی معنی دار هستند؟

ب - در این مسئله چند فرضیه صفر آزمون شدند؟

ج - ما چند تا از این فرضیه‌ها را می‌توانیم رد کنیم؟

۲۳- فهرستی از دانش‌آموزان یک دبیرستان دارید که اعداد ۱ تا ۱۰۰۰ به آنها تخصیص داده شده است. برای انتخاب یک نمونه ۵۰ نفری از این فهرست فرضی از جدول اعداد تصادفی ضمیمه کتاب استفاده کنید. اعدادی را که برای نمونه انتخاب کرده‌اید، ذکر کنید.

### /// پاسخها

۱- یک نمونه تصادفی بزرگتر، بیشتر از یک نمونه تصادفی کوچکتر محتمل است که معرف جامعه آماری باشد. یک نمونه بزرگ که روش انتخاب آن اجازه آریبی نظام‌مند را بدهد، بیشتر از یک نمونه آریب کوچک، معرف جامعه آماری نخواهد بود.

۲- برای به دست آوردن یک نمونه طبقه‌ای متناسب، ۵۰۰ آموزگار را به نسبت بازنمایی آنها در جامعه آماری تقسیم کنید، به صورت زیر عمل کنید:

$3500 / 7000 \times 500 = 250$	دوره دبستان
$2100 / 7000 \times 500 = 150$	دوره راهنمایی
$1400 / 7000 \times 500 = 100$	دوره دبیرستان
$500$	کل نمونه

۳- فهرستی از تمام دانشجویان تهیه کنید و به اسامی آنها شماره بدهید. سپس با استفاده از جدول اعداد تصادفی، یک نمونه تصادفی را به اندازه مورد نظر انتخاب کنید. یک نقطه جدول را به طور تصادفی انتخاب کنید و سپس روی ستون اعداد به بالا یا پایین بروید و دانشجویانی را که شماره آنها در ستون اعداد تصادفی وجود دارد، انتخاب کنید.

۴- الف - تمام مشترکین مجله

ب - مشترکین در پنج شهر

ج - ۵۰۰ نفر که مصاحبه می‌شوند.

۵- نمونه‌های احتمالاتی: الف، ج، د، ز

نمونه‌های غیر احتمالاتی: ب، ه، و

۶- الف - هیچ رابطه‌ای بین رنگ مو و لذت بردن از زندگی وجود ندارد.

ب - محققان دچار خطای نوع I می‌شوند، اگر اعلام کنند موطلائی‌ها بیشتر از غیر موطلائی‌ها از زندگی لذت می‌برند یا موطلائی‌ها کمتر از غیر موطلائی‌ها از زندگی لذت می‌برند، در حالی که در واقع هر دو گروه به یک اندازه از زندگی لذت می‌برند.

ج - محققان دچار خطای نوع II می‌شوند، اگر نتوانند نتیجه بگیرند که موطلائی‌ها لذت بیشتر یا کمتر می‌برند، در حالی که واقعیت چنین باشد.

د - محققى که از سطح معنی‌داری ۰/۰۵ استفاده می‌کند.

ه - محققى که از سطح معنی‌داری ۰/۰۰۱ استفاده می‌کند.

۷- ب

۸- فرضیه صفر باید درست باشد و محقق باید آن را رد کند.

۹- الف

۱۰- د

۱۱- ب

۱۲- با سبک و سنگین کردن پیامدهای خطای نوع I و خطای نوع II

۱۳- الف

۱۴- نتایج بسادگی می‌توانند تابعی از اتفاق باشند، شواهد برای توجیه یک نتیجه‌گیری

ناکافی است.

۱۵- در نمونه‌گیری طبقه‌ای، معرف بودن در هر طبقه، با استفاده از انتخاب تصادفی به دست

می‌آید، در حالی که در نمونه‌گیری سهمیه‌ای معرف بودن در طبقات مختلف از راه قضاوت حاصل می‌شود.

۱۶- ۳/۳۳ معنی‌دار نیست.

۱۷- ۰/۹۵۴ معنی‌دار نیست.

۱۸- الف - خی‌دو

ب - آزمون ۱ برای میانگینهای مستقل

ج - تحلیل واریانس عاملی

د - آزمون  $t$  برای میانگینهای نامستقل

ه -  $\chi^2$  دو

و - آزمون  $t$  برای میانگینهای مستقل

ز - تحلیل واریانس عاملی

۱۹- د

۲۰-  $df = 58$  و  $t = 4/11$  معنی‌دار در سطح  $0/001$

۲۱- با درجه آزادی ۲۲، به یک  $F$  مساوی با  $0/4227$  یا بزرگتر از آن نیاز داریم تا فرضیه صفر

را رد کنیم. بنابراین، فرضیه صفر رد نمی‌شود.

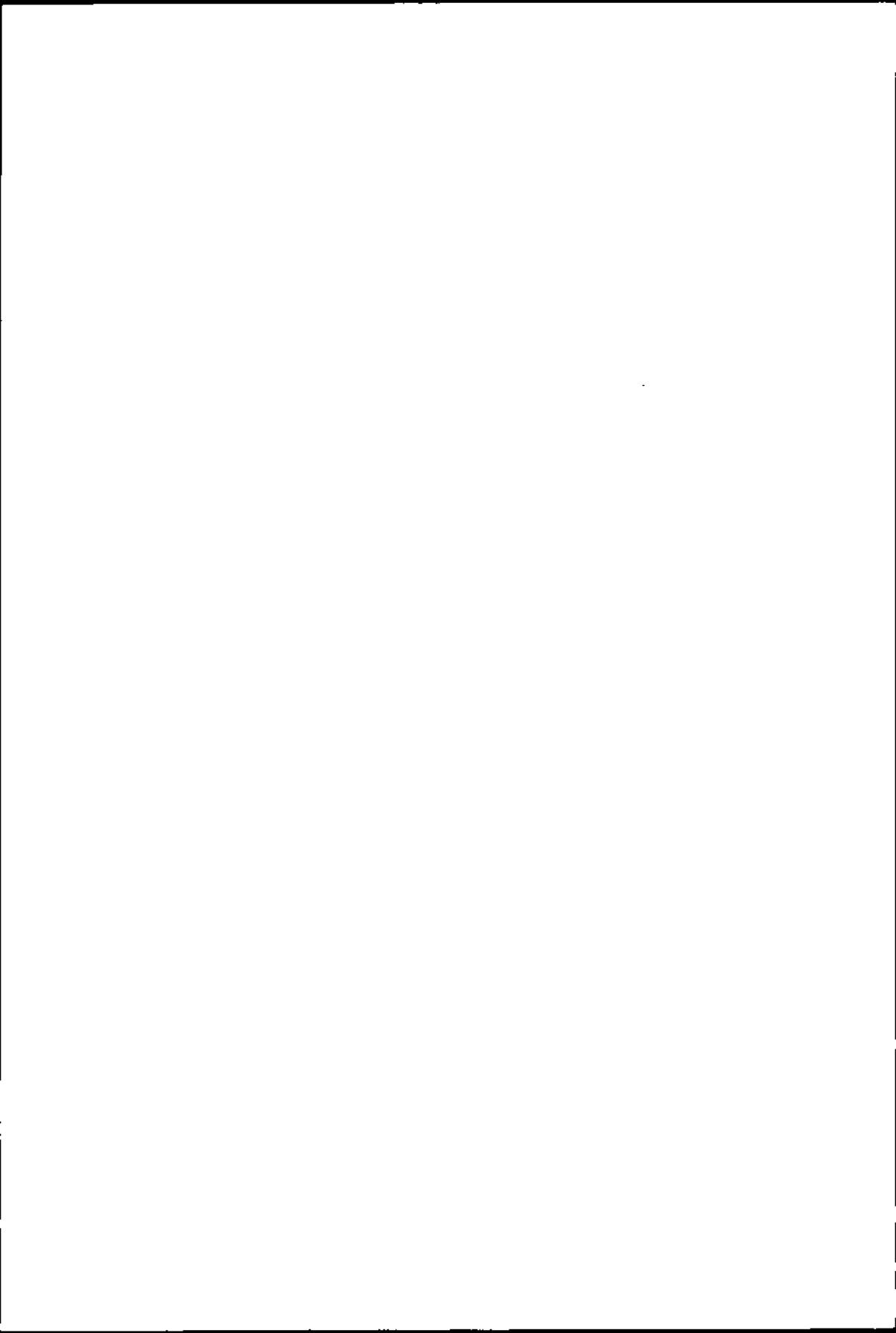
۲۲-

منبع واریانس	SS	df	MS	F
بین ستونها	۰	۲	۰	۰
بین ردیفها	۱۲۰	۱	۱۲۰	۱۱/۳۴
تعامل	۸۰	۲	۴۰	۳/۷۸
بین گروهها	۲۰۰	۵	۴۰	۳/۷۸
داخل گروهها	۲۵۴	۲۴	۱۰/۵۸	-

الف - بین ردیفها در سطح  $0/01$  و تعامل در سطح  $0/05$

ب - ۳

ج - ۲



# بخش چهارم

## مبانی اندازه‌گیری

### فصل هفتم

#### ابزار تحقیق

##### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- نقش اندازه‌گیری را در تحقیق توضیح دهد.
- ۲- از سالنامه اندازه‌گیریهای روانی بوروس برای دستیابی به داده‌های ضروری در ارزشیابی آزمونهای استاندارد شده و دیگر وسائل اندازه‌گیری استفاده کند.
- ۳- تفاوت بین آزمون و مقیاس را شرح دهد.
- ۴- تفاوت بین آزمونهای هنجاری<sup>(۱)</sup> و آزمونهای ملاکی<sup>(۲)</sup> را تشخیص دهد.
- ۵- تفاوت بین اندازه‌های استعداد و پیشرفت را تشخیص دهد.
- ۶- مراحل آماده‌سازی مقیاس لیکرت<sup>(۳)</sup> را برای اندازه‌گیری نگرشها توصیف کند.
- ۷- مراحل آماده‌سازی مقیاس تورستون<sup>(۴)</sup> را توصیف کند.
- ۸- ویژگیهای مقیاس لیکرت و تورستون را مقایسه کند.

1- Norm- referenced tests

2- Criterion Referenced Tests

3- Likert Scale

4- Thurstone scale

- ۹- مقیاس گاتمن<sup>(۱)</sup> و مزیت‌های آن را توصیف کند.
- ۱۰- ویژگی‌های مقیاس افتراق معنایی را توصیف کند.
- ۱۱- انواع خطاهای رایج در مقیاس‌های درجه‌بندی را بیان کند.
- ۱۲- مزایا و نقایص اندازه‌های خودسنجی شخصیت را بیان کند.
- ۱۳- حداقل پنج رهنمود را که یک محقق باید به هنگام استفاده از مشاهده مستقیم، به عنوان یک روش گردآوری داده‌ها، از آنها پیروی کند، فهرست کند.
- ۱۴- آزمون موقعیتی را تعریف کند و بگوید در تحقیق چه موقع می‌توان از آن استفاده کرد.
- ۱۵- ویژگی‌های اصلی یک روش فرافکن را بیان کند و حداقل دو روش فرافکن شناخته شده را نام ببرد.

یکی از هدفهای تحقیق در زمینه تعلیم و تربیت، کسب آگاهی بیشتر از روابط بین متغیرها، در جامعه‌های آماری مختلف است. برای مثال، ممکن است کسی بپرسد: «در بچه‌های شش ساله، چه رابطه‌ای بین هوش و خلاقیت وجود دارد؟» ما به طور مستقیم نه می‌توانیم هوش و خلاقیت را مشاهده کنیم و نه می‌توانیم همه کودکان شش ساله را مشاهده کنیم. اما این به آن معنا نیست که باید درباره این مسئله و مسائل شبیه آن در جهل باقی بمانیم. ما نشانگرهایی در دست داریم که ساختار هوش و خلاقیت را به طور تقریب برآورد می‌کنند، یعنی رفتارهای قابل مشاهده‌ای وجود دارند که به عنوان شاخصهای معتبر این ساختارها پذیرفته شده‌اند. یکی از مهم‌ترین وظایف محقق انتخاب یا تهیه مقیاسها و وسایلی است که می‌توانند ویژگیهایی مانند هوش، پیشرفت تحصیلی، شخصیت، انگیزش، نگرشها، استعدادها، علایق، خود پنداره<sup>(۱)</sup> و غیره را اندازه بگیرند. برای کمی کردن کیفیت، انواع مختلفی از وسایل مورد استفاده قرار می‌گیرند. برای به دست آوردن این اندازه‌ها دو روش اصلی وجود دارد: (۱) محقق از وسیله‌ای که قبلاً ساخته شده است، استفاده می‌کند و (۲) محقق، خود، وسیله اندازه‌گیری لازم را می‌سازد.

محقق برای انتخاب یک وسیله اندازه‌گیری، باید تحقیقاتی را که در زمینه مورد نظر او منتشر شده‌اند، مطالعه کند. او باید ببیند که محققان دیگر برای اندازه‌گیری سازه مورد نظر او از چه استفاده کرده‌اند. این محققان به طور کلی خاطر نشان خواهند کرد که وسیله اندازه‌گیری آنان خوب عمل کرده است، یا بهتر است از شیوه‌های دیگری استفاده شود. سالنامه اندازه‌گیریهای روانی بوروس و مجموعه آزمونهای منتشره شماره سه، که در فصل سوم توصیف شدند، منابع سودمند دیگری برای شناسایی وسایل اندازه‌گیری منتشر شده می‌باشند. منبع اطلاعاتی مفید دیگر در مورد آزمونها، مجموعه آزمونهای موسسه آزمونهای آموزشی (ETS)<sup>(۲)</sup> است. مجموعه آزمونهای ETS، مجموعه‌ای است مشتمل بر بیش از ۱۶۰۰۰ آزمون تجارتي و دیگر وسایل اندازه‌گیری که به منظور فراهم آوردن اطلاعات روز در زمینه آزمون سازی برای محققان آموزشی طراحی شده است. ETS بر خلاف بوروس، هم آزمونهای منتشر شده و هم آزمونهای منتشر نشده را فهرست بندی می‌کند. بخش جدیدی که اخیراً به این مجموعه اضافه شده است،

فهرست تجمعی آزمونها به صورت میکروفیش، ۸۷-۱۹۷۵، نام دارد، که آزمونهای محققان را برحسب مؤلف، عنوان و موضوع، طبقه‌بندی می‌کند. در فهرست موضوعی بیش از ۷۵۰ عنوان وجود دارد که هر چیزی را، از استدلال انتزاعی تا نگرشهای مربوط به شغل، شامل می‌شود. اگر محقق نتواند وسیله اندازه‌گیری را که قبلاً ساخته شده است، پیدا کند، آن‌گاه باید خود آن را بسازد. ساختن وسیله اندازه‌گیری مستلزم شناسایی و استفاده از رفتاری است که می‌تواند نشانگر حضور سازه مورد نظر محسوب شود. محقق برای یافتن این نشانگرها، ابتدا باید به نظریه‌ای که پشتوانه تحقیق است، روی آورد. یک نظریه خوب، به طور کلی گویای آن خواهد بود که سازه مورد نظر، خود را چگونه آشکار خواهد کرد و چه تغییراتی را می‌توان در آن مشاهده کرد؛ به عبارت دیگر راههایی را برای اندازه‌گیری سازه (ها) پیشنهاد خواهد کرد. محقق همچنین می‌تواند از تجربه و مهارت خود برای تصمیم‌گیری در مورد نشانگرهای مناسب سازه استفاده کند. ما در بخشهای بعدی، برخی از انواع وسایل اندازه‌گیری را که در تحقیقات آموزشی مورد استفاده قرار می‌گیرند، مورد بحث قرار خواهیم داد.

### /// وسایل اندازه‌گیری // آزمونها

آزمونها وسایل اندازه‌گیری ارزشمندی برای تحقیق در زمینه تعلیم و تربیت هستند. یک آزمون مجموعه‌ای از محرکها است که برای دریافت پاسخ به فرد ارائه می‌شود تا بتوان بر اساس آن پاسخها یک نمره عددی تعیین کرد. این نمره مبتنی بر یک نمونه معرّف از رفتار فرد و نشانگری است از این که آن فرد تا چه حد خصلت اندازه‌گیری شده را داراست. سودمندی این نمره‌ها، به عنوان نشانگرهای سازه مورد نظر، تا حد زیادی تابع عینیت، اعتبار و پایایی آزمونهاست. عینیت به میزان توافق در میان نمره‌ها اطلاق می‌شود. برخی از آزمونها، نظیر آزمونهای چند گزینه‌ای و درست - نادرست، آزمونهای عینی به شمار می‌آیند؛ زیرا نمره‌گذاری آنها از طریق مقایسه پاسخهای فرد با کلید نمره‌گذاری صورت می‌گیرد و نمره‌گذار خود در این مورد تصمیم نمی‌گیرد. آزمونهای انشایی از عینیت کمتری برخوردار هستند؛ زیرا نمره‌گذاری آنها تحت تأثیر قضاوت و نظرات نمره‌گذاران قرار می‌گیرد. اعتبار و پایایی در فصل ۸ مورد بحث قرار خواهند گرفت.



## / آزمونهای پیشرفت

متغیر وابسته، در اکثر تحقیقاتی که با مسئله تأثیر روشهای آموزشی سروکار دارند، پیشرفت است. بنابراین آزمونهای پیشرفت به فراوانی در تحقیقات مربوط به تعلیم و تربیت و همچنین در نظامهای مدرسه‌ای به کار می‌رود. این آزمونها تسلط و تبحر افراد را در حیطه‌های مختلف دانش می‌سنجند.

به طور کلی، آزمونهای پیشرفت به دو طبقه استاندارد شده (میزان شده) و ساخته آموزگار یا ساخته پژوهشگر تقسیم می‌شوند. آزمونهای استاندارد شده آزمونهای منتشرشده‌ای هستند که از یک فعالیت دقیق و ماهرانه حاصل شده‌اند و هدفهای آموزشی گسترده‌ای را پوشش می‌دهند که در بسیاری از نظامهای مدرسه‌ای مشترک هستند. برای این آزمونها، هنجارهای تطبیقی استخراج، اعتبار و پایایی تعیین و رهنمودهای اجرا و نمره‌گذاری تبیین شده است. رهنمودهای اجرایی آزمون در دستورالعمل‌هایی که ناشران آزمون تهیه کرده‌اند، درج شده است.

سازندگان آزمونها، برای آن که هنجارهایی برای آن آزمونها تعیین کنند، آنها را با یک نمونه معرف و مرتبط اجرا می‌کنند. گروه هنجاری ممکن است به گونه‌ای انتخاب شود که معرف تمام کشور یا ایالت، شهر، ناحیه یا مدرسه باشد.

میانگین نمونه برای یک سطح تحصیلی خاص، هنجار آن سطح تحصیلی برای تمام جامعه آماری می‌شود. مهارت‌های اندازه‌گیری شده لزوماً آن چیزهایی نیستند که باید در هر سطح تحصیلی آموخته شوند، بلکه استفاده از هنجارها به متخصصان تعلیم و تربیت مبنایی برای مقایسه گروه خود با برآورد میانگین همه کودکان آن سطح تحصیلی می‌دهد.

آزمونهای استاندارد شده برای موضوعهای مجزای مدرسه‌ای (دروس مجزای کلاسی) نظیر ریاضیات و شیمی و همچنین به شکل مجموعه آزمونهای جامع، که حیطه‌های مختلف پیشرفت را می‌سنجند، در دسترس هستند. برای مثال، آزمون پیشرفت کالیفرنیا شامل آزمونهایی در حیطه‌های خواندن، زبان و حساب است. آزمونهای زنجیره‌ای پیشرفت آموزشی<sup>(۱)</sup> شامل آزمونهایی برای هفت حیطه است.

پژوهشگر در انتخاب یک آزمون پیشرفت باید دقت به خرج دهد تا آزمونی را برگزیند که هم پایا و هم برای اندازه‌گیری آن جنبه از پیشرفت که منظور اوست، مناسب باشد. به هر حال، گاهی

از مواقع او قادر به انتخاب آزمون نخواهد بود، بلکه اجبار خواهد داشت از آزمونی که تشکیلات آموزشی قبلاً انتخاب کرده است، استفاده کند. سالنامه اندازه‌گیریهای روانی، فهرست جامعی را ارائه می‌کند که مروری بر آزمونهای پیشرفت موجود را به همراه دارد.

اگر آزمونی که در دسترس است، رفتار مورد نظر را اندازه‌گیری کند و اگر پایایی و هنجارهای آن برای منظور ما کفایت کند، استفاده از این ابزار استاندارد شده مزایایی دربرخواهد داشت. علاوه بر زمان و تلاشی که صرفه‌جویی می‌شود، پژوهشگر مزیت پیوستگی روش آزمون کردن را در خواهد یافت؛ یعنی می‌توان نتایج بررسی او را با نتایجی که دیگران با استفاده از همان ابزار به دست آورده‌اند، مقایسه و تفسیر کرد.

هنگامی که استفاده از یک آزمون پیشرفت استاندارد شده برای هدفهای خاص یک بررسی تحقیقاتی مناسب به نظر نرسد، امکان دارد پژوهشگر خود آزمون لازم را بسازد. ساختن آزمون مورد نظر بسیار بهتر است تا استفاده از آزمون استاندارد شده نامناسب، تنها به این دلیل که به آن دسترسی داریم. در این حالت، پژوهشگر باید دقت زیادی در آماده‌سازی آزمون به کار بندد؛ بویژه در تعیین اعتبار و پایایی آن قبل از این که آزمون را به کار برد. برای اطلاعات بیشتر در زمینه آزمون سازی می‌توان به متون تخصصی در مورد سنجش، نظیر کتابهای گرون لوند، ساکس، یا تورندایک و هاگن<sup>(۱)</sup> مراجعه کرد.

آزمونهای استاندارد شده و معلم ساخته را، بر مبنای نوع تفسیری که از آن به عمل می‌آید، می‌توان همچنین به آزمونهای هنجاری و آزمونهای ملاکی طبقه‌بندی کرد.<sup>(۲)</sup> آزمونهای هنجاری امکان مقایسه عملکرد یک فرد را با عملکرد افراد دیگر در آزمون فراهم می‌آورند. عملکرد فرد برحسب جایگاه نسبی او در یک گروه مشخص، که گروه هنجاری نامیده می‌شود، تفسیر و برحسب صدکها، نمره‌های استاندارد و غیره گزارش می‌شود.

از سوی دیگر، آزمونهای ملاکی این امکان را فراهم می‌آورند تا بدون آن که به عملکرد دیگران رجوع کنیم، به طور دقیق توصیف کنیم که یک فرد چه می‌تواند انجام دهد. عملکرد بر

1- Gronlund, N E, and R.L. Line (1990). *Measurement and Evaluation in Teaching* (6th ed.). New York: Macmillan; Sax, G. (1960). *Principles of Educational and Psychological Measurement and Evaluation in Psychology and Education* (4th ed.). New York: Wiley.

۲ - رای مقاله بیشتر درباره اندازه‌های ملاکی مراجعه کنید به:

R.A Berk (1984), *A Guide to Criterion Referenced Test Construction*, Baltimore: Johns Hopkins University Press; and W.J. Popham (1981), *Modern Educational Measurement*, Englewood Cliffs, NJ: Prentice - Hall.

حسب سطح تسلط بر یک حیطه محتوایی / مهارتی مشخص گزارش می‌شود. به طور نوعی، سطح تسلط با درصد پاسخهای درست نشان داده می‌شود. در تفسیر عملکرد فرد از نمره‌های حداقل، که از پیش تعیین شده‌اند، استفاده می‌شود.

محقق، قبل از طراحی وسیله اندازه‌گیری، باید از نوع تفسیری که قرار است به عمل آید، آگاه باشد. در آزمونهای هنجاری، پرسشهایی انتخاب می‌شوند که دامنه وسیعی از نمره‌ها را حاصل می‌کنند. دشواری پرسشها و توان ایجاد افتراق پرسشها در بین افراد، باید مورد توجه محقق باشد. در آزمونهای ملاکی، پرسشها صرفاً بر مبنای نحوه اندازه‌گیری مجموعه مشخصی از هدفهای آموزشی انتخاب می‌شوند. آزمون، بسته به این که چه چیزی را اندازه‌گیری می‌کند، می‌تواند آسان یا دشوار باشد. دلمشغولی اصلی، داشتن یک نمونه معرف از پرسشهایی است که هدفهای بیان شده را اندازه‌گیری کنند؛ به گونه‌ای که امکان توصیف عملکرد را برحسب آگاهی و مهارت مشخصی که افراد می‌توانند کسب کنند، به ما بدهد.

گاهی معدل عددی دروس مختلف به عنوان معیار موفقیت و پیشرفت تحصیلی به کار می‌رود. هنگام استفاده از معدل عددی دانش‌آموزان نظامهای مختلف درسی، احتیاط، شرط لازم است. نمرات الفبایی در مدارس مختلف لزوماً معنای همسان ندارند و نمی‌توان اطمینان داشت که تسلط و تبحر یکسانی را نشان می‌دهند.

### آزمونهای هوش /

آزمونهای هوش از این نظر با آزمونهای پیشرفت تفاوت دارند که در آزمونهای هوش، تلاش در جهت اندازه‌گیری عملکرد کلی است؛ در حالی که در آزمونهای پیشرفت، کوشش برای اندازه‌گیری عملکرد در حیطه‌های خاص صورت می‌گیرد. آزمونهای هوش سعی می‌کنند توانایی آزمودنی را در درک رابطه‌ها، حل مسئله‌ها و بکارگیری آگاهیها در زمینه‌های متنوع بسنجند.

آزمونهای هوش را نباید اندازه‌های هوش فطری یا «خالص» تصور کرد. عملکرد یک آزمودنی در چنین آزمونهایی، تا حدودی به زمینه فردی و سابقه تحصیلی او بستگی دارد. در سالهای اخیر استفاده از واژه هوش برای توصیف این آزمونها، به دلیل اختلاف نظر درباره مفهوم هوش و نیز به این دلیل که مردم تمایل دارند هوش را با تواناییهای ارثی مرتبط سازند، رو به کاستی گذارده است. اکنون، اصطلاح استعداد تحصیلی جایگزین واژه هوش شده است - یک

اصطلاح بیشتر توصیفی که به طور مشخص به کارکرد اصلی این آزمونها، که پیش‌بینی عملکرد تحصیلی است، اشاره دارد.

متخصصان تعلیم و تربیت این آزمونها را برای پیش‌بینی موفقیت تحصیلی مفید و به طور کلی معتبر تشخیص داده‌اند. محققان نیز به طور گسترده از این آزمونها استفاده می‌کنند. هوش یک متغیر مستقل است که معمولاً باید در آزمایشهای آموزشی کنترل شود. محقق برای کنترل این متغیر، از نمره‌های چند آزمون استعداد تحصیلی استفاده می‌کند. تعدادی از آزمونهای موجود برای استفاده انفرادی و بقیه، برای استفاده گروهی طراحی شده‌اند.

**آزمونهای انفرادی هوش آزمون استانفورد - بینه و سه آزمون وکسلر رایج‌ترین آزمونهای انفرادی هوش هستند.** آزمون استانفورد - بینه، که امروز معمول است، حاصل تجدید نظرهای متعدد در آزمونی است که برای اولین بار آلفرد بینه در فرانسه برای شناسایی آن دسته از کودکانی که احتمالاً نمی‌توانند از آموزشهای کلاسی عادی استفاده کنند، تهیه کرد. این آزمون، که اول بار درست در آستانه ورود به قرن حاضر ساخته شد، در اصل برای اندازه‌گیری سن عقلی فرد به کار می‌رفت. بعدها مفهوم هوشیهر<sup>(۱)</sup> مطرح شد. این بهر، از تقسیم سن عقلی (MA)<sup>(۲)</sup> بر سن تقویمی (CA)<sup>(۳)</sup> و ضرب آن در ۱۰۰ حاصل می‌شود. نسخه فعلی استانفورد - بینه سن عقلی را به دست می‌دهد، اما نسبت MA/CA را برای تعیین IQ به کار نمی‌برد. به جای آن، IQ را از طریق مقایسه عملکرد یک فرد (نمره او) با هنجارهای حاصله از همان گروه سنی با استفاده از نمره‌های استاندارد تعیین می‌کند (به فصل ۵ مراجعه کنید). یک ویژگی اساسی آزمونهای استانفورد - بینه آن است که یک اندازه عمومی از هوش به دست می‌دهد. این آزمون برای اندازه‌گیری تواناییهای مجزا، آن گونه که برخی دیگر از آزمونها انجام می‌دهند، تلاشی نمی‌کند. آزمونهایی که دیوید وکسلر برای اندازه‌گیری هوش تهیه کرد، اکنون در اشکال مختلفی در دسترس قرار دارد: مقیاس هوش وکسلر برای کودکان (WISC)<sup>(۴)</sup>، که در سال ۱۹۷۴ در آن تجدید نظر شد و اکنون WISC - R نامیده می‌شود، مقیاس هوش وکسلر برای بزرگسالان (WAIS)<sup>(۵)</sup>، و مقیاس هوش وکسلر برای کودکان پیش دبستانی و دبستانی (WPPSI)<sup>(۶)</sup> که برای کودکان ۴ تا ۶/۵ سال عرضه شد.

1- Intelligence Quotient (IQ)

3- Chronological age

5- Wechsler Adult Intelligence Scale

2- Mental age

4- Wechsler Intelligence Scale for Children

6- Wechsler Preschool and Primary Scale of Intelligence

یک خصلت مهم آزمونهای وکسلر تقسیم آنها به دو خرده آزمون<sup>(۱)</sup> است که امکان دستیابی به دو نمره برای هر آزمودنی را فراهم می‌آورد، یک نمره برای هوشبهر کلامی و دیگری برای هوشبهر غیر کلامی. این خرده آزمونها سپس به خرده مقیاسهایی<sup>(۲)</sup> تقسیم می‌شوند که عملکرد امتحان شونده را در تکالیف ویژه نشان می‌دهند.

آزمونهای گروهی هوش آزمون استانفورد - بینه یا وکسلر را باید یک متخصص روان‌سنجی، که برای این کار تربیت شده است، با یک آزمودنی منفرد اجرا کند و این روشی است که به زمان طولانی و هزینه زیاد نیازمند است. هنگامی که سنجش هوش برای گروه بزرگی از افراد مورد نظر باشد، این آزمونها غیر عملی هستند. در چنین موقعیتی، آزمونهای گروهی هوش مورد استفاده قرار می‌گیرند. اولین آزمون گروهی توانایی ذهنی<sup>(۳)</sup> در دوران جنگ اول جهانی برای اندازه‌گیری هوش مردانی که به خدمت نظام احضار می‌شدند، ساخته شد. یک نمونه از این آزمون (آزمون آلفای نظامی)<sup>(۴)</sup> بعد از جنگ برای استفاده غیر نظامی منتشر و سرمشقی برای تعدادی آزمون گروهی شد. امروز، آزمونهای گروهی هوش بسیاری در دسترس است. برای اطلاعات بیشتر درباره آزمونهای هوش به کتاب آناستازی<sup>(۵)</sup> مراجعه کنید.

### // اندازه‌های شخصیت

دستیابی به اندازه‌های شخصیت، حیطة مورد علاقه دیگری برای محققان مسائل تعلیم و تربیت است. برای اندازه‌های شخصیت انواع مختلفی وجود دارد که هر یک دیدگاه نظریه‌ای متفاوتی را منعکس می‌کنند. برخی از آنها انعکاسی از نظریه و ویژگی و نظریه سنخ هستند، در حالی که برخی دیگر ریشه در نظریه‌های روانکاوی و انگیزشی دارند. پژوهشگر باید دقیقاً بداند که چه چیزی را می‌خواهد اندازه بگیرد و سپس ابزار آن را انتخاب کند، در عین حال به شواهد مربوط به اعتبار آن ابزار توجه ویژه‌ای مبذول دارد. دو نوع اندازه شخصیت که گسترده‌ترین استفاده را در تحقیق دارند، عبارتند از: پرسشنامه‌ها و روشهای فراقکن.

1- Subtest

2- SubScale

3- Mental capacity

4- The Army Alpha

5- Anastasi, A. (1988). Psychological Testing (8th ed.). New York: Macmillan.

## پرسشنامه‌ها

در یک پرسشنامه، مجموعه جامعی از عباراتی که الگوهای رفتاری را توصیف می‌کنند، به آزمودنی ارائه و از او خواسته می‌شود که نشان دهد آیا هر عبارت جزو ویژگیهای اوست یا خیر. او می‌تواند «بله»، «خیر»، یا «نامشخص» را علامت بزند.

نمرهٔ آزمودنی با شمارش پاسخهای موافق در مورد خصوصیتی که آزمایشنده در تلاش اندازه‌گیری آن است، محاسبه می‌شود. برای مثال، انتظار خواهد رفت که یک شخص پارانویایی به عبارت «مردم همیشه پشت سر من حرف می‌زنند»، پاسخ «بله» و به عبارت «ماموران پلیس عادل و منطقی باشند» پاسخ «خیر» بدهد. البته، چنین پاسخهایی به تنها دو پرسش، تمایلات پارانویایی را نشان نخواهد داد. به هر ترتیب، پاسخهایی این چنین، به تعداد زیادی از این نوع پرسش‌ها، می‌توانست نشانگر پارانویا تصور شود. برخی از پرسشنامه‌های خودسنجی<sup>(۱)</sup> تنها یک ویژگی را اندازه‌می‌گیرند، نظیر مقیاس F کالیفرنیا<sup>(۲)</sup>، که قدرت طلبی را می‌سنجد. پرسشنامه‌های دیگر، نظیر پرسشنامه شانزده عامل شخصیتی کتل<sup>(۳)</sup> تعدادی از ویژگیها را اندازه می‌گیرند. دیگر پرسشنامه‌های مشهور در امر تحقیق عبارتند از: شخصیت سنج چند وجهی مینه سوتا،<sup>(۴)</sup> خوی سنج گیلفورد-زیمرمن،<sup>(۵)</sup> مشکل سنج مونی<sup>(۶)</sup>، مقیاس ادواردز برای رجحانهای شخصی<sup>(۷)</sup>، سنخ نمای مایرز-بریگز<sup>(۸)</sup> و پرسشنامه رغبت استرانگ.<sup>(۹)</sup>

در تحقیقات آموزشی از پرسشنامه‌ها برای دستیابی به توصیف ویژگیهای گروههای معین، نظیر کم‌آموزها، ترک تحصیل‌کنندگان، اعضای گروههای اقلیت و غیره استفاده شده است. همچنین این پرسشنامه‌ها برای تحقیق در مسائلی مثل روابط متقابل بین صفات شخصیتی و متغیرهایی مانند هوش، پیشرفت و نگرشها مورد استفاده قرار می‌گیرند. پرسشنامه‌ها مزایای اقتصادی بودن، سادگی و عینیت را دارا هستند. بیشترین عیوب آنها به مسئله اعتبار مربوط می‌شود. اعتبار پرسشنامه‌ها، تا حدی به توانایی پاسخگو در خواندن و فهمیدن پرسشها، درک او از خودش و بویژه تمایل او به ارائه پاسخهای صریح و صادقانه بستگی دارد. به عنوان نتیجه، اطلاعاتی که از پرسشنامه‌ها به دست می‌آید، ممکن است سطحی یا اریب باشند. هنگامی که از نتایج حاصله از این ابزارها استفاده می‌کنیم، باید چنین امکانی را در نظر بگیریم.

1- Self-report inventories

3- Cattell's Sixteen Personality Factor Questionnaire

5- Guilford-Zimmerman Temperament Survey

7- Edwards Personal Preference Schedule

9- Strong Interest Inventory

2- California F-Scale

4- Minnesota Multiphasic Personality Inventory

6- Mooney Problem Check list

8- Myers-Briggs Type Indicator

## / روشهای فرافکن

روشهای فرافکن اندازه‌هایی هستند که در آن از یک فرد خواسته می‌شود به یک محرک مبهم یا بی‌ساخت پاسخ دهد. این روشها از آن رو «فرافکن» خوانده می‌شوند که انتظار می‌رود فرد، نیازها، خواسته‌ها، ترسها، اضطرابها و... خود را بر محرک منعکس کند. آزماینده تلاش می‌کند بر مبنای تفسیر و پاسخهای آزمودنی تصویر جامعی از ساخت شخصیتی فرد بنا کند.

روشهای فرافکن را بیشتر روان‌شناسان بالینی برای مطالعه و تشخیص افرادی که مشکلات هیجانی دارند، به کار می‌برند. این روشها اغلب در تحقیقات آموزشی مورد استفاده قرار نمی‌گیرد؛ زیرا برای اجرا و نمره‌گذاری به آموزش تخصصی نیاز دارد و اجرای فردی آن نیز هزینه زیادی به بار می‌آورد. بعلاوه، بسیاری از پژوهشگران احساس می‌کنند که اعتبار آنها به گونه رضایت بخشی تبیین نشده است. دو روش مشهور فرافکن، آزمون رورشاخ<sup>(۱)</sup> و آزمون اندریافت موضوع<sup>(۲)</sup> است. در آزمون رورشاخ از لکه‌های جوهر به عنوان محرک استفاده می‌شود و در آزمون اندریافت موضوع به پاسخگو تصاویری به عنوان محرک نشان داده و از او خواسته می‌شود درباره هر یک از آنها داستانی تعریف کند. هر گونه بحث بیشتر درباره این آزمونها و روشهای آنها را می‌توان در کتاب آناستازی<sup>(۳)</sup> یافت.

مطلب زیر مثالی است از یک بررسی تحقیقاتی که برای اندازه‌گیری متغیرها از پرسشنامه استفاده کرده است.

### رابطه بین مقیاس سنخ‌نمای مایرز - بریگز و ارجحیت شیوه مشاوره دانشجویان سال اول<sup>(۴)</sup>

#### پیش زمینه

تحقیقات قبلی نشان داده است که مشاوره تحصیلی در تحصیلات عالی، یک راهبرد مهم برای تداوم تحصیلی دانشجویان است و بهبود کیفیت مشاوره می‌تواند یادداری دانشجویان را افزایش دهد. همچنین

1- Rorschach

2- Thematic Apperception test (TAT)

۳- Anastasi، همان، فصل ۱۹.

4- Crackett, J.B., and Crawford, R.L (1989). The relationship between Myers-Briggs Type Indicator (MBTI) scale scores and advising style preferences of college Freshmen. *Journal of College Student Development*, 30:154-161

تحقیقات نشان داده است که نیازها و ارجحیت‌های مشاوره‌ای دانشجویان با توجه به نمره‌های شخصیتی خاص هر دانشجو تغییر می‌کند. برای تحقیق بیشتر در مورد رابطه بین ویژگی‌های دانشجویان و ارجحیت‌های شیوه مشاوره، باید وسایلی برای اندازه‌گیری هر دوی این متغیرها پیدا کرد. مرور پیشینه تحقیق نشان داد که سنخ‌نمای مایرز - بریگز (MBTI) یک وسیله اندازه‌گیری است که بخوبی تبیین شده است و به طور گسترده مورد استفاده قرار می‌گیرد.<sup>(۱)</sup> یک وسیله اندازه‌گیری دیگر به نام پرسشنامه مشاوره تحصیلی (AAI) برای سنجش ارجحیت‌های شیوه مشاوره تحصیلی انتخاب شد.<sup>(۲)</sup>

### سؤال تحقیق

(۱) آیا رابطه‌ای بین نمره‌های مقیاس MBTI و ارجحیت‌های شیوه مشاوره وجود دارد؟ (۲) آیا رابطه‌ای بین سن، نژاد، جنس یا سطح زبان انگلیسی دانشجویان به هنگام ورود به دانشگاه، و ارجحیت‌های شیوه مشاوره وجود دارد؟

### نمونه

۲۰۱ نفر از دانشجویان سال اول یک دانشگاه بزرگ در جنوب کشور، که هر دو پرسشنامه را تکمیل کردند، گروه نمونه بررسی را تشکیل می‌دادند. پرسشنامه‌ها در کلاسهای زبان انگلیسی سال اول اجرا شد.

### وسایل اندازه‌گیری

MBTI پرسشنامه‌ای است که مبنای آن نظریه ارجحیت‌ها و

1-Briggs,K.C., and I.B. Myers (1983). Myers-Briggs Type Indicator, From G. Palo Alto. Consulting Psychologists Press.  
2-Winston,R.B., and J.A.Sandor (1984) Academic Advising Inventory. Athens, GA: Student Development Associates.



سرخ‌های کارل یونگ<sup>(۱)</sup> می‌باشد. از افراد خواسته شد ارجحیت خود را بین مبنای احساسات و ارزشهای فردی، و قضاوت-ادراک (JP)<sup>(۲)</sup> نشان دهند، که ارجحیت را در مورد نتیجه‌گیری و سازماندهی زندگی (قضاوت) یا آزادی ذهن نسبت به اطلاعات دریافتی و تغییرات اندازه‌گیری می‌کند.

پرسشنامه مشاوره تحصیلی یک چشم انداز مشاوره تحصیلی مبتنی بر نظریه دارد که بر این فرض استوار است که ارجحیت‌های مشاوره بر روی پیوستاری قرار دارد که دامنه آن از کاملاً تجویزی تا کاملاً توسعه‌ای است. مشاوره تجویزی نوع سنتی مشاوره است که در آن مشاور، مسئولیت اصلی تصمیمها را بر عهده می‌گیرد. مشاوره توسعه‌ای گستره بزرگتری دارد و تجربه زندگی دانشجویی را در بر می‌گیرد. در این بررسی ۱۴ عبارت زوج شده از پرسشنامه انتخاب و از دانشجویان خواسته شد ارجحیت خود را در مورد هر یک از اجزای زوجها نشان دهند. نمره‌ها در سه مقیاس به دست آمد: (۱) خصوصی کردن تعلیم و تربیت، که گستره فعالیت‌های مشاوره‌ای را اندازه می‌گیرد، (۲) تصمیم‌گیری تحصیلی، که نحوه تقسیم مسئولیت تصمیم‌گیری را بین مشاور و دانشجو اندازه‌گیری می‌کند و (۳) انتخاب دروس، که چگونگی انتخاب دروس را می‌سنجد.

## روش

با استفاده از نمره‌های MBTI و AAI و داده‌های مربوط به متغیرهای جمعیت شناختی (نژاد، سن، جنس و سطح زبان انگلیسی دانشجویان به هنگام ورود به دانشگاه) تحلیل‌های همبستگی و رگرسیون چند متغیری انجام شد.

## نتایج

مشاوره توسعه‌ای به طور کامل مورد تأیید همه دانشجویان شرکت کننده در بررسی قرار گرفت. نمره‌های مقیاس های SN و TF مایرز - بریگز بطور مثبت و معنادار با ارجحیت‌های شیوه مشاوره دانشجویان سال اول مرتبط بود. دانشجویانی که بینش شهودی و درک احساسی بیشتری داشتند، ارجحیت بالاتری برای مشاوره توسعه‌ای قایل بودند. سطح زبان انگلیسی دانشجویان به هنگام ورود به دانشگاه تنها متغیر جمعیت‌شناختی معنادار مرتبط با ارجحیت شیوه مشاوره بود.

## // مقیاسهای نگرش

یک مقیاس، مجموعه‌ای از ارزشهای عددی است که به منظور کمی کردن و اندازه‌گیری کیفیتها، به موضوعها، اشیاء یا رفتارها نسبت داده می‌شود. مقیاسها برای اندازه‌گیری نگرشها، ارزشها و علایق مورد استفاده قرار می‌گیرند. مقیاسها از این نظر با آزمونها تفاوت دارند که نتایج آنها، بر خلاف نتایج آزمونها، موفقیت و شکست، یا قوت و ضعف را نشان نمی‌دهند. مقیاسها این مسئله را می‌سنجند که یک فرد تا چه حد ویژگی مورد نظر را داراست. برای مثال، ما می‌توانیم از یک مقیاس برای سنجش نگرش دانشجویان نسبت به مذهب یا هر موضوع دیگر استفاده کنیم.

پژوهشگران اغلب خود مقیاسهای خویش را برای اندازه‌گیری نگرشها می‌سازند. ما در این مطلب تلاش خواهیم کرد برخی از شیوه‌های مورد استفاده در مقیاس‌بندی نگرشها را معرفی کنیم.

نگرش را می‌توان چنین تعریف کرد: یک عاطفه مثبت یا منفی نسبت به یک گروه، نهاد، مفهوم یا شیء اجتماعی خاص. اندازه‌گیری نگرشها بر این فرض مبتنی است که می‌توان مکان افراد را بر روی یک پیوستار مطلوبیت - عدم مطلوبیت نسبت به شیئی نگرشی تعیین کرد. چهار نوع اصلی از مقیاسهای نگرش وجود دارند: (۱) مقیاسهای مجموع درجه‌ها (مقیاسهای لیکرت)، (۲) مقیاسهای فاصله‌های یکسان نما (مقیاسهای ترستون)، (۳) مقیاسهای تراکمی (مقیاسهای گاتمن)، و (۴) مقیاسهای افتراق معنایی.

## / مقیاسهای لیکرت: روش مجموع درجه‌ها

مقیاس لیکرت به عنوان یکی از شیوه‌های سنجش نگرش، بسیار مورد استفاده قرار گرفته و نتایج موفقیت آمیزی نیز به بار آورده است.<sup>(۱)</sup> یک مقیاس لیکرت نگرش نسبت به یک موضوع را از این طریق می‌سنجد که از پاسخگویان می‌خواهد که نشان دهند در مورد هر عبارت از یک مجموعه عبارات، که درباره آن موضوع باشد، کاملاً موافق، موافق، مردد، مخالف، یا کاملاً مخالف هستند. یک مقیاس لیکرت با جمع‌آوری تعدادی عبارت درباره یک موضوع، که تقریباً نیمی از آنان به وضوح نگرش مطلوب و نیم دیگر به وضوح نگرش نامطلوب را بیان می‌کنند، ساخته می‌شود. آنچه که در اینجا اهمیت دارد، آن است که این عبارات یک نمونه معرف از تمام نظرات یا نگرشهای ممکن درباره موضوع نگرش را به وجود آورند. روشی که می‌تواند در این راه به محقق کمک کند، این است که ابتدا درباره همه زیر موضوعهای مرتبط با موضوع نگرش تفکر کند و سپس در مورد هر زیر موضوع، عبارتهای لازم را بنویسد. اگر محقق از اشخاصی که درباره یک موضوع خاص آگاهی لازم و نگرشهای مشخصی دارند، درخواست کند که تعدادی عبارت مثبت و منفی درباره آن موضوع بنویسند، شاید تهیه آن مجموعه متنوع عبارات سهل تر شود. همچنین، سر مقاله‌هایی که نشریات مختلف درباره موضوع مورد نظر می‌نویسند، می‌تواند منبع خوبی از عبارات بالقوه، برای یک مقیاس نگرش باشد.

عبارات، همراه با مقولات پاسخ (به طور نوعی پنج مقوله)، بر روی یک پیوستار توافق - عدم توافق، به آزمودنی ارائه می‌شوند. ترتیب عبارات باید به طور تصادفی تنظیم شود تا مانع از آن شود که پاسخهای آزمودنی، به دلیل ترتیب عبارت، همه به یک جهت متمایل شوند. به آزمودنیها گفته می‌شود که پاسخی را انتخاب کنند که به بهترین نحو نمایانگر واکنش آنها نسبت به هر عبارت باشد: کاملاً موافق (SA)<sup>(۱)</sup>، موافق (A)<sup>(۲)</sup>، مردود (U)<sup>(۳)</sup>، مخالف (D)<sup>(۴)</sup> و کاملاً مخالف (SD)<sup>(۵)</sup>.

برای نمره دادن به مقیاس باید مقوله‌های پاسخ ارزش گذاری شوند. به مقوله‌های پاسخ عبارات مطلوب یا مثبت، به ترتیب از سمت مثبت آن، ارزشهای عددی ۵، ۴، ۳، ۲، ۱ داده

1- Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of Psychology, no. 140. This is Likert's Original Monograph.

3- Agree

5- Disagree

2- Strongly agree

4- Undecided

6- Strongly disagree

می‌شود. برای مثال، پاسخ «کاملاً موافقم» یک عبارت مطلوب، ارزش ۵، پاسخ «موافقم» ارزش ۴ و پاسخ «کاملاً مخالفم» ارزش ۱ خواهد گرفت. برای عبارات نامطلوب یا منفی، ارزش گذاری بر عکس می‌شود؛ زیرا از نظر روان شناختی، مخالفت با یک عبارت نامطلوب، معادل موافقت با یک عبارت مطلوب است. بنابراین «کاملاً موافقم» عبارت نامطلوب، ارزش ۱ و «کاملاً مخالفم»، ارزش ۵ خواهد گرفت (برروی مقیاسی که به پاسخگویان ارائه می‌شود، ارزش‌های وزنی درج نمی‌شوند). تصور کنید مثالهای زیر از مقیاسی که نگرش را نسبت به ریاضیات اندازه می‌گیرد، انتخاب شده‌اند:

SA	A	U	D	SD	
۵	۴	۳	۲	۱	ریاضیات درس دلخواه من در مدرسه است
۱	۲	۳	۴	۵	من ریاضیات را به هیچ شکلی دوست ندارم

شخصی که نگرش مطلوبی نسبت به ریاضیات دارد، با عبارت مثبت، موافقت و با عبارت منفی، مخالفت خواهد کرد. شخصی که ریاضیات را دوست ندارد، با عبارت منفی موافقت و با عبارت مثبت، مخالفت خواهد کرد. اگر آزمودنی برای عبارت اول A و برای عبارت دوم D را علامت بزند، مجموع نمره او برای این دو عبارت  $4 + 4$ ، یعنی ۸، خواهد بود. در یک مقیاس کامل، مجموع ارزشهای همه عباراتی که آزمودنی علامت زده است، نمره کل او را نشان خواهد داد.

این سیستم ارزش گذاری به معنای آن است که کسب نمره بالا در یک مقیاس نگرش (پاسخ «کاملاً موافقم» به عبارات مطلوب، و پاسخ «کاملاً مخالفم» به عبارات نامطلوب)، نگرش مثبت نسبت به موضوع را نشان می‌دهد. بالاترین نمره ممکن برای مقیاس نگرش،  $5 \times N$ ، و پایین ترین نمره ممکن  $1 \times N$  است (N تعداد عبارات است).

**تحلیل عبارت** بعد از آن که مقیاس نگرش با یک گروه مقدماتی از پاسخگویان اجرا شد، برای تشخیص بهترین عبارات، یک تحلیل عبارت انجام می‌شود. تحلیل عبارت، به طور نوعی، سه آماره برای هر عبارت به دست می‌دهد: (۱) شاخص افتراق، (۲) تعداد و یا درصد پاسخگويانی که گزینه‌های مختلف هر عبارت را انتخاب کرده‌اند، و (۳) میانگین و انحراف معیار

عبارت. شاخص افتراق عبارت، همانند نمره کل، میزان تمایزی را که هر عبارت در بین پاسخگویان ایجاد می‌کند، نشان می‌دهد. اگر دارندگان نمره بالا در یک عبارت، در کل مقیاس نیز نمره بالایی به دست آورده باشند و اگر دارندگان نمره پایین در یک عبارت، در کل مقیاس هم نمره پایینی به دست آورده باشند، پس به همان نحو که نمره کل در بین آزمودنیها تمایز ایجاد کرده است، آن عبارت نیز آزمودنیها را از هم متمایز خواهد کرد. شاخص افتراق عبارات، از طریق محاسبه همبستگی نمره هر عبارت با نمره کل مقیاس به دست می‌آید و این کار معمولاً به وسیله کامپیوتر صورت می‌گیرد. هر عبارت باید حداقل همبستگی  $0/25$  با نمره کل داشته باشد. عباراتی که با نمره کل، همبستگی پایین یا منفی دارند، باید حذف شوند؛ زیرا آنها همان چیزی را که نمره کل اندازه می‌گیرد، اندازه نمی‌گیرند، از اینرو در اندازه‌گیری نگرش دخالت ندارند. محقق باید آن عباراتی را که متمایز کننده نیستند، بررسی کند. این عبارات ممکن است مبهم یا دو پهلو (حاوی دو عقیده یا نظر در یک جمله) باشند یا ممکن است جملات واقعی باشند که احساسی را در مورد یک موضوع بیان نمی‌کنند. تجدید نظر در این عبارات، شاید آنها را قابل استفاده سازد.

آماره‌های دیگری که از تحلیل عبارات به دست می‌آیند (۲ و ۳ یاد شده)، میزان استفاده پاسخگویان را از گزینه‌های مختلف نشان می‌دهند. عباراتی که در آنها پراکندگی پاسخگویان بر روی مقوله‌های پاسخ زیاد است، بر عباراتی که در آنها پاسخها بر روی یک یا دو مقوله تجمع یافته‌اند، ارجحیت دارند. محقق پس از آن که بر مبنای نتایج تحلیل، عبارات خوب را انتخاب کرد، باید مقیاس تجدید نظر شده را با گروه دیگری از آزمودنیها اجرا و باز هم عبارات را از نظر ایجاد افتراق بررسی کند.

اعتبار یافتن معیارهایی برای تعیین اعتبار مقیاسهای نگرش، کاری دشوارتر است. برخی از محققان، از مشاهده رفتار واقعی به عنوان معیار نگرشی که اندازه‌گیری شده است، استفاده کرده‌اند، اما این شیوه همیشه قابل استفاده نیست؛ زیرا تعیین این که چه رفتاری بهترین معیار نگرش خواهد بود، دشوار است، و بعلاوه، حصول اطمینان از این که رفتار بدرستی اندازه‌گیری شده است نیز کار آسانی نیست.

یکی از آسانترین راههای گردآوری شواهدی در زمینه اعتبار مقیاسهای نگرش، تعیین میزان توانایی آنها در ایجاد تمایز بین دو گروه از افرادی است که مشخص شده است نگرشهای متفاوتی

دارند. اگر مقیاسی که نگرش را نسبت به مذهب اندازه‌گیری می‌کند، بین اعضای فعال کلیسا و افرادی که به کلیسا نمی‌روند و هیچ‌گونه وابستگی مذهبی ندارند، تمایز قایل شود، می‌توان گفت که آن مقیاس دارای اعتبار است. یک مقیاس معتبر که نگرش را نسبت به سقط جنین اندازه‌گیری می‌کند، باید اعضای گروههای ضد آزادی سقط جنین را از گروههای طرفدار آزادی سقط جنین متمایز کند، یعنی انتظار خواهد رفت که میانگین نمرات اعضای این دو گروه در مقیاس نگرش مذکور، از نظر آماری با هم تفاوت معنادار داشته باشند. روش دیگر تعیین اعتبار، محاسبه همبستگی بین نمره‌های دو مقیاس نگرش است که هر دو یک سازه را اندازه می‌گیرند و مشخص شده است که یکی از آنها دارای اعتبار است.

### / مقیاسهای ترستون: روش فاصله‌های یکسان نما

ترستون روشی را برای تخصیص ارزشهای مقیاسی ویژه به عبارات نگرشی ابداع کرد.<sup>(۱)</sup> در حالی که در شیوه لیکرت از پاسخگویان خواسته می‌شود تا موافقت یا مخالفت خود را با هر یک از عبارات ابراز کنند، مقیاس ترستون مجموعه‌ای از عبارات را که درباره یک موضوع باشند و دامنه آنها از خیلی مطلوب تا خنثی و از خنثی تا خیلی نامطلوب سیر کند، به پاسخگویان ارائه می‌کند و از آنها می‌خواهد که عباراتی را که بیشتر از بقیه با نگرش خود آنها مطابقت دارند، انتخاب کنند. تهیه یک مقیاس ترستون، مراحل زیر را شامل می‌شود:

- ۱- تعداد زیادی عبارت (۵۰ تا ۱۰۰ عبارت) در مورد موضوع نگرش جمع‌آوری کنید.
- ۲- عبارتها را در اختیار تعدادی داور بگذارید. معمولاً از پنجاه تا صد داور برای ارزیابی عبارتها استفاده می‌شود. داورها مستقلاً کار می‌کنند و از آنها خواسته می‌شود همه عبارات را بر مبنای درجه «مطلوبیت» هر عبارت در یازده طبقه دسته بندی کنند. هر داور عباراتی را که به نظرش مطلوبترین محتوا رانسبت به موضوع نگرش دارند، در مجموعه اول قرار می‌دهد. در مجموعه دوم عبارات کمتر «مطلوب» را جای می‌دهد و الی آخر. مجموعه ششم حاوی عبارات خنثی است و مجموعه یازدهم «نامطلوبترین» عبارات را در خود دارد. نکته مهم در این است که این طبقه بندی ربطی به نگرش داورها نسبت به موضوع ندارد، بلکه تنها معرف قضاوت آنها در مورد «مطلوبیت» عبارت‌ها است.

۳- ارزش مقیاسی هر عبارت را از طریق تعیین میانه و زنها یا مکانهای مقیاسی که داوران به هر عبارت تخصیص داده‌اند، پیدا کنید. عبارت‌هایی که تغییر پذیری بسیار زیاد نشان دهند، حذف می‌شوند.

۴- تعداد ۲۰ یا ۳۰ عبارت را که به طور یکنواخت از یک انتها تا انتهای دیگر مقیاس توزیع شده‌اند، انتخاب کنید. این عبارت‌ها مقیاس نگرش را تشکیل می‌دهند. عبارات زیر با ارزش مقیاسی آنها، از مقیاس نگرش ترستون نسبت به کلیسا گرفته شده‌اند.

عبارت	ارزش مقیاس
من بر این باورم که کلیسا در آمریکای امروز بزرگترین نهاد است.	۰/۲
من بر این باورم که عضویت در کلیسا برای گذران زندگی به بهترین نحو، امری اساسی است.	۱/۵
من خدمات کلیسا را هم آرام بخش و هم الهام بخش یافته‌م.	۲/۳
من از کلیسای خود لذت می‌برم، زیرا فضایی دوستانه بر آن حاکم است.	۳/۳
من به تعلیمات کلیسا معتقدم، اما به شرط عقل.	۴/۵
بعضی اوقات احساس می‌کنم کلیسا و مذهب ضروری هستند و گاهی در مورد آن شک می‌کنم.	۵/۶
من به صداقت و نیکی بدون تشریفات کلیسایی اعتقاد دارم.	۶/۷
من بر این باورم که هر چه تعلیم و تربیت پیشرفت می‌کند، مذهب زمینه خود را از دست می‌دهد.	۷/۴
من فکر می‌کنم که تعلیمات کلیسا بر رویهم بسیار سطحی‌تر از آن است که اهمیت اجتماعی داشته باشد.	۸/۳
من فکر می‌کنم کلیسا مانعی بر سر راه مذهب است، برای این که هنوز به جادو، خرافات و افسانه تکیه می‌کند.	۹/۶
من فکر می‌کنم کلیسا انگلی برای جامعه است.	۱۱/۰

برای اجرای یک مقیاس نوع ترستون، از آزمودنی خواسته می‌شود عباراتی را که معرف موضع او هستند، انتخاب کند، یا سه عبارت را که بیشتر از سایر عبارات به موضع او نزدیک‌اند،

برگزینند. البته ارزشهای مقیاسی در مقیاس نگرش قید نمی‌شوند و عبارات نه بر اساس ارزش مقیاسی، بلکه به طور تصادفی در مقیاس نگاشته می‌شوند. نمره هر آزمودنی میانگین ارزشهای عباراتی که او انتخاب کرده است، خواهد بود.

### / مقیاسهای گاتمن: روش تراکمی

منتقدان مقیاسهای نگرش لیکرت و ترستون خاطر نشان کرده‌اند که این مقیاسها حاوی عبارات ناهمگن درباره ابعاد مختلف یک موضوع نگرش است. برای مثال، در مقیاس ترستون که نگرش نسبت به جنگ را می‌سنجد، برای جدا سازی عبارات اخلاقی از عباراتی که درباره نتایج اقتصادی جنگ است یا عباراتی که دیگر جنبه‌های ممکن نگرش نسبت به جنگ را منعکس می‌کنند، هیچ تلاشی صورت نگرفته است. پیامد ترکیب ابعاد مختلف در یک مقیاس، آن است که ارائه تفسیری روشن از نمره‌های حاصله را دشوار می‌سازد. گاتمن روشی را در جهت تلاش برای غلبه بر این مشکل ابداع کرد. روش او، که به عنوان مقیاس تک بعدی مشخص شده است، می‌خواهد تعیین کند که نگرش مورد مطالعه واقعاً دارای یک بعد است. تنها نگرشی تک بعدی محسوب می‌شود که یک مقیاس تراکمی به بار آورد. مقیاسی که عبارات آن به گونه‌ای با هم مرتبط اند که اگر آزمودنی با عبارت ۲ موافق است، همچنین با عبارت ۱ نیز موافقت می‌کند. کسی که با عبارات ۳ موافق است، همچنین با عبارات ۱ و ۲ نیز موافقت می‌کند و الی آخر. بنابراین افرادی که یک عبارت خاص را در این نوع مقیاس تایید می‌کنند، از افرادی که آن عبارت را رد می‌کنند، در کل مقیاس، نمره بالاتری به دست خواهند آورد. برای مثال عبارات زیر را که از پاسخگویان خواسته می‌شود با آنها موافقت یا مخالفت کنند، در نظر بگیرید:

۱- PTA ارزش آن را دارد که برای آن وقت صرف شود.

۲- PTA در امر بهبود وضعیت مدارس، دارای نفوذ بسیار است.

۳- PTA در ایالات متحده بهترین سازمان در زمینه بهبود وضعیت مدارس است.

اگر این یک مقیاس تراکمی باشد، باید امکان مرتب کردن همه پاسخهای پاسخگویان به گونه‌ای که در جدول ۱-۷ نشان داده شده است، وجود داشته باشد. بنابراین اگر نمره یک فرد را بدانیم، این امکان را نیز داریم که دقیقاً بگوییم او کدام عبارتها را تأیید کرده است. برای مثال، همه پاسخگویانی که نمره ۲ به دست آورده‌اند، بر این باورند که PTA ارزش صرف وقت را دارد و



دارای نفوذ زیادی در امر بهبود وضعیت مدارس است، اما معتقد نیستند که PTA در ایالات متحده مهمترین سازمان در زمینه بهبود وضعیت مدارس است. آزمودنیها را می توان بر مبنای پاسخهایشان به عبارات مقیاس رتبه بندی کرد.

به هنگام ساختن یک مقیاس تراکمی، ابتدا تعیین این که عبارات آن یک مقیاس تک بعدی را تشکیل می دهند یا خیر، امری اساسی است.<sup>(۱)</sup> برای انجام این کار، تجدیدپذیری پاسخها را تحلیل می کنند - یعنی نسبت پاسخهایی را که واقعاً در الگویی شبیه جدول ۷-۱ جای می گیرند، تعیین می کنند. بر مبنای نمره کل، الگوی پاسخها به عبارات خاص، پیش بینی می شود. سپس الگوی واقعی پاسخها مطالعه می شود و بر اساس میزانی که پاسخها از روی نمره کل تجدیدپذیرند، اندازه ای ساخته می شود. یک روش آن است که تعداد کل خطاها بر تعداد کل پاسخها تقسیم و از یک کم شود. گاتمن ۰/۹ را به عنوان حداقل ضریب تجدیدپذیری، که برای شناسایی مجموعه ای از عبارات به عنوان اجزای مشکله یک مقیاس تک بعدی لازم است، پیشنهاد می کند.

جدول ۱ - ۷ مثال روش تراکمی\*

نمره	موافقت با عبارت			مخالفت با عبارت		
	۳	۲	۱	۳	۲	۱
۳	x	x	x	o	o	o
۲	o	x	x	x	o	o
۱	o	o	x	x	x	o
o				x	x	x

\* پاسخگو برای هر پاسخ موافق یک نمره می گیرد.

بسیاری بر این باور هستند که مقیاس گاتمن بیش از آن که اهمیت کاربردی داشته باشد؛ اهمیت نظریه ای دارد؛ زیرا جمع آوری عبارتهایی که معیار تجدیدپذیری را برآورده سازند، دشوار

۱- برای بحث مفصل تر مراجعه کنید به :

A. L. Edwards (1952), *Techniques of Attitude Scale Construction*, New York: Irvington.

است. همچنین این انتقاد بر شیوه گاتمن وارد است که راه یا راههایی را برای تهیه یا انتخاب عبارتها پیشنهاد نمی‌کند. تنها پس از این که عبارتها انتخاب شدند، می‌توان درباره تجدیدپذیری آنها قضاوت کرد.

### مقیاسهای افتراق معنایی

رویکرد دیگر برای اندازه‌گیری نگرش نسبت به اشیاء، موضوعها و رویدادها روش افتراق معنایی است. این روش را آزگود، سوسی، و تانن‌باوم<sup>(۱)</sup> ابداع و از آن استفاده کردند. روش افتراق معنایی بر این مفروضه بنا شده است که اشیاء برای افراد دو معنای مختلف دارند، معنای صریح و معنای ضمنی، که می‌توان آنها را به طور مستقل ارزیابی کرد. معنای صریح یک کلمه به معنایی که در فرهنگ لغات برای آن کلمه آمده است، اطلاق می‌شود، در حالی که معنای ضمنی تداعی‌ها و اشاراتی است که یک کلمه به ذهن متبادر می‌کند. افراد بسادگی می‌توانند معنای صریح یک شیء را بیان کنند، اما در بیان معنای ضمنی آن با مشکل مواجه می‌شوند. اندازه‌گیری معنای ضمنی اشیاء به طور غیر مستقیم، با استفاده از تعدادی صفات دو قطبی و درخواست از افراد برای رتبه بندی اشیاء در قبال این صفات، هم امکان پذیر و هم مفید است. بنابراین، معنای یک شیء برای فرد، الگوی رتبه بندی آن شیء بر روی مقیاسهای صفات دو قطبی خواهد بود.

مقیاسهای نگرش با انتخاب زوجهای صفات دو قطبی که معرف بعد ارزشی باشند، ساخته می‌شوند. زوجهای صفات به همراه یک پیوستار پاسخدهی هفت تایی ارائه می‌شوند و از پاسخگو خواسته می‌شود که با گذاردن × بر روی یکی از فضاها، هفت تایی، نشان دهد که صفت مورد نظر تا چه حد شیء نگرشی را توصیف می‌کند. برای مثال، فرض کنید فردی می‌خواهد نگرشهای دانش‌آموزان دبیرستانی را نسبت به مدرسه اندازه بگیرد.

آزگود، سوسی، و تانن‌باوم از طریق مطالعات تحلیل عاملی، به سه دسته از صفات دست پیدا کرده‌اند: ارزشی<sup>(۲)</sup> (شامل صفاتی نظیر خوب و بد، ارزشمند و بی ارزش)، توانایی<sup>(۳)</sup> (شامل صفاتی نظیر قوی و ضعیف یا سنگین و سبک)، فعالیت<sup>(۴)</sup> (شامل صفاتی نظیر فعال و وارفته یا سریع و کند)، در میان این سه دسته، به نظر می‌رسد که بعد ارزشی برای اندازه‌گیری نگرش، بیشترین اهمیت را دارا باشد.

1- Osgood, G.E., G.J. Suci, and P.H. Tannenbaum (1967). The Measurement of Meaning. Urbana: University of Illinois Press.

2- Evaluative

3- Potency

4- Activity

مدرسه	
بد	خوب
فعال	وارفته
تیزهوش	خنک
خوشایند	ناخوشایند
بی‌ارزش	ارزشمند
سخت	نرم
سنگین	سبک
ضعیف	قوی
سریع	کند

به مقیاس بالا توجه کنید، که در آن، برای این که گرایش پاسخگو به یک سمت مقیاس به حداقل برسد، زوجهای صفات در هر دو جهت مثبت و منفی گنجانده شده‌اند. ممکن است فردی انتهای سمت راست مقیاس را انتخاب کند و برای همه موارد، آنجا را علامت بزند. به هر ترتیب، اگر جهت مقیاس به طور تصادفی تغییر داده شود، به گونه‌ای که سمت راست مقیاس، همیشه پاسخ مثبت را در بر نداشته باشد، پاسخگو و اداری می‌شود که هر مورد را ابتدا بخواند و پاسخ خود را برحسب محتوای آن ارائه کند. به هر روی، در نمره دهی مقیاس افتراق معنایی، به پاسخها از ۱ تا ۷ نمره داده می‌شود؛ به طوری که ۷، مثبت‌ترین پاسخ را بنمایاند. بنابراین، مورد اول مقیاس بالا، یعنی صفت بد، اولین گزینه از سمت راست، نمره ۱، دومین گزینه از سمت راست، نمره ۲، و اولین گزینه از سمت چپ، نمره ۷ خواهد گرفت. برای مورد دوم، یعنی صفت فعال، ترتیب نمره دهی برعکس می‌شود، به طوری که گزینه اول سمت راست، نمره ۷، و گزینه اول سمت چپ، نمره ۱ می‌گیرد. سپس، تمام نمره‌ها با هم جمع می‌شوند و یک نمره میانگین محاسبه می‌شود.

### // مقیاسهای درجه‌بندی

یکی از رایجترین ابزارهای اندازه‌گیری، مقیاس درجه‌بندی است. مقیاسهای درجه‌بندی عبارتند از سنجش رفتار یا عملکرد یک فرد به وسیله فردی دیگر. به طور نوعی، از فرد درجه‌بندی کننده خواسته می‌شود که جایگاه فرد درجه‌بندی شونده را بر روی یک پیوستار یا در

یک طبقه‌بندی که ویژگیهای رفتاری او را توصیف می‌کند، مشخص کند. به این درجه‌بندیها ارزش عددی تخصیص می‌یابد. چنین فرض می‌شود که درجه‌بندی کنندگان با رفتار نوعی افراد آشنا هستند. درجه‌بندیها در تحقیق بر روی رشد کودکان و سایر جنبه‌های رفتاری مورد استفاده قرار گرفته‌اند.

شکل ۲-۷ مثالی از یک مقیاس نگراره‌ای		
پایین	متوسط	بالا
		وضع ظاهری
		پذیرش اجتماعی
		مهارتهای بیانی
		همیشه ایده‌های خلاق دارد
		ایده‌های خلاق بسیاری دارد
		گاهی ایده‌های خلاق دارد
		به ندرت ایده‌های خلاق دارد

انواع مختلفی از مقیاسهای درجه‌بندی وجود دارند. یکی از مقیاسهایی که بیشترین مورد استفاده را دارد، مقیاس نگراره‌ای است که در آن درجه‌بندی کننده فقط نقطه مناسب را بر روی خطی که دو حد نهایی رفتار مورد نظر را نشان می‌دهد، علامت می‌زند. شکل ۲-۷ یک مثال نوعی از مقیاس نگراره‌ای است. فردی که درجه‌بندی می‌کند، می‌تواند هر نقطه‌ای را روی خط پیوستار علامت بزند. در برخی از مقیاسهای نگراره‌ای، سازنده آزمون، به نقاط توصیفی، ارزشهای عددی تخصیص می‌دهد. به چنین مقیاسهایی، مقیاسهای درجه‌بندی عددی گفته می‌شود. مورد مهارتهای بیانی در شکل ۲-۷، در یک مقیاس عددی به صورت زیر در می‌آید.

۱	۲	۳	۴	۵	۶	۷
یکی از			بک سخنان			یکی از
ضعیف‌ترین			متوسط			متن
سخنرانان						سخنرانان

نوع دوم مقیاس درجه‌بندی، مقیاس طبقه‌ای است که شامل طبقاتی است که بر مبنای نظم خاصی مرتب شده‌اند. غالباً پنج الی هفت طبقه مورد استفاده قرار می‌گیرند. فرد درجه‌بندی کننده، یکی از طبقات را که به بهترین نحو رفتار فرد درجه‌بندی شونده را توصیف می‌کند، انتخاب می‌کند. فرض کنید تواناییهای یک دانش‌آموز درجه‌بندی شده است و یکی از ویژگیهای درجه‌بندی شده، خلاقیت است. مورد خلاقیت به صورت زیر در می‌آید.

این فرد تا چه حد خلاق است؟ (یکی را علامت بزنید)

فوق‌العاده خلاق است

خیلی خلاق است

خلاق است

خلاق نیست

ابداً خلاق نیست

درجه‌بندی کنندگان، به هنگام استفاده از مقیاسهای نگراره‌ای و طبقه‌ای، بدون آن که فرد درجه‌بندی شونده را مستقیماً با افراد یا گروههای دیگر مقایسه کنند، قضاوت خود را انجام می‌دهند. از سوی دیگر، در مقیاسهای درجه‌بندی مقایسه‌ای، به درجه‌بندی‌کنندگان آموزش داده می‌شود که با عطف به جایگاه کسانی که قابل مقایسه هستند، قضاوت خود را انجام دهند. در مقیاس درجه‌بندی، جایگاه برحسب یک جمعیت معین با مشخصه‌های شناخته شده، تعریف می‌شود. در شکل ۷-۳ یک مقیاس درجه‌بندی مقایسه‌ای نشان داده شده است. چنین مقایسه‌ای را می‌توان در انتخاب داوطلبان ورود به دانشکده مورد استفاده قرار داد. از درجه‌بندی کنندگان خواسته می‌شود در مورد توانایی داوطلبان برای تحصیل در دانشکده، در مقایسه با همه دانشجویانی که از قبل می‌شناسند، قضاوت کنند.

همه روشهای درجه‌بندی در معرض خطاهای قابل توجهی هستند. از رایجترین خطاهای سیستماتیک در درجه‌بندی افراد «اثر هاله‌ای»<sup>(۱)</sup> است. این خطا هنگامی اتفاق می‌افتد که فرد درجه‌بندی کننده اجازه دهد تا برداشت تعمیم یافته او از آزمودنی بر درجه‌بندی جنبه‌های بسیار خاص رفتار آزمودنی تأثیر بگذارد. این برداشت کلی به تمام مقیاس انتقال می‌یابد. برای مثال، ممکن است یک آموزگار، دانش‌آموز درسخوانی را از نظر هوش، محبوبیت، صداقت، امانتداری و همه جنبه‌های دیگر شخصیتی در حد عالی درجه‌بندی کند.

شکل ۳-۷ مثالی از یک مقیاس درجه بندی مقایسه ای

به منظور غیر عادی پایین	صبیح روزا کبر دانش آموزان	حد متوسط در میان دانش آموزان	بهروز اکبر دانش آموزان	واقار برار دیگران	غیر قابل مقایست	حیطه شایستگی (باید درجه بندی شود)
						۱- آیا این فرد شراهد روشن و شایسته ای از هدفهای حرفه ای نشان می دهد؟
						۲- آیا این فرد با روش سازنده ای به جنگ مشکلات می رود؟
						۳- آیا او انتقادهای مثبت را می پذیرد و از آن به طور سازنده ای استفاده می کند؟

یک نوع خطای دیگر، «خطای سخاوتمندی»<sup>(۱)</sup> است. این خطا، گرایش به در نظر گرفتن سوی مثبت هر شک و تردیدی، در مورد آزمودنی است، یعنی هنگامی که درجه بندی کننده مطمئن نیست، احتمالاً در مورد فردی که درجه بندی می کند، با دید موافق صحبت خواهد کرد. در مقابل آن، «خطای سخت گیری»<sup>(۲)</sup> است که گرایش به درجه بندی کردن بسیار پایین همه افراد، در همه ویژگیها دارد. منبع خطای دیگر، «خطای گرایش مرکزی»<sup>(۳)</sup> است که به اجتناب از دو حد نهایی مقیاس و درجه بندی کردن همه افراد در حد میانی مقیاس اطلاق می شود.

یک راه برای کاهش چنین خطاهایی آن است که افراد درجه بندی کننده را قبل از انجام کار به طور کامل آموزش دهیم. آنها باید از امکان ارتکاب این نوع خطاها آگاه باشند. این یک امر اساسی است که افراد درجه بندی کننده، قبل از انجام درجه بندی، وقت کافی برای مشاهده رفتار فردی که قرار است درجه بندی شود، داشته باشند. یک راه دیگر برای به حداقل رساندن خطا آن است که مطمئن شویم رفتاری که قرار است درجه بندی شود و نقاطی که روی مقیاس هستند، به طور روشن، تعریف شده اند. نقاط روی مقیاس باید برحسب رفتارهای آشکاری که قابل مشاهده هستند، توصیف شوند، نه برحسب رفتارهایی که نیازمند استنباط از سوی فرد درجه بندی کننده

1- Generosity Error

2- Error of Severity

3- Error of Central tendency

می‌باشند. به خواننده توصیه می‌کنیم برای یک مبحث عالی در مورد شیوه‌های رویارویی با خطاهای فرد درجه‌بندی کننده به کتاب گیلفورد<sup>(۱)</sup> مراجعه کند.

معمولاً اگر افراد متعددی به طور مستقل یک فرد را درجه‌بندی کنند، پایایی درجه‌بندی افزایش می‌یابد. این درجه‌بندیهای مستقل یک کاسه می‌شوند، یا میانگین آنها گرفته می‌شود تا یک نمره مستقل به دست آید.

### // روشهای گروه سنجی

روشهای گروه سنجی برای مطالعه سازمان گروههای اجتماعی مورد استفاده قرار می‌گیرند. روش اصلی، اگر چه ممکن است به گونه‌های مختلف تغییر یابد، چنین است که از اعضای یک گروه خاص خواسته می‌شود که انتخاب اول، دوم و انتخابهای بعدی خود را از میان سایر افراد گروه، برای همراهی در یک فعالیت خاص و بر مبنای یک معیار ویژه نشان دهند. برای مثال، از افراد یک گروه پرسیده می‌شود: «با چه کسی می‌خواهید درس بخوانید، یا در کنارش بنشینید، یا با او ناهار بخورید، یا بعد از مدرسه با او بازی کنید؟» اساساً این روش، مطالعه انتخابهایی است که هر فرد در یک گروه انجام می‌دهد. انتخابهایی که به دست می‌آید، به صورت یک نگاره گروهی ترسیم می‌شود که الگوی روابط میان فردی را در یک گروه نشان می‌دهد. شکل ۴-۷ یک نگاره گروهی را نشان می‌دهد، که در آن ج بیش از سایر افراد گروه انتخاب شده است. می‌توان از او به عنوان «ستاره» نام برد. توجه کنید که و، ز و ه یکدیگر را انتخاب کرده‌اند. این نوع انتخابها نشانه‌ای از وجود یک دسته است، یعنی سه فرد یا بیشتر که دو بدو یکدیگر را انتخاب می‌کنند. ب اصلاً انتخاب نشده است، اگر چه او خود در گروه انتخابهایی انجام داده است، او یک «منزوی» است. انتخابهایی که بر روی یک نگاره گروهی آشکار می‌شود، می‌تواند به صورت کمی درآید و برای مقاصد تحقیقاتی مورد استفاده قرار گیرد.

روشهای گروه سنجی برای تحقیق در روان‌شناسی اجتماعی و همچنین در تحقیقات آموزشی، که در آن امکان مطالعه رابطه وضع فرد در گروه با سایر متغیرها نظیر توانایی ذهنی، پیشرفت، رجحانهای آموزگاران در مورد کودکان وجود دارد، به طور وسیع مورد استفاده قرار گرفته است.<sup>(۲)</sup>

1- Gullford, J.P. (1954). *Psychometric Methods* (ch. 11). New York: McGraw - Hill.

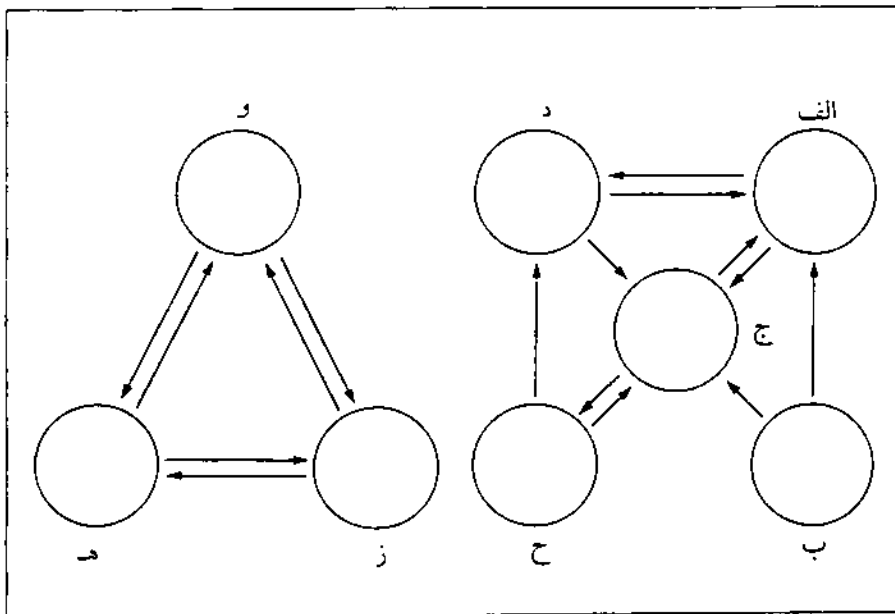
2- G. Lindzey (1968). *Handbook of Social Psychology* (Vol. 1). Reading, MA: Addison - Wesley.

## // مشاهده مستقیم

در بسیاری از حالات، مشاهده مستقیم و نظام‌مند رفتار، مطلوبترین روش اندازه‌گیری است. یک پژوهشگر رفتار مورد نظر را معین می‌کند و سپس روش نظام‌مندی برای تشخیص، طبقه‌بندی و ثبت آن رفتار، هم در محیط طبیعی و هم در محیط مصنوعی، ابداع می‌کند. مشاهده نظام‌مند به طور گسترده برای تحقیق بر روی نوزادان و کودکان پیش دبستانی مورد استفاده قرار گرفته است.

تحقیق اوربان<sup>(۱)</sup> با عنوان «تغییرات رفتاری حاصله از مطالعه بیماریهای واگیر» نمونه ممتازی از کاربرد مشاهده مستقیم در محیط طبیعی است. مشاهده‌گران تعداد موارد رفتارهای نامطلوب، نظیر گذاشتن انگشت یا اشیای دیگر را در دهان و رفتارهای مطلوب، نظیر استفاده از دستمال به

شکل ۴-۷ یک نگاره گروهی



شبه شبکه



هنگام سرفه یا عطسه را ثبت کردند. در پی آن، یک گروه آزمایشی انتخاب شد و یک دوره شش هفته‌ای را درباره بیماریهای واگیر، که هم برای تغییر رفتار آسکار افراد گروه و هم برای آگاه سازی آنها از این بیماریها طراحی شده بود، گذراند. در پایان دوره، رفتارهای نامطلوب دوباره ثبت شدند. نتایج حاصله مشخص کرد که در گروه آزمایشی رفتارهای نامطلوب به طور چشمگیری کاهش و رفتارهای مطلوب به طرز چشمگیری افزایش یافته‌اند، در حالی که در گروه گواه، که آن دوره را نگذرانند، تغییر کمی صورت گرفته است. مشاهده‌هایی که در طول ۱۲ هفته انجام شد، نشان داد که تفاوت‌های بین دو گروه گواه و آزمایشی ماندگار است.

قبل از استفاده از مشاهده مستقیم پنج مرحله مقدماتی مهم باید طی شود.

۱- آن جنبه از رفتار که قرار است مشاهده شود، باید انتخاب شود. از آنجا که یک مشاهده‌گر نمی‌تواند هر چیزی را که اتفاق می‌افتد، مورد توجه قرار دهد، محقق از قبل باید تصمیم بگیرد که کدام رفتارها ثبت و کدام رفتارها حذف شوند.

۲- رفتارهایی که در یک طبقه‌بندی قرار می‌گیرند، باید بروشنی تعریف شوند. مشاهده‌گران باید از پیش بدانند که برای مثال، چه اعمالی به عنوان رفتار جمعی یا رفتار خودخواهانه طبقه‌بندی خواهد شد.

۳- کسانی که کار مشاهده را انجام می‌دهند، باید آموزش داده شوند. مشاهده‌گران باید آموزش ببینند و تمرین کنند تا یکدستی تفسیرها و کاربست همسان طبقه‌بندیهای مشاهده حاصل شود.

۴- برای کمی کردن مشاهده‌ها باید سیستمی تهیه شود. محقق باید یک روش استاندارد برای شمارش رفتارهای مشاهده شده مشخص کند. برای مثال، محقق باید از قبل تعیین کند که آیا یک کنش و واکنش را باید یک رفتار منفرد شمارش کرد یا دو بخش از یک رفتار. یک شیوه پیشنهادی آن است که مدت زمان مشاهده به قطعات زمانی کوتاه تقسیم شود و برای هر دوره زمانی - به عنوان مثال، ۳۰ ثانیه - بروز یا عدم بروز آن رفتار ثبت شود.

۵- برای ثبت رفتار باید شیوه‌های مبسوطی تهیه شود. معمولاً حافظه اکثر مشاهده‌گران برای یک تحقیق هدفمند، به اندازه کافی قابل اعتماد نیست. بهترین راه عبارت است از تهیه یک طرح رمزگردانی که مشاهده‌گران را قادر سازد تا مشاهدات خود را، نه به صورت نوشتاری - که هم آهده‌گو را می‌گیرد و هم توجه او را منحرف می‌سازد - بلکه با یک حرف مفرد یا عدد،

کاربرد دیگر مشاهده مستقیم، مطالعه رفتار کلاسی است. فلاندرز تعامل کلامی را در کلاس درس مطالعه و یک برنامه رمزگردانی برای ثبت مشاهده‌ها تهیه کرد که آن را می‌توان در شکل ۵-۷ دید. سیستم فلاندرز طبقه‌بندیهای جامع و نامتناخلی ارائه می‌کند، که هر یک را می‌توان با یک عدد مفرد ثبت کرد. مشاهده‌گران آموزش دیده قادرند در هر ۳ ثانیه یک عدد را ثبت کنند. زنجیره اعداد حاصله را می‌توان به آسانی تحلیل کرد تا نه تنها پیشینه نسبتهای رفتار کلامی هر طبقه‌بندی، بلکه همچنین تصویری از تقدم و تأخر آن رفتارها نیز به دست آید.

فلاندرز گفته‌های معلم را به دو بخش گفتار پاسخی، که آن را نفوذ غیر مستقیم نامید، و گفتار آغازی، که آن را نفوذ مستقیم نامید، تقسیم کرد. فلاندرز در یک بررسی، با استفاده از طبقه‌بندی مشاهده خود، تأثیر گفته‌های نوع نفوذ غیرمستقیم معلم را در مقایسه با گفته‌های از نوع نفوذ مستقیم او، مورد مطالعه قرار داد.<sup>(۱)</sup> کلاسهای علوم اجتماعی و ریاضیات مورد مشاهده قرار گرفتند و کلاسهایی که معلمان آنها از روش غیرمستقیم استفاده می‌کردند، با کلاسهایی که معلمان آنها از روش مستقیم استفاده می‌کردند، هم از نظر پیشرفت تحصیلی و هم از نظر نگرش نسبت به معلم با هم مقایسه شدند. فلاندرز دریافت که کلاسهای معلمانی که از روش غیرمستقیم استفاده می‌کنند، هم در زمینه نگرش و هم در زمینه پیشرفت تحصیلی، عملکرد بهتری دارند. نتیجه این بررسی آن بود که اگر به معلمان آموزش داده شود که از مقوله غیرمستقیم رفتار کلامی بیشتر استفاده کنند، کار آنها بهبود می‌یابد.

در یک مطالعه کلاسی دیگر، رالینز و دیگران مشاهده نظام‌مند را قبل و بعد از یک دوره آموزشی که طی آن معلمان شیوه‌های تقویت مثبت را فرا گرفتند و به استفاده از این شیوه‌ها تشویق شدند، به کار بردند.<sup>(۲)</sup> مشاهده برای این انجام شد که میزان استفاده معلمان از شیوه تقویت مثبت و اثر تقویت مثبت بر رفتار دانش‌آموزان، تعیین شود. مشاهده در دوره‌های ۵ دقیقه‌ای انجام شد. در اولین دوره ۵ دقیقه‌ای رفتارهای تقویتی مثبت و منفی معلمان شمارش شد، در دومین دوره ۵ دقیقه‌ای، رفتارهای اختلال برانگیز دانش‌آموزان شمارش شد، در سومین دوره ۵ دقیقه‌ای، رفتارهای توجه‌آمیز دانش‌آموزان مورد مشاهده قرار گرفت. رالینز دریافت

1. Rollins, N.A. (1965). *Teacher Influence, Pupil Attitudes, and Achievement*. U.S. Office of Education Cooperative Research Monograph. Washington: U.S. Government Printing Office.

2. Rollins, H.A., B.R. McCandless, and M. Thompson (1974). Project Success Environment: An Extended Application of Contingency Management in Inner-city Schools. *Journal of Educational Psychology*, 66, pp. 167 - 178.

معلمانی که آموزش دیده بودند، تقریباً چهار برابر معلمان گروه کنترل، نسبت به هر دانش‌آموز، از رفتار تقویتی استفاده کردند و خیلی کمتر مجبور به استفاده از تنبیه شدند. مشاهداتی که بر روی دانش‌آموزان صورت گرفت، کاهش چشمگیری در رفتارهای اخلال برانگیز و افزایش در رفتار توجه‌آمیز را نشان داد. بعلاوه، دانش‌آموزانی که در گروه آزمایشی قرار داشتند، در یک دوره ۸ ماهه نمره متوسط بالاتری در آزمون خواندن به دست آوردند. در این مطالعه، از مشاهده نظام‌مند به طور خیلی مؤثر برای ثبت تغییرات رفتار کلاسی، که با تغییرات چشمگیر در پیشرفت تحصیلی دانش‌آموزان همبسته بود، بهره گرفته شد.

### / ارزشیابی مشاهده مستقیم

در مشاهده مستقیم ما قادر هستیم که پیشینه‌ای از رفتار واقعی فرد را در موقعیتهای طبیعی به دست آوریم. استفاده از این روش در مورد کودکان کم سن و سال که نمی‌توانند ارتباط زبانی برقرار کنند، بی‌نهایت ارزشمند است.

به هر روی، مشاهده مستقیم به لحاظ مدت زمانی که برای مشاهده نیاز دارد، یک شیوه پرهزینه است. علاوه بر این، همیشه این امکان وجود دارد که حضور مشاهده‌گر رفتار فرد را تغییر دهد. برای غلبه بر این مشکل، در برخی از موقعیتهای، می‌توان از شیشه‌های یک سو به استفاده کرد. در بسیاری از موارد، آزمودنیهایی که مورد مشاهده قرار می‌گیرند، بعد از یک واکنش اصلی، توجه چندانی به مشاهده‌گر نمی‌کنند، بویژه آن که مشاهده‌گر بدون جلب نظر آزمودنی به کار خود مشغول باشد. برای آن که مشاهده‌گران به افراد کارآمدی بدل شوند و بدانند که چه چیزی را مشاهده کنند و نتیجه مشاهده خود را چگونه گزارش دهند، به آموزشهای گسترده‌ای نیاز دارند.

### / مشاهده‌های ابداعی

در مشاهده‌های ابداعی، محقق تدارک کار را به نحوی فراهم می‌آورد که آزمودنیها را در شبیه‌سازی موقعیتهای زندگی واقعی مورد مشاهده قرار دهد. شرایط به گونه‌ای فراهم می‌شود که رفتار دلخواه بروز کند.

یک شکل از مشاهده ابداعی، آزمون موقعیتی است. یک مثال کلاسیک از آزمون موقعیتی - اگر چه در زمان خود به این نام خوانده نشد - در مجموعه مطالعات هارتشورن و می دربار

## شکل ۵-۷ طبقه‌بندی تحلیل تعاملی فلاندرز (FIAC)\*

گفتار معلم:

۱- احساسات را می‌پذیرد؛ نگرش باحال و هوای احساس یک دانش‌آموز را به شیوه‌ای ملاطفت‌آمیز می‌پذیرد و توضیح می‌دهد. احساسات ممکن است مثبت یا منفی باشند. پیش‌بینی و به یادآوری احساسات را نیز شامل می‌شود.

۲- تحسین یا تشویق می‌کند؛ کنش با رفتار دانش‌آموز را تحسین یا تشویق می‌کند. لطیفه‌هایی می‌گوید تا تشنگی را کم کند، اما لطیفه‌های او دانش‌آموز دیگری را نمی‌آزارد؛ سرنگان دادن یا گفتن «وام هوم» یا ادامه بدنه را نیز شامل می‌شود.

پاسخ

۳- ایده‌های دانش‌آموزان را می‌پذیرد یا مورد استفاده قرار می‌دهد؛ ایده‌های پیشنهادی دانش‌آموز را توضیح می‌دهد، می‌پروراند یا بهبود می‌بخشد. بسط ایده‌های دانش‌آموز توسط معلم را نیز شامل می‌شود، اما اگر ایده‌ها، بیشتر متعلق به خود معلم باشد، به طبقه‌بندی شماره ۵ متعلق می‌شود.

۴- سؤال مطرح می‌کند؛ درباره شیوه یا محتوا سؤالی مطرح می‌کند که بر ایده‌های خود معلم مبتنی باشد، با این نیت که دانش‌آموز پاسخ دهد.

۵- سخنرانها؛ واقعیت‌ها یا نظرات را درباره محتوا و شیوه ارائه می‌دهد، ایده‌های خود را بیان می‌کند، توضیحات خود را ارائه می‌دهد، یا سنی به غیر از یک دانش‌آموز مثال می‌زند.

۶- رهنمود می‌دهد؛ رهنمود، فرمان، یا دستوری می‌دهد که انتظار می‌رود دانش‌آموزان از آن اطاعت کنند.

۷- اقتدار را نقد یا توجیه می‌کند؛ به قصد تغییر رفتار دانش‌آموز، از یک الگوی ناپذیرتنی به الگوی پذیرفتنی، عباراتی بیان می‌کند؛ کسی را دعوا می‌کند، دلیل انجام کار خود را شرح می‌دهد، منتها درجه خود مدار است.

گفتار دانش‌آموز

۸- پاسخ دانش‌آموز؛ گفته‌های دانش‌آموزان در پاسخ به معلم. معلم ارتباط را برقرار می‌کند، اظهارات دانش‌آموزان را خواستار می‌شود، یا به موقعیت شکل می‌دهد، آزادی بیان ایده‌های شخصی محدود است.

آغاز

۹- آغاز به وسیله خود دانش‌آموز؛ گفته‌های دانش‌آموزان که خود آغازگر آن باشند. ایده‌های خود را شرح می‌دهند، موضوع جدیدی را آغاز می‌کنند، این آزادی را دارند که نظرات یا بکت خط فکری را بسط دهند، نظیر طرح سؤالات فکروانه. از ساختار موجود فراتر می‌روند.

سکوت

۱۰- سکوت یا سردرگمی؛ دوره‌های کوتاه سکوت و دوره‌های سردرگمی که مشاهده‌گر نمی‌تواند ارتباط را درک کند.

بررسی تعلیم منش (CEI) مورد استفاده قرار گرفت.<sup>(۱۱)</sup> این آزمونها برای استفاده در مطالعه بسط و ویژگیهای رفتاری، نظیر درستکاری، خویشتنداری، راستگویی و همکاری طراحی شدند. هارتشورن و می‌مشاهداتی بر روی فعالیتهای روزمره مدرسه‌ای کودکان انجام دادند، اما همچنین موقعیتهایی را نیز برای تمرکز بر رفتار خاص آنان فراهم آوردند. برای مثال، آنها آزمونهای واژگان و خواندن به کودکان دادند، آزمونها را جمع‌آوری کردند و بدون آن که کودکان

\* این مقیاس اسمی است. هر عدد یک طبقه‌بندی را نشان می‌دهد، یعنی صابانگر یک نوع رویداد ارتباطی است. نوشتن اعداد با این ترتیب، به معنای ارزشگذاری هر رویداد ارتباطی نیست، بلکه این کار صرفاً به منظور طبقه‌بندی کردن این رویدادها صورت می‌گیرد.

Source: From N. A. Flanders (1970). *Analyzing Teaching Behavior* (p. 34). Boston: Addison - Wesley.

1- Hartshorne, H., M. A. May, and F.K. Shuttlesworth. *Studies in the Organization of Character*. New York: Macmillan.

آگاه باشند از پاسخهای آنان رونوشت برداشتند. سپس کلید پاسخها را به کودکان دادند و از آنها خواستند که برگه‌های اصلی آزمون خود را تصحیح کنند و نمره بدهند. تفاوت بین نمره‌هایی که کودکان گزارش دادند و نمره‌های واقعی که از تصحیح رونوشت پاسخها به دست آمد، میزانی برای تقلب به دست داد.

در آزمون دیگر، از کودکان خواسته می‌شد در حالی که چشمان خود را بسته‌اند، در داخل هر یک از ده دایره کوچکی که به طور نامنظم بر برگه آزمون ترسیم شده‌اند، علامتی بگذارند. آزمونه‌های کنترل شده قبلی در شرایطی که از زیرچشمی نگاه کردن کودکان جلوگیری می‌شد، نشان داد که در سه بار آزمون، کسب نمره بیش از ۱۳، یعنی ۱۳ علامتی که به درستی در داخل دایره‌ها جای داده شده باشند، بسیار نامحتمل است. بنابراین، نمره بالاتر از ۱۳ به عنوان نشانه‌ای از زیرچشمی نگاه کردن کودکان ثبت شد.

هارت‌شورن و می، به طور عملی هیچ همبستگی بین تقلب در موقعیت‌های متفاوت، نظیر تقلب در آزمون و در مسابقات ورزشی، نیافتند. آنها نتیجه گرفتند که پاسخهای کودکان از نظر موقعیتی، خاص هستند. به این معنا که، اعم از این که کودکان بسته به فعالیت خاص، حضور معلم و دیگر موقعیتها تقلب کردند، اما در صفات شخصیتی عمومی، نشانی از تقلب نداشتند.

### /// خلاصه

یک وظیفه مهم محققان علوم رفتاری، انتخاب ابزار اندازه‌گیری قابل اعتماد به منظور کمی کردن اطلاعات تحقیق است. در تحقیق آموزشی، آزمونها رایجترین ابزار اندازه‌گیری هستند. یک آزمون، عبارت است از مجموعه‌ای از محرکها که به یک فرد ارائه می‌شود تا پاسخی را از او بگیرد که بر مبنای آن بتوان یک نمره عددی تعیین کرد. آزمونه‌های پیشرفت تحصیلی، نمونه اصلی این نوع ابزار اندازه‌گیری هستند. آزمونه‌های پیشرفت تحصیلی متعددی وجود دارند که هنجارهایی را فراهم می‌آورند که می‌توان از آن هنجارها به عنوان مبنای مقایسه استفاده کرد. آزمونه‌های هوش ابزار سنجش قابلیت‌های کلامی و غیر کلامی یک فرد هستند. پرسشنامه‌های شخصیت برای اندازه‌گیری ویژگیهای فردی و عملکرد نوعی آزمودنی طراحی شده‌اند. مقیاسهای نگرش، ابزاری برای اندازه‌گیری باورها، احساسها و واکنشهای فرد نسبت به موضوعات معین هستند. انواع اصلی مقیاسهای نگرش عبارتند از: مقیاسهای نوع لیکرت، مقیاسهای ترستون،

مقیاسهای گاتمن و افتراق معنایی. روشهای گروه سنجی، ابزار سنجش موقعیت فرد در میان همسالان خود می‌باشند. از طریق این روشها، یافتن افراد محبوب گروهها (ستاره‌ها)، منزویها و گروهکها امکان‌پذیر می‌شود.

تعدادی از روشهایی که رفتار آزمودنیها را به صورت نظام‌مند مورد مشاهده قرار می‌دهند، به عنوان روشهای مشاهده مستقیم و مشاهده ابداعی شکل گرفته‌اند.

### // مفاهیم اصلی

آزمون استاندارد شده	آزمون
آزمون پیشرفت تحصیلی	آزمون استعداد تحصیلی
اثر هاله‌ای	آزمون هوش
خطای سخاوتمندی	پرسشنامه
خطای گرایش مرکزی	خطای سختگیری
مشاهده ابداعی	روش فراقکن
مقیاس افتراق معنایی	مشاهده مستقیم
مقیاس ترستون	مقیاس تراکمی
مقیاس گاتمن	مقیاس
مقیاس مجموع درجه‌ها	مقیاس لیکرت
نگاره گروهی	مقیاس نگرش

### /// تمرینها

- ۱- معنای اصطلاح استاندارد شده، هنگامی که به ابزارهای اندازه‌گیری اطلاق می‌شود، چیست؟
- ۲- تفاوت بین مقیاسهای درجه‌بندی مقایسه‌ای و مقیاسهای نگاره‌ای و طبقه‌ای چیست؟
- ۳- برخی از منابع رایج آریبی در مقیاسهای درجه‌بندی را فهرست کنید.
- ۴- چرا مقیاس گاتمن، در مقایسه با مقیاسهای ترستون و لیکرت، تک‌بعدی محسوب

می‌شود؟

- ۵- استفاده اصلی مقیاس افتراق معنایی در اندازه‌گیری چیست؟
- ۶- برخی از شیوه‌های افزایش دقت در روشهای مشاهده مستقیم چه هستند؟
- ۷- یک مقیاس لیکرت، با پنج عبارت، برای اندازه‌گیری نگرش معلمان در مورد آموزش زبان خارجی در مدارس ابتدایی، بسازید.

- ۸- مزیت‌های اصلی مقیاس لیکرت نسبت مقیاس ترستون چیست؟
- ۹- یک مقیاس درجه‌بندی نگاره‌ای با پنج عبارت بسازید که برای ارزشیابی یک برنامه تحقیقاتی مفید باشد.
- ۱۰- داده‌های زیر، درجه‌بندی‌هایی است که از سوی ۲۰۰ داور به یک عبارت مقیاس ترستون تخصیص یافته است. چه ارزش مقیاسی به این عبارت تعلق خواهد گرفت؟

تعداد داورها	طبقه‌بندی
۰	۱۱
۰	۱۰
۰	۹
۳۰	۸
۵۰	۷
۶۰	۶
۳۴	۵
۱۶	۴
۱۰	۳
۰	۲
۰	۱
-----	
۲۰۰	

- ۱۱- اعداد زیر، ارزشهای مقیاسی پنج عبارت یک مقیاس ترستون هستند که دو دانشجو آنها را علامت زده‌اند. نمره‌های نگرش آنها را محاسبه کنید و نمره‌های آنها را برحسب این که نگرشهای آن دو را تا چه حد مطلوب و روشن تعریف می‌کنند، تفسیر کنید.

جان	جین
۱۰/۵	۵/۵
۸/۲	۴/۶
۷/۱	۴/۱
۳/۹	۳/۸
۲/۸	۳/۱

## /// پاسخها

۱- استاندارد شده به پرسشنامه‌ای گفته می‌شود که برای آن هنجارهای تطبیقی به دست آمده باشد، پایایی و اعتبار آنها تعیین شده باشد و دستورالعملهایی برای اجرا و نمره دهی آنها مقرر شده باشد.

۲- در مقیاسهای نگره‌ای و طبقه‌ای، درجه بندی کننده به هنگام داوری بر روی یک فرد، او را با افراد دیگر مقایسه مستقیم نمی‌کند. در مقیاس مقایسه‌ای، به هنگام داوری بر روی یک فرد، درجه بندی کننده باید گروهی را که آن فرد با آن مقایسه می‌شود، بشناسد.

۳- هنگامی که گرایشهایی نظیر اثر هاله‌ای، خطای سخاوتمندی، خطای سختگیری یا خطای گرایش مرکزی تأثیر خود را بر فرد درجه بندی کننده گذارده باشند، داوری او از عینیت کمتری برخوردار خواهد بود.

۴- مقیاسهای ترستون و لیکرت ممکن است حاوی عباراتی در مورد ابعاد مختلف موضوع نگرش باشند. مقیاس گاتمن تک بعدی محسوب می‌شود؛ زیرا سوالاتی درباره درجات واکنش فرد نسبت به یک موضوع منفرد را مطرح می‌کند. بنابراین، مقیاس گاتمن تراکمی است، یعنی می‌توان پیش‌بینی کرد که اگر فردی با یک عبارت مقیاس موافقت کند، با تمام عبارات بعدی نیز موافقت خواهد کرد.

۵- مقیاس افتراق معنایی برای اندازه‌گیری معنای ضمنی که فرد برای یک شیء نگرشی قایل می‌شود، به کار می‌رود.

۶- رفتارهایی که قرار است مشاهده شوند، باید مشخص گردند، رفتارهایی که در یک طبقه قرار می‌گیرند، باید تعریف شوند، برای کمی کردن نتایج مشاهده باید سیستمی تهیه شود و مشاهده‌گران باید آموزش ببینند تا کار مشاهده را مطابق با شیوه‌های وضع شده به انجام رسانند. ۹- پاسخها متفاوت خواهند بود.

۶/۱۷-۱۰

۱۱- جین با میانگین  $4/2$  و میانه  $4/1$ ، یک نگرش نسبتاً مثبت را نشان می‌دهد. انتخابهای او از نظر ارزش مقیاسی به هم نزدیک هستند، که این نکته، همسانی را نشان می‌دهد. میانگین جان  $6/5$  و میانه او  $7/1$  است، که نشان‌دهنده نگرش کمی نامطلوب اوست. پراکندگی زیاد ارزشهای مقیاسی، نگرشی را که به طور روشن تعریف نشده است، نمایان می‌کند.



## فصل هشتم

### اعتبار و پایایی

#### اهداف آموزشی

دانشجو بعد از مطالعه این فصل قادر خواهد بود:

- ۱- بین اعتبار و پایایی تمایز قایل شود.
- ۲- سه نوع اصلی اعتبار و روشهای مورد استفاده در سنجش آنها را توصیف کند.
- ۳- آن نوع اعتباری را که برای اندازه‌گیری خاصی مناسب است، توضیح دهد.
- ۴- بین اعتبار همگرا و اعتبار افتراقی تفاوت قایل شود.
- ۵- رابطه بین پایایی و مفهوم خطای تصادفی اندازه‌گیری را توضیح دهد.
- ۶- منابع مختلف خطا در اندازه‌های آموزشی و روان‌شناختی را بیان کند.
- ۷- شیوه‌های مختلف (بازآزمایی، صورت‌های هم ارز، دو نیمه سازی، کوردر - ریچاردسون، و سایر) برای برآورد پایایی یک اندازه را توصیف کند.
- ۸- بر اساس داده‌های ارائه شده، ضریب پایایی را محاسبه کند.
- ۹- فرمول اسپیرمن - براون را برای تعیین اثر مطول کردن یک آزمون بر پایایی آزمون به کار

ببرد.

- ۱۰- عوامل مؤثر بر مقدار ضریب پایایی را شرح دهد.
- ۱۱- خطای معیار اندازه‌گیری را تفسیر کند و رابطه آن را با پایایی آزمون توضیح دهد.
- ۱۲- خطای معیار اندازه‌گیری را محاسبه و دسته‌های نمره‌ها را تفسیر کند.
- ۱۳- شاخصهایی را برای نشان دادن پایایی یک آزمون ملاکی محاسبه کند.

تحقیق همواره به اندازه گیری وابسته است. دو ویژگی مهم وجود دارند که هر ابزار اندازه گیری باید آنها را دارا باشد: اعتبار و پایایی. اعتبار، به این معنا اطلاق می شود که یک ابزار اندازه گیری تا چه حد آن چیزی را اندازه می گیرد که مورد نظر بوده است. از سوی دیگر، پایایی میزان ثباتی است که یک وسیله اندازه گیری، در اندازه گیری هر آنچه که اندازه می گیرد، از خود نشان می دهد. روشنتر بگوئیم، اعتبار و پایایی به اطلاعاتی که یک ابزار اندازه گیری تولید می کند، مربوط می شود تا به خود ابزار اندازه گیری. ما برای سهولت، از اعتبار و پایایی یک آزمون صحبت می کنیم، اما اگر بخواهیم دقیقتر سخن بگوئیم، باید از اعتبار و پایایی نمره های آزمون صحبت کنیم. اگر چه آزمون ایستاست، اما نتایج آن (نمره ها) برحسب موقعیت تغییر می کنند. یک محقق باید در جستجوی اعتبار و پایایی ابزار اندازه گیری ای باشد که در بررسی خود از آن استفاده کرده است و اطلاعات مربوط به آن را در گزارش تحقیقاتی خود بگنجانند. اگر داده های یک محقق با ابزار اندازه گیری پایا و دارای اعتبار به دست نیامده باشد، به نتایج حاصله از آن با تردید نگریده خواهد شد.

شواهد مربوط به اعتبار و پایایی، بویژه در تحقیقات آموزشی، اهمیت دارند؛ زیرا اکثر اندازه گیریها در این حیطه به طور غیر مستقیم انجام می شوند. سنجش این امر برای محقق ضروری است که یک ابزار اندازه گیری آموزشی یا روان شناختی، تا چه حد آنچه را که قرار بوده است اندازه بگیرد، با دقت و اطمینان اندازه می گیرد.

### /// اعتبار

امر اعتبار با این مسئله سروکار دارد که یک ابزار اندازه گیری تا چه حد آن چیزی را اندازه می گیرد که ما تصور می کنیم. پرسیدن این سؤال از سوی محقق یک امر کاملاً اساسی است. ابزار اندازه گیری آموزشی و روان شناختی برای ارزیابی سازه هایی نظیر پیشرفت تحصیلی، هوش، خلاقیت، استعداد تحصیلی، نگرش، انگیزش و نظایر آنها طراحی شده اند. به هر روی، برخلاف علوم تجربی که می توان ویژگیهایی نظیر طول، حجم و وزن را به طور مستقیم اندازه گیری کرد، برای اندازه گیری این سازه ها هیچ راه مستقیمی وجود ندارد. محققان باید راههای غیر مستقیمی برای اندازه گیری ویژگیهای پیچیده ابداع کنند. بنابراین، این سؤال مطرح می شود که: «شیوه های غیرمستقیم، چیزی را که قرار است اندازه بگیرند، با چه کیفیتی اندازه می گیرند؟» پژوهشگران

باید چنین پرسشهایی مطرح کنند: آیا این آزمون واقعاً انگیزه پیشرفت را اندازه می‌گیرد؟ آیا این آزمون سایر کیفیتها را نیز به همان خوبی اندازه می‌گیرد؟ آیا این آزمون خلاقیت واقعاً می‌تواند افراد بسیار خلاق را از افراد عادی متمایز کند؟ آیا می‌توان بر مبنای نمره‌های این آزمون استعداد تحصیلی، پیش‌بینی‌های مفیدی انجام داد؟ آیا این آزمون ابزار مناسبی برای استفاده در مورد همه دانش‌آموزان است، یا آن را تنها باید در مورد گروههای خاص به کار برد؟ برای مثال، آزمونهای استعداد تحصیلی، برای پیش‌بینی پیشرفت تحصیلی دانشجویان متعلق به اقلیتهای قومی، چقدر مناسب هستند؟ این پرسشها همه به معنادار بودن و سودمند بودن استنباطهایی که بر مبنای نمره‌ها صورت خواهد گرفت، یعنی به اعتبار آزمون، مربوط می‌شوند.

اعتبار همیشه مختص مقصود و ویژه‌ای است که ابزار اندازه‌گیری برای آن مقصود مورد استفاده قرار می‌گیرد. برای مثال، آزمونی که در یک موقعیت و برای یک مقصود دارای اعتبار است، ممکن است در یک موقعیت دیگر و برای یک مقصود دیگر اعتبار نداشته باشد. مثلاً، یک آزمون تاریخ استاندارد شده، که بر درک و تفسیر تاریخ تأکید دارد، ممکن است از نظر یک آموزگار اعتبار داشته باشد، اما برای آموزگار دیگری که تأکیدش بر زمان رویدادهای تاریخی و واقعتهای تاریخی است، اعتبار نداشته باشد. مقصودی که یک آزمون برای آن مورد استفاده قرار می‌گیرد نیز عامل مهمی برای اعتبار آزمون است. یک آزمون شیمی استاندارد شده ممکن است برای اندازه‌گیری پیشرفت تحصیلی در درس شیمی در پایان سال مورد استفاده قرار گیرد. در این مورد، سؤالاتی که مطرح می‌شود، این است که آزمون چه چیزی را اندازه می‌گیرد و این اندازه‌گیری را چقدر خوب انجام می‌دهد. امکان دارد همان آزمون برای پیش‌بینی پیشرفت تحصیلی در درس شیمی دوره دانشگاه به کار رود. در چنین موردی علاقه‌مندیم بدانیم که این آزمون، آینده تحصیلی را چقدر خوب پیش‌بینی می‌کند. مقاصد مختلفی که برای آزمونها در نظر گرفته می‌شود، حاکی از آن است که به انواع مختلفی از شواهد نیاز داریم تا اعتبار آزمون برای هر کاربرد خاص تأیید شود.

هر چند اعتبار، یک مفهوم یگانه است، اما انواع مختلفی از شواهد وجود دارند که می‌توانند برای تأیید استنباطهایی که بر مبنای نمره‌های یک ابزار اندازه‌گیری به عمل می‌آیند، گردآوری شوند. کمیته مشترک انجمن آمریکایی روان‌شناسی (APA)<sup>(۱)</sup>، انجمن آمریکایی تحقیقات

آموزشی (AERA)<sup>(۱)</sup> و شورای ملی سنجش در تعلیم و تربیت (NCME)<sup>(۲)</sup> راههای گردآوری شواهد اعتباری را در مقوله‌های زیر طبقه‌بندی می‌کند: شواهد اعتباری مرتبط با محتوا، مرتبط با ملاک و مرتبط با سازه.<sup>(۳)</sup>

### // شواهد مرتبط با محتوا

آزمونهای کلاسی به طور کلی به منظور سنجش دانش و مهارتهای دانش‌آموزان در یک حیطه محتوایی تعریف شده مورد استفاده قرار می‌گیرند. شیوه آزمانی برای تحقق این امر، استفاده از آزمونی است که همه سؤالات احتمالی را درباره آن محتوا در بر بگیرد. شیوه مفید دیگر، نمونه‌گیری از کل حیطه محتوایی و سپس استفاده از آن نمونه به عنوان مبنایی برای استنباط درباره آگاهی دانش‌آموزان از کل آن حیطه محتوایی است. از آنجا که استنباطها تنها بر مبنای یک نمونه صورت می‌گیرند، نکته مهم آن است که این نمونه معرف تمام حیطه محتوایی باشد، یعنی یک نمونه دارای اعتبار باشد.

این امر ما را به شواهد مرتبط با محتوا رهنمون می‌سازد. این نوع شواهد نشان می‌دهند که نمونه پرسشهای یک آزمون تا چه حد معرف یک حیطه محتوایی تعریف شده هستند. این شواهد از طریق بازبینی دقیق و موشکافانه محتوای آزمون به وسیله داوران خبره گردآوری می‌شود تا رابطه بین آزمون و حیطه محتوایی تعریف شده تعیین شود. این روش از دیدگاه خاصی که برای کاربرد نتایج آزمون در نظر گرفته شده است، انجام می‌گیرد.

البته، این حیطه محتوایی نظری است. اگر کسی در پی تهیه آزمونی باشد، ابتدا باید رئوس موضوعات، مهارتها و تواناییهای موجود در آن حیطه را که قرار است اندازه‌گیری شود، تهیه و اهمیت هر یک را با علایمی مشخص کند. تعداد زیادی از پرسشهای آزمون را می‌توان با استفاده از این رئوس مطالب به عنوان راهنما نوشت. پرسشهای آزمون را می‌توان از هر یک از طبقات رئوس مطالب، به طور تصادفی بیرون کشید؛ به گونه‌ای که تعداد پرسشها منعکس کننده وزن نسبی هر طبقه در کل حیطه باشند. نمونه پرسشهای حاصله باید معرف آن حیطه محتوایی باشند. برای مثال، حیطه محتوایی آزمونی را که درباره جنگ داخلی آمریکا باشد، می‌توان به

1- American Educational Research Association

2- National Council on Measurement in Education

3- American Psychological Association (1985). *Standards for Educational and Psychological Testing*. Washington: APA.

صورت شناخت و درک موضوعاتی نظیر دلایل جنگ، استراتژیهای نظامی و عملیات جنگی، شخصیت‌های مهم، تأثیرات اقتصادی، اثرات آن بر تاریخ پس از جنگ آمریکا و نظایر آن تعریف کرد. پرسشهای آزمون درباره هر یک از این موضوعات، نه تنها باید آگاهی از موضوع، بلکه درک، تفسیر، تحلیل و هر هدف شناختی دیگری را که در درس بر آن تأکید شده است، اندازه بگیرد. فرض کنید یک معلم ادبیات انگلیسی بر درک ایده‌های نویسندگانی معین و کاربرد آن ایده‌ها در قرن بیستم تأکید کرده است. اگر آزمون این معلم، اکثراً حاوی پرسشهایی باشد که در آنها از دانش‌آموزان خواسته شده است نام نویسندگان را یا نام آثار آنها تطبیق دهند، یا تاریخ تولد آنها را به یاد آورند، این معلم در دستیابی به یک نمونه معرف از حیطه محتوایی درس خود ناکام مانده است و آن آزمون برای مقصود موردنظر، یعنی اندازه‌گیری درک دانش‌آموزان از ایده‌های نویسندگان، اعتبار کمی خواهد داشت.

شواهد مرتبط با محتوا معمولاً به صورت عددی بیان نمی‌شوند. گردآوری چنین شواهدی اساساً و ضرورتاً مبتنی بر داوری است و چنین داوریایی باید برای هر مقصودی به طور جداگانه انجام پذیرد. گردآوری این شواهد شامل بازبینی دقیق و موشکافانه‌ای است تا تعیین شود که آیا محتوا و هدفهایی که آزمون اندازه‌گیری می‌کند، معرف آن مطالبی است که حیطه محتوا را تشکیل می‌دهند یا خیر. سازنده آزمون باید تعیین کند که آیا پرسشهای آزمون، معرف دوره درسی و هدفهای آن، که در راهنمای برنامه درسی و کتابهای درسی بیان شده است، هستند یا خیر. سازنده آزمون برای آن که به یک ارزشیابی بیرونی از اعتبار محتوا دست یابد، باید از تعدادی از آموزگاران یا دیگر افراد خبره بخواهد که محتوای آزمون را به طور نظام‌مند بازبینی و ارتباط آن را با حیطه محتوایی مشخص شده ارزیابی کنند. اگر تمام این افراد توافق کنند که پرسشهای آزمون حیطه محتوایی را به اندازه کافی منعکس می‌کند، پس می‌توان گفت که آزمون، اعتبار محتوا دارد. همچنین این نکته نیز باید مشخص شود که آزمون، از عواملی که با مقصود اندازه‌گیری نامرتبط هستند، تأثیر نپذیرفته است. برای مثال، عملکرد دانش‌آموز در یک آزمون ریاضیات نباید تحت تأثیر سرعت خواندن و واژه‌شناسی قرار گیرد. وجود چنین عواملی در یک آزمون ریاضیات، اعتبار آن را کاهش خواهد داد؛ زیرا در چنین حالتی آزمون چیزی را اندازه‌گیری خواهد کرد که نیت اندازه‌گیری آن را ندارد. شواهد مرتبط با محتوا، بویژه در ارزشیابی اعتبار آزمونهای پیشرفت تحصیلی، اهمیت دارند.

محقق همیشه باید شواهد مرتبط با محتوا را در مورد اعتبار یک آزمون پیشرفت تحصیلی خود ساخته یا استاندارد شده، که قرار است در بررسی او مورد استفاده قرار گیرد، گردآوری کند. به طور کلی، ناشران آزمونها، شواهد اعتباری گسترده‌ای از این نوع را ارائه می‌کنند. به هر روی، باید دوباره تأکید کنیم که یک آزمون پیشرفت تحصیلی، که برای سازنده آزمون اعتبار زیادی دارد، ممکن است برای استفاده کننده که شاید حیطه محتوایی را به گونه‌ای دیگر تعریف کرده است، اعتبار نداشته باشد. در نهایت تنها استفاده کننده یک آزمون پیشرفت تحصیلی می‌تواند در مورد اعتبار آن آزمون برای مقصود خود، قضاوت کند.

### // شواهد مرتبط با ملاک

شواهد مرتبط با ملاک نشان می‌دهند که نمره‌های یک ابزار اندازه‌گیری تا چه اندازه با یک متغیر مستقل بیرونی (ملاک)، که گمان می‌رود رفتار یا خصیصه مورد نظر را به طور مستقیم اندازه می‌گیرد، مرتبط هستند. هنگامی که کسی رابطه بین نمره‌های یک آزمون استعداد تحصیلی و معدل دانشگاه (GPA) را بررسی می‌کند، در واقع شواهد مرتبط با ملاک را در مورد اعتبار آزمون گردآوری می‌کند. میزان رابطه نمره‌های آزمون استعداد تحصیلی با موفقیت در دانشگاه، که با GPA اندازه‌گیری می‌شود، مقدار اعتباری است که آزمون استعداد تحصیلی برای پیش‌بینی GPA دارد.

در این نوع شواهد، همان‌طور که از نام آن بر می‌آید، تأکید بر ملاک و شیوه‌های اندازه‌گیری مورد استفاده برای دستیابی به نمره‌های ملاکی است. بنابراین، انتخاب ملاک برای حصول موفقیت در این نوع بررسی، امری حیاتی است. ویژگیهای متعددی وجود دارند که یک اندازه ملاکی باید دارا باشد. یکی از مهمترین آنها، ربط<sup>(۱)</sup> است. محقق باید قضاوت کند که آیا ملاک انتخابی واقعاً بر عملکرد موفقیت آمیز رفتار مورد نظر دلالت می‌کند یا خیر. اگر آن ملاک، خصیصه مورد مطالعه را منعکس نکند، استفاده از آن به عنوان مبنایی برای اعتباریابی یک ابزار دیگر، بی‌معنا خواهد بود. چنین فرض می‌شود که GPA یک اندازه مرتبط با موفقیت در دانشگاه است و به طور کلی به عنوان ملاکی برای اعتباریابی آزمونهای استعداد تحصیلی، که برای

گزینش دانشجو از میان داوطلبان ورود به دانشگاه ساخته می‌شوند، به کار می‌رود. به منظور اعتباریابی آزمونی که برای انتخاب فروشنده طراحی شده است، ملاک مرتبط می‌تواند مقدار فروش در یک زمان مشخص باشد. در برخی از موارد، یافتن ملاکهای مرتبط شاید به دشواری صورت گیرد. برای مثال، یافتن ملاکهایی برای اعتباریابی اندازه‌هایی که برای پیش‌بینی اثربخشی معلمان به کار می‌رود، دشوار بوده است. بدون آن که توصیفی مورد توافق همگان از اثر بخشی معلم و روش مؤثری برای اندازه‌گیری آن متغیر در دست داشته باشیم، عملاً تعیین اعتبار هر ابزاری که برای شناسایی داوطلبان خوش آتیه شغل معلمی طرح شده، غیر ممکن است.

ویژگی دوم آن است که یک ملاک باید پایا<sup>(۱)</sup> باشد. این بدان معناست که ملاک باید اندازه‌ها یا ثباتی از خصیصه مورد نظر در طول زمان یا از موقعیتی به موقعیتی دیگر باشد. اگر ملاک، خود با ثبات نباشد، انتظار نخواهد رفت که همواره با پیش‌بینی کننده مورد نظر مرتبط شود.

یک ملاک باید عاری از جهت‌گیری باشد، یعنی نمره دهی یک اندازه ملاکی نباید از عاملی به جز عملکرد واقعی بر مبنای آن ملاک تأثیر پذیرد. فرض کنید ملاک مورد استفاده برای اعتباریابی آزمونی که برای انتخاب داوطلبان یک شغل خاص به کار می‌رود، درجه‌بندی یک سرپرست است. اگر آن سرپرست اجازه دهد که یک نظر کلی از فرد مورد نظر، یا هر عامل دیگری بجز عملکرد واقعی او، بر درجه‌بندی تأثیر بگذارد، پس نمره ملاکی جهت‌دار خواهد بود. وقتی که ملاک ما درجه‌بندی است، برای اجتناب از جهت‌گیری، باید آموزشهای صریحی درباره ویژگیهایی که قرار است درجه‌بندی شوند و نحوه انجام درجه‌بندی ارائه کنیم. شیوه درجه‌بندی هر چه عینی‌تر باشد، جهت‌گیری کمتری در ملاک وارد خواهد شد. منبع احتمالی دیگر جهت‌گیری در ملاک، ناخالصی<sup>(۲)</sup> است. ناخالصی هنگامی به وجود می‌آید که نمره یک فرد در ملاک مورد نظر، تحت تأثیر آگاهی فرد نمره‌گذار از نمره پیش‌بینی کننده او قرار گیرد. برای مثال، فرض کنید کسی یک آزمون استعداد هنری دارد که باید با استفاده از نمره‌های کلاس هنر به عنوان ملاک، اعتباریابی شود. آگاهی معلمان از نمره شاگردان در آزمون استعداد هنری، ممکن است بر ارزیابی آنها از شاگردان تأثیر بگذارد. اگر به فرد نمره دهنده یا درجه‌بندی کننده اجازه دیدن نمره‌های پیش‌بینی کننده را ندهیم، می‌توانیم از ناخالصی ملاک جلوگیری کنیم.



هنگامی که ملاک بیرونی تعریف شد، داده‌های تجربی را برای سنجش رابطه بین نمره‌های حاصل از ابزار اندازه‌گیری (X) و نمره‌های ملاکی (Y) گردآوری می‌کنیم. سپس، ابزار اندازه‌گیری را که قرار است اعتباریابی شود، با گروهی از افراد، که معرف آنهايي باشند که ابزار اندازه‌گیری در مورد آنها استفاده خواهد شد، به کار می‌گیریم. نمره‌هایی را که توسط این آزمودنیها در پیش‌بینی کننده (X) کسب می‌شود، به منظور جلوگیری از ناخالصی نمره‌های ملاکی، در جایی حفظ می‌کنیم و این نمره‌ها را در هیچ یک از تصمیم‌گیریهایی که ممکن است بر سرنوشت آتی آنها تأثیر بگذارد، دخالت نمی‌دهیم. سپس وقتی که داده‌های ملاکی (Y) حاضر شد، به سراغ آزمونهای اصلی می‌رویم و همبستگی نمره‌های آن آزمونها را با نمره‌های ملاکی محاسبه می‌کنیم. ضریب همبستگی بین این دو مجموعه از اندازه‌ها، ضریب اعتبار ( $r_{yy}$ ) نامیده می‌شود و نشان می‌دهد که نمره‌های آزمون (X) با چه دقتی ملاک (Y) را پیش‌بینی می‌کنند. هر چه ( $r_{yy}$ ) بزرگتر باشد، آزمون (X) با دقت بیشتر ملاک (Y) را پیش‌بینی می‌کند.

نمونه‌ای از این شیوه در اعتباریابی آزمون استعداد تحصیلی (SAT) به کار رفت. آزمون استعداد تحصیلی در بررسیهای متعددی به گروه بزرگی از دانش‌آموزان سال آخر دبیرستان داده شد و نتایج آن تا زمانی که این دانش‌آموزان وارد دانشگاه شدند و سال اول دانشگاه را به پایان رساندند، بایگانی شد. در این زمان، برای به دست آوردن ضریب اعتبار آزمون، همبستگی نمره‌های SAT با معدل سال اول دانشگاه یا GPA (ملاک) محاسبه شد. از آنجا که در مطالعات مکرر این نتیجه حاصل شد که SAT، اعتبار مرتبط با ملاک سودمندی دارد، اکنون به طور معمول برای پیش‌بینی عملکرد دانشگاهی مورد استفاده قرار می‌گیرد. دانش‌آموزان معمولاً در سال آخر دبیرستان در آزمون استعداد تحصیلی شرکت و نمره‌های SAT خود را به دانشگاهها ارسال می‌کنند. مسئولان پذیرش دانشگاهها، با آگاهی از اعتبار پیش‌بینی، SAT، نمره‌های داوطلبان را بررسی می‌کنند و برای پذیرش دانشجوی، حداقل تا حدی، بر مبنای نمره‌های SAT تصمیم‌گیری می‌کنند. هر چه نمره SAT بالاتر باشد، احتمال موفقیت در دانشگاه بیشتر است. در بسیاری از موارد، پیش‌بینی موفقیت در دانشگاه بر پایه ترکیبی از نمره SAT و برخی از اندازه‌های پیشرفت تحصیلی در دبیرستان، نظیر رتبه کلاسی، صورت می‌گیرد. ترکیب نمره SAT و رتبه دبیرستانی، معدل دانشگاهی را دقیق‌تر از هر یک از این دو اندازه پیش‌بینی می‌کند. شواهد مرتبط با ملاک، برای آزمونهایی که به منظور انتخاب و طبقه‌بندی به کار می‌روند،

یک ضرورت است. قبل از آن که بتوان از آزمونی در امر انتخاب استفاده کرد، باید شواهدی در دست داشت که آن آزمون براستی می‌تواند عملکرد را در مورد آن ملاک مشخص پیش‌بینی کند. پرسش اساسی این است که نمره‌های آن آزمون یا چه دقتی می‌توانند عملکرد ملاکی را پیش‌بینی کنند. این که نمره آزمون بالا باشد یا پایین، مفید باشد یا بی‌فایده، به زمینه‌ای که آزمون در آن مورد استفاده قرار می‌گیرد، بستگی دارد. در مواردی که از قبل هیچ ابزار پیش‌بینی وجود نداشته است، یک ضریب همبستگی  $0/4$  می‌تواند بسیار سودمند باشد. در سایر موارد، اگر پیش‌بین‌های دیگری در دسترس باشند که با ملاک، رابطه قوی‌تری دارند، یک همبستگی  $0/65$  شاید کم و ناکافی به حساب آید. به طور کلی، یک وسیله هنگامی به عنوان ابزار پیش‌بینی، اعتبار «مطلوبی» دارد که شواهد حاکی باشند که همبستگی آن با ملاک، بالاتر از ابزار پیش‌بینی رقیب است. انباشتگی شواهد مرتبط با ملاک برای اعتبار یک آزمون، نیاز به زمان و شکیبایی دارد. در برخی از موارد، برای تعیین این که آیا عملکرد در یک اندازه برای پیش‌بینی موفقیت در یک ملاک مفید است یا خیر، به سالهای متمادی صبر نیاز است.

برای کسب شواهد مرتبط با ملاک می‌توان دو طرح را از هم تمیز داد: بررسیهای پیش‌بینی<sup>(۱)</sup> و بررسیهای همزمان<sup>(۲)</sup>. این طرحها، هر دو، با رابطه تجربی بین نمره‌های آزمون و یک ملاک سروکار دارند، اما از نظر زمان گردآوری داده‌های ملاکی، بین آنها تفاوت وجود دارد. یک بررسی پیش‌بینی، اطلاعاتی را درباره همبستگی بین نمره‌های آزمون و یک ملاک گردآوری می‌کند که آن ملاک در زمان آینده به منصفه ظهور خواهد رسید. یک بررسی همزمان، اطلاعاتی را درباره همبستگی بین نمره‌های آزمون و یک اندازه ملاکی که در همان زمان در دسترس است، گردآوری می‌کند. برای مثال، بررسی همزمان یک آزمون قرائت که در پایان کلاس چهارم به اجرا در می‌آید، به رابطه بین نمره‌های آزمودنیها در آن آزمون و نمره‌هایی که آموزگاران به همان آزمودنیها در درس قرائت کلاس چهارم داده‌اند، نظر دارد؛ در حالی که یک بررسی پیش‌بینی، همبستگی نمره‌های همان آزمون را با نمره‌هایی که دانش‌آموزان، بعداً، در درس قرائت کلاس پنجم خواهند گرفت، مدنظر قرار می‌دهد. بررسیهای پیش‌بینی، به طور کلی در اعتباریابی آزمونهایی که در تعلیم و تربیت و صنعت به منظور انتخاب به کار می‌روند، ارجحیت دارند.

بررسیهای همزمان، به طور کلی برای آزمونهای پیشرفت تحصیلی، آزمونهایی که برای اندازه‌گیری سازه‌ها طراحی شده‌اند، یا آزمونهایی که برای تأیید یا تشخیص مورد استفاده قرار می‌گیرند، ارجحیت دارند.

از آنجا که ضریب اعتبار یک ضریب همبستگی است، مقدار آن تحت تأثیر همان عواملی قرار می‌گیرد که بر هر ضریب همبستگی تأثیر می‌گذارند، یعنی می‌توان گفت که خطی بودن رابطه بین آزمون و ملاک و دامنه تفاوت‌های فردی در گروه، بر ضریب اعتبار تأثیر دارند.

### // شواهد مرتبط با سازه

در شواهد مرتبط با سازه، کانون توجه بر نمره‌های آزمون، به عنوان اندازه‌ای از یک صفت یا سازه روان‌شناختی قرار دارد. همچنان که از فصل اول به خاطر دارید، واژه سازه به چیزی اطلاق می‌شود که خود به طور مستقیم قابل اندازه‌گیری نیست، اما اثرات قابل مشاهده را توضیح می‌دهد. سازه بلوغ اجتماعی «ساخته» شد تا بر الگوهای رفتاری مشاهده شده دلالت کند. بلوغ اجتماعی را نمی‌توان به طور مستقیم اندازه‌گیری کرد، اما می‌توان بسیاری از رفتارهایی را که ما معتقدیم و جوهی از این سازه هستند، توصیف و اندازه‌گیری کرد و مجموع مقادیر آنها نیز اندازه غیرمستقیمی از سازه مجرد بلوغ اجتماعی به ما می‌دهد. برخی از دیگر نمونه‌های متداول سازه‌ها عبارتند از: اضطراب، هوش، انگیزش، توانایی استدلال، تفکر نقادانه، استعداد در عرصه‌های مختلف، درک مطلب، و خود پنداره.

انسانها در طول تاریخ، تجزیدات پیچیده‌تر را از مفاهیم خود کنار هم‌گذارند یا ساختند. همان طور که یک کودک قطعات چوبی اسباب بازی را کنار هم می‌چیند و از آنها «اسب» یا «آدم» درست می‌کند، انسانها نیز با ترکیب مفاهیم و سازه‌های ساده‌تر، در قالبهای معنی‌دار، سازه‌ها را خلق می‌کنند. انگیزه اعتباریابی سازه‌ها از نظریه شخصیت و نیاز محققان به روشی برای اعتباریابی ابزاری که در ساخت نظریه به کار می‌رود، نشئت گرفته است. نه شواهد مرتبط با محتوا و نه شواهد مرتبط با ملاک، سازه‌ای را که یک آزمون اندازه‌گیری می‌کند، مورد توجه قرار نمی‌دهند. هدف گردآوری شواهد مرتبط با سازه، تعیین این امر است که یک آزمون کدام سازه روان‌شناختی را اندازه‌گیری می‌کند و این اندازه‌گیری را تا چه حد خوب انجام می‌دهد.

بررسی سازه، هم یک رویکرد منطقی و هم یک رویکرد تجربی را به کار می‌گیرد. یک وجه

از رویکرد منطقی این است که پرسیم آیا عناصری که آزمون اندازه می‌گیرد، همان عناصری هستند که سازه را تشکیل می‌دهند؟ برای مثال، هنگامی که دال در ۱۹۳۵ مقیاس بلوغ اجتماعی واینلند<sup>(۱)</sup> را ابداع کرد، سازه بلوغ اجتماعی را ترکیبی از عناصر وابسته به هم خودیاری، خودفرمانی، حرکت، مشغولیت، ارتباطات و روابط اجتماعی تعریف کرد. آنهایی که اولین نسخه تجدید نظر شده این آزمون را در سالنامه اندازه‌گیریهای روانی بوروس<sup>(۲)</sup> مطالعه کردند، به این نتیجه رسیدند که این عناصر و جوهی از سازه مذکور هستند که باید در یک آزمون بلوغ اجتماعی گنجانده شوند. گاهی در مورد این که چه عناصری یک سازه را تشکیل می‌دهند، توافقی وجود ندارد. برای مثال، اگر کسی تصور کند که سازه هوش، در درجه اول ترکیبی از مهارتهایی است که فرد را قادر می‌سازد تا خود را با یک محیط تحصیلی هماهنگ سازد، بنابراین انتظار خواهد داشت که چنین مهارتهایی در یک آزمون هوش اندازه‌گیری شود. اگر کسی هوش را به عنوان مجموعه‌ای از مهارتها تعریف کند، که با محیطهای دیگر همان مقدار رابطه دارد که با محیط درسی، انتظار نخواهد داشت که مهارتهای خاص مدرسه‌ای در آزمون گنجانده شود.

یک وجه دیگر رویکرد منطقی، بررسی پرسشهای آزمون برای تعیین این نکته است که آیا آنها برای سنجش عناصر موجود در سازه مناسب به نظر می‌رسند یا خیر. برای مثال، در مقیاس واینلند از پدر یا مادر یک کودک شش ساله پرسیده می‌شود آیا آزمودنی از اسکیت، سورتمه یا چهار چرخه استفاده می‌کند (مشغولیت)، بدون کمک دیگری به رختخواب می‌رود (خودیاری)، کلمات ساده را با حروف مجزا می‌نویسد (ارتباطات)، بازیهای ساده روی میزی را انجام می‌دهد (حرکت) و آیا پول در اختیار او گذاشته می‌شود (خودفرمانی)؟ این پرسشها برای اندازه‌گیری عناصر بلوغ اجتماعی مناسب به نظر می‌رسند. اگر آزمون اولیه شامل پرسشهایی در مورد تمایل کودک به غذاهایی خاص یا راست دست بودن او بود، این بار این پرسشها حذف شدند؛ زیرا به طور مستقیم به عناصر موجود در سازه ربط ندارند.

همچنین، داده‌های تجربی نیز به عنوان شواهد گردآوری می‌شوند. از نظر درونی، روابط درون آزمون باید به صورتی باشند که به وسیله سازه پیش‌بینی شده است و از نظر بیرونی، روابط بین

1- Doll, E. A. (1935, 1949, 1965). *Vineland Social Maturity Scale*. Circle Pines, MN: American Guidance Service. The Scale was revised in 1984 by S.S. Aparrow, D.A. Balla, and D.C. Cicchetti as the Vineland Adaptive Behavior Scales.

2- Buros, O.K. (1949). *Mental Measurements Yearbook* (Vol. 4). Lincoln: University of Nebraska Press.

نمره‌های آزمون و سایر مشاهدات باید با سازه سازگار باشند. دال توانست نشان دهد که نمره‌های مقیاس واینلند برای مشغولیت، خودیاری و غیره، به طور مثبت با یکدیگر همبستگی دارند. این مشاهدات یک پشتوانه درونی برای این نظریه که سازه بلوغ اجتماعی شامل عناصر به هم وابسته است، فراهم کرد و شواهدی به دست داد که مقیاس و اینلند در اندازه‌گیری این عناصر به هم وابسته موفق عمل کرده است.

اگر روابط عناصر درون یک آزمون آن چیزی نباشند که سازه پیش‌بینی کرده است، پس یا سازه خود نامناسب است، یا آزمون در اندازه‌گیری عناصر موجود در سازه ناموفق است. برای مثال، ممکن است محققى بخواهد سازه‌ی غریزه اجتماعی - زیستی را اندازه‌گیری کند و ابتدا فرض نماید که سازه مزبور شامل این عناصر به هم وابسته است: (۱) میل به بقا، (۲) میل به تولید مثل، (۳) اشتیاق به انتخاب جفت سالم، (۴) تمایل به از خود گذشتگی برای حفظ بقای فرزندان خود یا خویشان نزدیک. هنگامی که آزمون ساخته و اجرا شد، اگر محقق دریابد که این عناصر به طور مثبت با هم مرتبط نیستند، نتیجه خواهد گرفت که اندازه مزبور فاقد اعتبار است و آزمون یا سازه باید مورد تجدیدنظر قرار گیرند.

نمره‌های یک آزمون باید به شیوه‌ای هماهنگ با سازه، با اندازه‌های بیرونی مرتبط باشند. دال و دیگران نشان داده‌اند که نمره‌های مقیاس واینلند با سن تقویمی، سن عقلی و سنجشهای مستقل بلوغ اجتماعی همبستگی دارند. بنابراین می‌توان گفت که نمره‌های مقیاس واینلند با اندازه‌های بیرونی روابطی را نشان می‌دهند که باید در یک مقیاس بلوغ اجتماعی، که برای اندازه‌گیری این سازه اعتبار دارد، انتظار آن را داشت.

اندازه‌ای که به سازه خاصی مربوط است، باید تا حد امکان مستقل از اندازه‌های سازه‌های دیگر باشد. برای مثال، اگر ما آزمونی تهیه کنیم که برای اندازه‌گیری مهارت‌های مسئله‌گشایی درس حساب طراحی شده باشد و دریابیم که نمره‌های این آزمون همبستگی بسیار بالایی با نمره‌های آزمون قرائت دارند، نتیجه خواهیم گرفت که به جای یک آزمون مسئله‌گشایی حساب، یک آزمون قرائت ساخته‌ایم. به هر روی، اگر ما دو آزمون مسئله‌گشایی حساب در دست داشته باشیم که هر دو با یک آزمون حساب همبستگی داشته باشند ( $r = 0/7$ )، اما اولی با یک آزمون قرائت همبستگی  $0/8$  و دومی با آن آزمون همبستگی  $0/6$  داشته باشد، حکم خواهیم کرد که آزمون دوم اعتبار بیشتری برای اندازه‌گیری مسئله‌گشایی حساب دارد؛ زیرا نسبت به آزمون اول از آزمون قرائت مستقل‌تر است.

## / روشهای مورد استفاده در گردآوری شواهد مرتبط با سازه

هیچ روش منفردی برای گردآوری شواهد در زمینه تفسیر سازه یک آزمون وجود ندارد. هر دلیلی که بر سودمندی یا معنای نمره‌های آزمون، چیزی بیفزاید، به سازه مرتبط است. برخی از متداولترین شیوه‌های گردآوری شواهد مرتبط با سازه در این بخش ارائه می‌شوند.

**همبستگی با سایر اندازه‌ها** یکی از ساده‌ترین شیوه‌ها این است که نشان دهیم نمره‌های حاصل از یک آزمون جدید با نمره‌های حاصل از یک آزمون متداول، که گمان می‌رود اندازه معتبری از سازه مورد نظر است، همبستگی دارند. برای مثال، می‌توان همبستگی آزمون هوشی را که به تازگی ساخته شده است، با یک آزمون هوش کاملاً تثبیت شده، نظیر استانفورد - بینه یا وکسلر محاسبه کرد. اگر این همبستگی بالا باشد، می‌توان فرض کرد که آزمون جدید نیز همان سازه آزمون تثبیت شده را اندازه می‌گیرد.

داده‌های مرتبط با تفسیر سازه نیز از طریق ضریب اعتبار ارائه می‌شوند. نمره‌های یک آزمون استعداد تحصیلی باید همبستگی بالایی با معدل درسی، نمره‌های آزمون پیشرفت تحصیلی، یا رتبه‌بندی آموزگاران در مورد استعداد دانش‌آموزان داشته باشد. یک آزمون روابط فضایی باید موفقیت در ترسیم درس مکانیک را پیش‌بینی کند. تأیید چنین پیش‌بینی‌هایی خود پشوانه‌ای برای حصول اطمینان از این که آزمون مورد نظر، اندازه مناسبی از سازه است، فراهم می‌آورد.

مسبک در مورد استفاده از همگرایی شاخصهای سازه و نیز تمایز پذیری یک سازه از دیگر سازه‌ها صحبت می‌کند.<sup>(۱)</sup> همگرایی یعنی این که اندازه مورد نظر با سایر اندازه‌هایی که فرض می‌شود شاخصهای معتبری از همان سازه باشند، مرتبط است. محقق برای دریافت همگرایی شاخصهای سازه، ابتدا اندازه‌های دیگری را که سازه مورد نظر باید به طور نظریه‌ای با آنها همبستگی داشته باشد، پیدا می‌کند و سپس نشان می‌دهد که آنها با هم چه همبستگی‌ای دارند. یک آزمون استدلال ریاضی باید با نمره‌های درس ریاضی همبستگی داشته باشد. اگر برآستی آن آزمون همبستگی بالایی با نمره‌های ریاضی داشته باشد، این خود شهادتی بر وجود همگرایی خواهد بود.

1- Messick, S. (1989). Validity. In R. Linn (ed.). *Educational Measurement* (pp. 13 - 103). New York: American Council on Education and Macmillan.

به هر روی، همگرایی شاخصها ادله کافی نیست. مسیك به نیاز برای شواهدی که می توان با آن سازه را به طور تجربی از دیگر سازه‌ها متمایز کرد، اشاره می‌کند. برای تعیین افتراق پذیری، محقق در پی شواهدی می‌گردد دال بر این که سازه مورد نظر اساساً با ابزاری که برای اندازه‌گیری سازه‌های دیگر به کار می‌رود، همبستگی ندارد؛ به این معنا که محقق اندازه‌هایی را شناسایی می‌کند که آن سازه اساساً نباید با آنها همبستگی داشته باشد. یک آزمون استدلال ریاضی باید همبستگی بسیار کم و بی‌اهمیتی با یک آزمون قرائت داشته باشد؛ زیرا در آزمون استدلال ریاضی، قرائت یک متغیر نامرتبط است. اگر بین آزمون ریاضی و آزمون قرائت همبستگی کم و بی‌اهمیتی یافت شود، این خود شواهدی بر افتراق پذیری خواهد بود. البته، همبستگی کم یا صفر با هر سازه‌ای، ادله مناسبی نخواهد بود. سازه مورد استفاده باید سازه‌ای باشد که حداقل نمایانگر برخی از جنبه‌های بالقوه سازه مورد بررسی باشد. هر چند انتخاب آزمون قرائت، به عنوان اندازه‌ای که آزمون استدلال ریاضی نباید با آن همبستگی قابل توجهی داشته باشد، منطقی خواهد بود، همبسته کردن نمره‌های ریاضی با پرتاب توپ در فوتبال، برای تهیه شواهد مرتبط با سازه، بی‌معنا خواهد بود. اگر بر روی یک پیوستار متشکل از سازه‌های مرتبط، سازه مورد نظر را به صورت نقطه‌ای مشخص کنیم؛ به نحوی که سازه‌های همگرایی و افتراق نقاط دیگر همان پیوستار باشند، شاید درک مطلب آسانتر شود. برای مثال، انتظار خواهیم داشت که یک اندازه جامعه‌پذیری، همبستگی مثبت با اندازه برونگرایی (همگرایی) و همبستگی منفی با اندازه درونگرایی (افتراق) داشته باشد. خواننده را برای بحث کاملتری درباره نقش همگرایی و افتراق، به عنوان شواهد سازه مورد اندازه‌گیری، به مقاله کلاسیک کمپبل و فیسک ارجاع می‌دهیم.<sup>(۱)</sup>

وجه دیگر رویکرد همبستگی در گردآوری شواهد، تحلیل عاملی است. تحلیل عاملی یک روش آماری برای مطالعه همبستگی داخلی ما بین مجموعه‌ای از نمره‌های آزمون است که به منظور تعیین تعداد عوامل (سازه‌های) لازم برای اثبات وجود همبستگی داخلی انجام می‌گیرد. این روش، همچنین اطلاعاتی در مورد این که عوامل تعیین کننده عملکرد در هر آزمون کدامند و نیز درصد واریانس نمره‌های آزمون که آن عوامل موجب آنها هستند، به دست می‌دهند. محقق کار را با تعداد زیادی از اندازه‌های مختلف آغاز می‌کند، اما با بررسی همبستگی داخلی بین آنها و

1- Campbell, D.T., and D.W. Fiske (1959). Convergent and Discriminant Validation by the Multitrait - multimethod Matrix. *Psychological Bulletin*, 56, 81 - 105.

تعیین این که کدام اندازه‌ها با هم مطابقت (همبستگی) دارند، می‌تواند نتایج زیاد اندازه‌گیریها را به تعداد کمتری از عوامل که واقعاً اندازه‌گیری می‌شوند، کاهش دهد.<sup>(۱)</sup> همبستگی‌های داخلی نه تنها نشان می‌دهند که کدام آزمون‌ها یک عامل همسان را اندازه می‌گیرند، بلکه همچنین مقدار این اندازه‌گیری‌رانی‌نشان می‌دهند. با بررسی آن بخش از محتوای آزمون‌ها که بر یک عامل بار شده‌اند، می‌توان ماهیت سازه‌ای را که قرار است اندازه‌گیری شود، استنباط کرد.<sup>(۲)</sup>



بررسیهای آزمایشی می‌توان چنین فرض کرد که هنگامی که انواع خاصی از تیمارهای آزمایشی اعمال شوند، نمره‌های آزمون تغییر خواهند کرد. برای مثال، در اعتباریابی یک مقیاس اضطراب، می‌توان فرض کرد که اگر افراد در شرایط اضطراب‌زا قرار داده شوند، نمره‌های حاصل از آن مقیاس تغییر خواهند کرد. اگر در یک آزمایش کنترل شده، اضطراب دستکاری شود و نمره‌های حاصل از مقیاس اضطراب به گونه‌ای پیش‌بینی شده تغییر یابند، شواهدی در دست خواهیم داشت که آن مقیاس، اضطراب را اندازه‌گیری می‌کند.

مقایسه نمره‌های گروه‌های تعریف شده محقق می‌تواند گروه‌هایی را مورد استفاده قرار دهد که از قبل مشخص شده است که با هم تفاوت دارند و فرض کند که نمره‌های حاصل از ابزار اندازه‌گیری مورد نظر، آن گروه‌ها را از هم متمایز خواهد کرد. محقق انتظار خواهد داشت که نمره‌های یک آزمون استعداد موسیقی، بین دانشجویانی که اخیراً در مدرسه موسیقی ثبت نام کرده‌اند و یک گروه غیر منتخب از دانشجویان دانشکده، تمایز قایل شود. به همین نحو، اگر بتوان بر مبنای نمره‌های یک آزمون استعداد مکانیکی، مکانیک‌ها و غیر مکانیک‌ها را از هم متمایز کرد، این خود تأییدی بر اعتبار آن آزمون به عنوان اندازه‌ای از استعداد مکانیکی خواهد بود. اگر پرسشنامه‌ای سازگاری روانی را اندازه‌گیری بکند، پس می‌توان انتظار داشت که نمره‌های آن آزمون بین گروه‌هایی که از قبل به عنوان سازگار و روان رنجور شناسایی شده‌اند، تمایز قایل شود. گروه‌های مختلفی که در بررسی اعتبار مورد استفاده قرار می‌گیرند، می‌توانند گروه‌های

۱ - برای مثال مراجعه کنید به:

J. Kim and C. Mueller (1978), *Introduction to Factor Analysis: What It Is and How to Do It*, Beverly Hills: Sage Publications.

۲ - واژه بار شدن (Loading) به معنای بین آزمون و عامل اطلاق می‌شود. بار زیاد نشان می‌دهد که آزمون اندازه خوبی از عامل مورد نظر (سازه) است.



سنی، گروه‌های جنس، گروه‌هایی با مقادیر متفاوت آموزش در حیطه‌های مرتبط با سازه، گروه‌های بهنجار، ناسازگار و غیره باشند. اگر نمره‌های آزمون، تفاوت‌های پیش‌بینی شده را تایید کنند، پس اعتبار آزمون به عنوان یک اندازه از سازه مورد نظر، تایید می‌شود.

**تحلیل درون آزمونی** روش تحلیل درون آزمونی، خود آزمون را بررسی و اطلاعاتی درباره محتوای آزمون، فرایندهای پاسخگویی به سؤالات آزمون و همبستگی بین سؤالات آزمون گردآوری می‌کند. داده‌هایی که از بررسی‌های مرتبط با محتوا به دست می‌آید، ممکن است اطلاعاتی درباره سازه‌ای که آزمون اندازه‌گیری می‌کند، فراهم آورد. اگر جامعه آماری رفتارهایی که آزمون از میان آنها نمونه‌گیری کرده است، مشخص باشد، فرد محقق پیشنی در مورد ماهیت سازه‌ای که قرار است آزمون آن را اندازه‌گیری کند، کسب می‌کند. برای مثال، اگر محقق جامعه رفتاری یک آزمون استدلالی را با توصیف توانایی‌هایی که آزمون از میان آنها نمونه‌گیری کرده است، تعریف کند (نظیر توانایی استدلال با قیاس‌های کمی و کلامی)، درباره اعتبار آن آزمون پیشنی به دست خواهد آورد.

همچنین، محقق می‌تواند فرایندها و مهارت‌های ذهنی را که افراد به هنگام پاسخگویی به سؤالات آزمون مورد استفاده قرار می‌دهند، بررسی کند. برای مثال، می‌توان از دانشجویان خواست که به هنگام پاسخ به یک آزمون استدلال کلامی، «با صدای بلند فکر کنند». چنین شیوه‌ای می‌تواند نشان دهد که آیا آزمون همان توانایی استدلال را، که ادعا می‌کند، اندازه می‌گیرد یا عوامل دیگری نظیر درک واژگان یا درک مطلب را می‌سنجد.

همچنین، محقق باید همگنی محتوای آزمون را به منظور اطمینان یافتن از این که آن آزمون یک صفت یا کیفیت منفرد را اندازه می‌گیرد، بررسی کند. اندازه‌های همسانی درونی، نظیر ضریب آلفا، یا ضریب کودر - ریچاردسون، شواهدی در زمینه همگنی ارائه می‌کنند. این اندازه‌ها در بخش بعدی این فصل مورد بحث قرار خواهند گرفت. اندازه‌های همگنی، شواهدی در مورد سازه به دست می‌دهند؛ زیرا به توصیف این که یک صفت یا سازه منفرد تا چه حد اندازه‌گیری می‌شود، کمک می‌کنند. به هر روی، برای اعتباریابی یک آزمون، داده‌های درون آزمونی هرگز کافی نیستند. ما برای تعیین این که یک آزمون چه چیزی را اندازه می‌گیرد، به داده‌های بیرونی نیاز داریم.

### سهم بررسیهای مرتبط با سازه در اعتباریابی آزمونها

در بررسی اعتباریابی، بررسیهای مرتبط با سازه جامعترین بررسیها هستند؛ زیرا آنها هم ارتباط محتوایی و معرف بودن و هم شواهد مرتبط با ملاک را در بر میگیرند. برای اعتباریابی آزمون رویکرد مرتبط با سازه اهمیت دارد؛ زیرا توجه خود را بر نقش نظریه در ساخت آزمون و نیاز به صورتبندی فرضیهها متمرکز می‌کند، که به عنوان بخشی از فرآیند اعتباریابی می‌توانند مورد بررسی قرار گیرند.

حتی زمانی که یک نظریه مشخص دخیل نباشد، این رویکرد بر پی‌ریزی سازه موردنظر برای یک آزمون خاص در یک چهارچوب مفهومی، که معنای سازه را مشخص می‌کند، آن را از سایر سازه‌ها متمایز می‌کند و نشان می‌دهد که چگونه اندازه‌های سازه باید با دیگر متغیرها مرتبط شوند، تاکید دارد. بنابراین، محقق می‌تواند داده‌هایی از منابع گوناگون گردآوری کند که پشتوانه‌ای برای اعتبار آزمون باشند.

### // مقایسه منابع شواهد اعتبار

انواع شواهد اعتبار یکدیگر را تکمیل می‌کنند، همه آنها اساساً با تفاسیر معتبر معنای نمره‌ها و نحوه استفاده از آن نمره‌ها سروکار دارند. دلیل اصلی شناخت اعتبار به عنوان یک مفهوم یگانه همین است.

فرض کنید آموزگاری می‌خواهد یک آزمون قرائت بسازد تا آن را در مورد دانش‌آموزان کلاس ششم به کار گیرد. این آموزگار چه نوع شواهدی باید گردآوری کند تا تأییدی بر اعتبار بکارگیری این آزمون باشد؟ اگر قرار باشد این آزمون برای انجام استنتاجهایی درباره پیشرفت خواندن دانش‌آموزان به کار رود، آموزگار ابتدا باید در مورد جامعه محتوایی آزمون، که باید از آن نمونه‌گیری شود، تصمیم بگیرد. به طور اخص، کتابهای درسی، مطالب خواندنی خارج از کتب درسی، تمرینهای کلاسی و غیره شناسایی خواهند شد. از این جامعه محتوایی، نمونه‌ها به گونه‌ای انتخاب خواهند شد که آزمون، یک نمونه معرف از تمام محتوا و هدفهای دوره درسی باشد. از آموزگاران دیگر نیز خواسته خواهد شد با توجه به هدف آزمون، در مورد کفایه محتوای آن قضاوت کنند.

در مورد رابطه بین نمره‌های آزمون دانش‌آموزان و عملکرد آنها در یک ملاک مناسب، شواهدی گردآوری خواهد شد. برای تعیین این که آیا بین نمره‌های آزمون و پیشرفت خواندن در

کلاس هفتم همبستگی وجود دارد، یک بررسی پیگیری می‌تواند صورت گیرد. یک همبستگی بالا، شواهدی به دست می‌دهد حاکی از این که آزمون قرائت برای پیش‌بینی پیشرفت خواندن در کلاس هفتم اعتبار دارد. از طریق محاسبه همبستگی نمره‌های آزمون با نمره‌های دانش‌آموزان در کلاس قرائت یا با نمره‌های حاصل از یک آزمون قرائت استاندارد شده معتبر، می‌توان شواهد اضافی را به دست آورد.

آموزگار برای گردآوری شواهد اعتباری مرتبط با سازه، سعی خواهد کرد که تعیین کند آیا فرضیه‌هایی که در مورد ماهیت مهارت در خواندن ساخته است، به وسیله عملکرد دانش‌آموزان در آزمون تأیید می‌شود یا خیر. تفاوت‌های موجود در عملکرد دارندگان نمره بالا، در مقابل دارندگان نمره پایین، مطالعه خواهد شد و برای تعیین این که عملکرد در آزمون با یک نظریه آموزشی خواندن رابطه دارد یا نه، کوشش به عمل خواهد آمد و الی آخر.

روش دیگر، مطالعه در مورد همگرایی نمره‌های قرائت با متغیرهای دیگر خواهد بود. یک اندازه‌همسانی درونی نشان خواهد داد که آیا آزمون، یک سازه منفرد را اندازه‌گرفته است یا خیر. خلاصه‌ای از همه انواع این شواهد روشن خواهد کرد که استنتاج‌هایی که از نمره‌های آزمون قرائت به عمل آمده است، چقدر مناسب است؛ به عبارت دیگر، اعتبار آزمون چقدر است. جدول ۱-۸ انواع شواهد اعتباری را با در نظر گرفتن هدفهای متفاوت آزمون‌ها خلاصه می‌کند.

جدول ۱-۸ اعتباریابی یک آزمون قرائت برای هدفهای متفاوت

هدف	نوع شواهد	سؤالاتی که باید طرح شوند
آزمون پیشرفت تحصیلی در کلاس قرائت ششم	محتوا	آزمون تا چه حد از آنچه که یاد گرفته شده است، خوب نمونه‌گیری می‌کند؟
آزمون استناد برای پیش‌بینی عملکرد در کلاس قرائت هفتم	مرتبط با ملاک: بررسی پیش‌بینی	آزمون تا چه حد پیشرفت در درس قرائت کلاس هفتم را خوب پیش‌بینی می‌کند؟
آزمون تشخیصی برای تشخیص مشکلات خواندن	مرتبط با ملاک: بررسی همزمان	آزمون تا چه حد مشکلات کلی در درس قرائت را خوب تشخیص می‌دهد؟
آزمونی برای اندازه‌گیری درک مطلب	سازه	آزمون تا چه حد درک مطلب را خوب اندازه‌می‌گیرد؟ آیا داده‌ها، فرضیه‌های موجود در مورد درک مطلب را تأیید می‌کنند؟

## // کاربرد مفهوم اعتبار

اگر چه ما اعتبار را به شیوه‌ای کلی به عنوان وحد و اندازه‌ای که یک آزمون، چیزی را که قصد اندازه‌گیری آن را دارد، اندازه می‌گیرد، تعریف کردیم، اما اعتبار، یک ویژگی عمومی، که هر آزمونی باید دارای آن باشد، نیست. اعتبار، خاص ویژه کاری است که می‌خواهیم یک آزمون آن ویژه کار را انجام دهد. یک آزمون یا مقیاس باید تنها با یک هدف ساخته شود. آزمونی که قصد اندازه‌گیری هدفهای متعددی را داشته باشد، معمولاً به هیچیک از آنها به طور کامل دست نخواهد یافت. محقق باید هدف آزمون را بداند و زمینه‌ای را که آزمون در آن مورد استفاده قرار خواهد گرفت، بشناسد تا بتواند اعتبار آزمون را برای آن شرایط خاص بسنجد.

## /// پایایی

پایایی یک ابزار اندازه‌گیری، عبارت است از درجه ثبات آن در اندازه‌گیری هر آنچه که اندازه می‌گیرد. این کیفیت، در هر نوع اندازه‌گیری، یک امر اساسی است. مدیر پستخانه‌ای که متوجه شود ترازوی او وزن بسته‌های پستی را گاهی کم و گاهی زیاد نشان می‌دهد، خیلی زود در صدد تعمیر آن برخواهد آمد. روان‌شناسان و متخصصان تعلیم و تربیت نیز که در تلاش برای اندازه‌گیری خصیصه‌های پیچیده‌ای نظیر استعداد تحصیلی، پیشرفت تحصیلی، انگیزش، اضطراب و نظایر آنها می‌باشند، به همان نسبت دربارهٔ ثبات وسیله اندازه‌گیری خود دل‌نگران هستند. اگر یک آزمون استعداد تحصیلی، هر بار که با آزمودنی یکسانی به کار می‌رود، نتایج بسیار متفاوتی به بار آورد، از نظر روان‌شناسان و متخصصان تعلیم و تربیت بی‌ارزش قلمداد خواهد شد. اشخاصی که این وسایل اندازه‌گیری را به کار می‌برند، باید شیوه‌هایی را برای تعیین این که وسایل اندازه‌گیری آنها تا چه حد باثبات و پایا هستند، شناسایی کنند و مورد استفاده قرار دهند.

## // نظریه پایایی

شناسایی خطای تصادفی اندازه‌گیری و خطای نظام‌مند (سیستماتیک) اندازه‌گیری راهی است برای آن که بین مفهوم پایایی و مفهوم اعتبار تمایز قایل شویم. خطای تصادفی به خطایی اطلاق می‌شود که حاصل احتمال محض باشد. خطای تصادفی اندازه‌گیری، ممکن است به

طریقی غیر قابل پیش‌بینی، نمره هر آزمودنی را افزایش دهد یا کم کند. برای مثال، یکی از اجزای آزمون آمادگی جسمانی پرزیدنت<sup>(۱)</sup> برای دانش‌آموزان دوره ابتدایی، پرتاب توپ بیس‌بال است. به آزمودنیها آموزش داده می‌شود که یک توپ بیس‌بال را تا آنجا که می‌توانند دورتر پرتاب کنند و سپس مسافت این پرتاب اندازه‌گیری می‌شود. اگر چه هدف این آزمون کسب نمره‌ای است که خاص عملکرد هر آزمودنی می‌باشد، اما مطمئناً اگر ما تنها یک آزمودنی داشته باشیم که توپ بیس‌بال را به دفعات متعدد پرتاب کند، در خواهیم یافت که پرتابهای او همه در مسافت یکسانی فرود نمی‌آیند.

فرض کنید از هر دانش‌آموز خواسته می‌شود که در دو روز متوالی، هر روز یک پرتاب انجام دهد. پس از آن، اگر ما دو نمره (مسافت پرتاب شده) را برای هر دانش‌آموز مقایسه کنیم، در خواهیم یافت که آن دو نمره تقریباً هرگز یکسان نیستند. اکثر تفاوتها جزئی خواهند بود، اما برخی از تفاوتها نسبتاً زیاد و تعداد کمی از تفاوتها نیز بسیار زیاد خواهند بود. نتایج پرتابها در آن دو روز بی‌ثبات هستند. پرتاب کردن تنها یک توپ، به عنوان اندازه‌ای از توانایی پرتاب دانش‌آموز کاملاً پایا نیست.

سه نوع تأثیر اتفاقی، یا تصادفی وجود دارد که به بی‌ثباتی بین نمره‌های حاصله در آن دو روز منجر می‌شود:

۱- شاید دانش‌آموز طی آن دو روز واقعاً تغییر کرده باشد. شاید او در یکی از آن دو روز احساس کند که بهتر است. شاید در یکی از آن دو روز دانش‌آموز انگیزه بیشتری داشته یا کمتر خسته باشد. شاید پدر دانش‌آموز، با شنیدن تکلیفی که به فرزندش داده‌اند، آموزش لازم برای پرتاب توپ بیس‌بال را به او داده باشد.

۲- شاید تکلیف خواسته شده از دانش‌آموزان طی دو اندازه‌گیری تغییر کرده باشد. برای مثال، تویی که در روز اول مورد استفاده قرار می‌گیرد، ممکن است سفت باشد، در حالی که در روز دوم خیس و مرطوب باشد. شاید در روز اول ممتحن به دانش‌آموزان اجازه دورخیز بلند بدهد، در حالی که در روز دوم دانش‌آموزان مجاز باشند که فقط دو قدم دورخیز کنند. این تغییرات شاید به برخی از دانش‌آموزان بیشتر از سایرین کمک کند.

۳- نمونه محدود شده رفتاری به نمره ناپایداری منتهی می‌شود. یک نمونه کوچک رفتاری،

موضوع انواع بسیاری از تأثیرات احتمال و اتفاق است. شاید به هنگام پرتاب توپ باد شدید بوزد. شاید دانش‌آموز موقع پرتاب توپ تعادل خود را از دست بدهد، یا شاید هنگامی که توپ را محکم گرفته است، انگشتانش بلغزد.

پایایی، با تأثیر این خطاهای تصادفی اندازه‌گیری بر ثبات نمره‌ها سروکار دارد.

از سوی دیگر، برخی از خطاهای موجود در اندازه‌گیری قابل پیش‌بینی یا نظام‌مند هستند. حالا به مثال پرتاب توپ بیس بال بر می‌گردیم، تصور کنید که دستورالعمل‌های پرتاب توپ به زبان انگلیسی ارائه می‌شود، اما همه آزمودنی‌ها انگلیسی بلد نیستند. نمره‌های آزمودنی‌های غیر انگلیسی زبان به طور نظام‌مند می‌تواند کم شود؛ زیرا آنها درک نمی‌کنند که چه کاری باید بکنند. این خطاهای نظام‌مند اندازه‌گیری، به مسئله اعتبار بر می‌گردند. هر گاه که نمره‌ها، به لحاظ تأثیر هر چیزی به جز آنچه که ما درصدد اندازه‌گیری آن هستیم، به طور نظام‌مند تغییر یابند، اعتبار آزمون کاهش می‌یابد. در این مثال، ما نه تنها توانایی پرتاب توپ بیس بال، بلکه تاحدی، درک زبان انگلیسی را نیز اندازه می‌گیریم.

برای این که بدانیم با مسئله پایایی سروکار داریم یا با اعتبار، باید تعیین کنیم که آیا خطاهای تصادفی را در نظر می‌گیریم یا خطاهای نظام‌مند را. اگر دانش‌آموزان یک کلاس را در آزمون پرتاب توپ بیس بال شرکت می‌دهیم و برای این آزمون دو توپ را در اختیار داریم که یکی سفت و دیگری خیس و مرطوب است و تنها بخت و اقبال تعیین می‌کند که چه کسی با کدام یک از این توپ‌ها پرتاب خود را انجام خواهد داد، تغییرپذیری نمره‌ها، که ناشی از توپ مورد استفاده است، یک مسئله پایایی می‌باشد. تغییرپذیری ناشی از توپ، نمایانگر خطای تصادفی است که بر ثبات اندازه‌گیری تأثیر می‌گذارد.

اما اگر از افراد کلاس بخواهیم که به ترتیب حروف الفبا در آزمون شرکت کنند و آزمون را هم در یک روز بارانی انجام دهیم، به نحوی که توپ بیس بال با هر پرتاب مرطوب‌تر شود، تغییرپذیری نمره‌ها ناشی از خیسی رو به ازدیاد توپ است و ما با مسئله اعتبار سروکار داریم. در این حالت نمره‌های افرادی که در آغاز ترتیب الفبایی قرار دارند، افزایش می‌یابد و نمره‌های افرادی که در انتهای ترتیب الفبایی هستند، کم می‌شود. بنابراین اعتبار نمره‌های پرتاب توپ کاهش می‌یابد؛ زیرا این نمره‌ها نه تنها مهارت پرتاب توپ، بلکه ترتیب الفبایی را نیز منعکس می‌کنند. این مثالی از خطای نظام‌مند است که بر اعتبار اندازه‌گیری تأثیر می‌گذارد.

پایایی با این امر سروکار دارد که ما چیزی را که می‌خواهیم اندازه بگیریم، تا چه حد باثبات اندازه‌گیری می‌کنیم. پایایی با این امر سروکار ندارد که آیا ما همان چیزی را که می‌خواهیم، اندازه‌گیری می‌کنیم یا خیر، این مسئله اعتبار است. امکان دارد که یک ابزار اندازه‌گیری پایا باشد، بی‌آنکه اعتبار داشته باشد؛ هر چند که این ابزار نمی‌تواند معتبر باشد، مگر آن که پایا باشد. برای مثال، فردی می‌تواند تصمیم بگیرد که میزان هوش دیگران را با اندازه‌گیری دور سر آنها تعیین کند. این اندازه‌گیری می‌تواند در زمانهای مختلف، بسیار باثبات (پایا) باشد، اما این روش، اندازه‌گیری معتبری از میزان هوش محسوب نمی‌شود؛ زیرا اندازه‌دورسر یا هیچ یک از ملاکهای هوش همبستگی ندارد و نیز با هیچ یک از نظریه‌های هوش قابل پیش‌بینی نیست.

پایایی از خطاهای تصادفی تأثیر می‌پذیرد که همه عواملی هستند که موجب تفاوت بین نمره‌ها، در اجراهای مکرر یک وسیله اندازه‌گیری می‌شوند.

خطاهای تصادفی از منابع چندی پدید می‌آیند. امکان دارد که خطاها، ذاتی خود وسیله اندازه‌گیری باشند. برای مثال، اگر آزمونی خیلی کوتاه باشد، آن گروه از آزمودنیهایی که به طور اتفاقی پاسخ آن چند سؤال معدود را می‌دانند، نمره‌هایی بیشتر از آنچه که شایسته‌اند، می‌گیرند، در عین حال افرادی که آن چند پاسخ معدود را نمی‌دانند، نمره‌ای کمتر از آنچه سزاوارند، کسب می‌کنند.

برای مثال، اگر قرار باشد آزمونی آگاهی دانش‌آموزان را در مورد اسامی پایتختهای ۵۰ ایالت آمریکا بسنجد، اما فقط دارای پنج سؤال باشد، امکان دارد دانش‌آموزی که فقط نام ده پایتخت را می‌داند، به هر پنج سؤال پاسخ درست بدهد، در حالی که دانش‌آموزی که نام ۴۰ پایتخت را می‌داند، به همه پنج سؤال پاسخ اشتباه بدهد. بخت و اقبال در یک آزمون کوتاه نقش بیشتری دارد تا در یک آزمون طولانی.

اگر آزمونی آن قدر ساده باشد که هر کسی پاسخ اکثر سؤالهای آن را بداند، باز هم نمره‌های نسبی دانش‌آموزان فقط به چند سؤال بستگی پیدا می‌کند و بخت و اقبال یک عامل اصلی خواهد بود. اگر سؤالات آزمون مبهم باشند، آزمودنیهای خوش اقبال به شیوه‌ای پاسخ خواهند داد که مورد نظر ممتحن است، در حالی که آزمودنیهای بد اقبال به شیوه دیگری پاسخ می‌دهند که هر چند آن هم درست است، اما ممتحن پاسخهای آنان را نادرست محسوب می‌کند. شیوه نمره دهی نیز بر پایایی تأثیر می‌گذارد. شیوه‌های دقیق نمره دهی پایایی را افزایش می‌دهند، در حالی

که شیوه‌های نمره دهی مبهم آن را کم می‌کنند.

امکان دارد خطاها، ذاتی نحوه اجرای آزمون باشند. یک فرد بی تجربه ممکن است در اجرا یا نمره دهی آزمون از شیوه‌های استاندارد شده فاصله بگیرد. شاید شرایط انجام آزمون، نظیر نور، گرما و تهویه، بر عملکرد آزمودنیها تأثیر بگذارد، یا شاید دستورالعملهای انجام آزمون مبهم باشند.

همچنین خطاهای دانش‌آموزی نیز وجود دارند، یعنی تغییرات در انگیزه، علاقه، خستگی، شرایط بدنی، اضطراب و دیگر عوامل ذهنی و هیجانی بر نتایج آزمون تأثیر می‌گذارند. اگر در یک آزمون زمان دار، نوک مداد دانش‌آموزی بشکند، مولفه خطا در نمره آزمون افزایش خواهد یافت.

### // معادله‌های پایایی

این امر به طور کلی پذیرفته شده است که همه اندازه‌گیریهای کیفیهای بشری، شامل مقداری خطا است. شیوه‌های پایایی با تعیین میزان بی‌ثباتی نمره‌ها به دلیل خطای تصادفی، سروکار دارند.

هنگامی که آموزگاری از یک دانش‌آموز آزمونی می‌گیرد، نمره‌ای به دست می‌آورد که می‌توان آن را نمره مشاهده شده نامید. اگر این آموزگار همان دانش‌آموز را در موقعیت دیگری با همان وسیله اندازه‌گیری آزمون کند، به احتمال زیاد دقیقاً آن نمره مشاهده شده را به دست نخواهد آورد. نمره مشاهده شده حاوی یک خطای اندازه‌گیری است. بنابراین، می‌توان نتیجه گرفت که نمره هر آزمونی شامل دو مؤلفه است: نمره واقعی به اضافه مقداری خطای اندازه‌گیری. همان طور که قبلاً اشاره شد، این مؤلفه خطا ممکن است نتیجه یک عامل یا ترکیبی از عوامل همبسته با تغییرات درونی آزمودنی در زمانهای مختلف یا مرتبط با اجرای آزمون با آن آزمودنی باشد.

پایایی یک آزمون را از نظر ریاضی می‌توان چنین بیان کرد: بهترین برآورد از این که چه نسبتی از کل واریانس نمره‌های آزمون، واریانس واقعی است. همان طور که در فصل ۵ توضیح داده شد، واریانس شاخصی از پراکندگی یک مجموعه از نمره‌هاست. اگر ما آزمونی را با گروهی از دانش‌آموزان اجرا کنیم، مقداری از پراکندگی (واریانس)، ناشی از تفاوت‌های واقعی بین آن گروه



از دانش آموزان و مقداری از پراکندگی، ناشی از خطاهای اندازه‌گیری است.  
از نظر ریاضی، ایده مؤلفه خطا و مؤلفه واقعی در یک نمره آزمون را می‌توان این‌گونه ارائه داد:

$$X = T + E \quad (۸-۱)$$

$X$  = نمره مشاهده شده

$T$  = مؤلفه واقعی

$E$  = مؤلفه خطا

مؤلفه واقعی را می‌توان نمره‌ای در نظر گرفت که در صورت استفاده از یک وسیله اندازه‌گیری کامل به دست خواهد آمد. مؤلفه خطای اندازه‌گیری می‌تواند مثبت و یا منفی باشد. اگر مثبت باشد، نمره مشاهده شده فرد، برآورد بالایی از نمره واقعی اوست. اگر منفی باشد، نمره واقعی او کم برآورد شده است. از آنجا که فرض شده است که خطای اندازه‌گیری همان اندازه که احتمال دارد مثبت باشد، می‌تواند منفی هم باشد، پس می‌توان نتیجه گرفت که اگر ابزار اندازه‌گیری بی‌نهایت بار با یک آزمودنی اجرا شود، جمع جبری خطاها و میانگین خطاها هر دو صفر خواهند شد. در این شرایط، مؤلفه واقعی، نمره میانگین فرد در بی‌نهایت بار اندازه‌گیری خواهد بود. بنابراین نمره واقعی یک مفهوم نظری است؛ زیرا بی‌نهایت بار اجرای آزمون با یک آزمودنی امکان‌پذیر نیست.

محقق، در یک موقعیت تحقیقاتی معمولی، به ازای هر آزمودنی یک اندازه در دست دارد. به سخن دیگر، محقق با یک مجموعه منفرد از نمره‌های آزمون سروکار دارد. هر نمره مشاهده شده شامل یک مؤلفه نمره واقعی و یک مؤلفه نمره خطا است. به طور ریاضی نشان داده شده است که واریانس نمره‌های مشاهده شده یک گروه بزرگ از آزمودنیها ( $\sigma_x^2$ ) مساوی است با واریانس نمره‌های واقعی آنها ( $\sigma_t^2$ ) به اضافه واریانس خطاهای اندازه‌گیری آنها ( $\sigma_e^2$ )، یا:

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \quad (۸-۲)$$

پایایی را می‌توان به طور نظری، نسبت واریانس نمره واقعی به واریانس نمره مشاهده شده در یک مجموعه از نمره‌ها تعریف کرد. یعنی پایایی برابر است با:

$$r_{xx} = \frac{\sigma_t^2}{\sigma_x^2} \quad (۸-۳)$$

$r_{xx}$  = پایایی آزمون

$\sigma_t^2$  = واریانس نمره‌های واقعی

$\sigma_x^2$  = واریانس نمره‌های مشاهده شده

پایایی، نسبتی از واریانس نمره‌های مشاهده شده است که عاری از خطا باشد. این بیان را می‌توان به صورت این فرمول، که از فرمولهای ۸-۲ و ۸-۳ مشتق شده است، نشان داد.

$$r_{xx} = 1 - \frac{\sigma_e^2}{\sigma_x^2} \quad (۸-۴)$$

دامنه ضریب پایایی ( $r_{xx}$ ) می‌تواند از یک، یعنی هنگامی که هیچ خطایی در اندازه‌گیری وجود ندارد، تا صفر، یعنی هنگامی که تمام اندازه‌گیری خطا است، نوسان داشته باشد (زمانی که هیچ خطایی در اندازه‌گیری نیست،  $\sigma_e^2$  در معادله قبل صفر است و  $r_{xx} = 1$  اگر تمام اندازه‌گیری خطا باشد  $\sigma_e^2 = \sigma_x^2$  و  $r_{xx} = 0$ ). این میزان خطا با میزان انحراف ضریب همبستگی از یک نشان داده می‌شود. بنابراین، هر چه خطا بیشتر باشد، ضریب پایایی از یک کمتر می‌شود و پایایی کمتر است. برعکس، اگر ضریب پایایی به یک نزدیک باشد، وسیله اندازه‌گیری خطای نسبتاً کمتر و پایایی بیشتر دارد.

### // رویکردهایی در مورد پایایی

یک آزمون تا آنجا پایایی دارد که نمره‌هایی که یک فرد در اندازه‌گیریهای مکرر کسب می‌کند، تقریباً یکسان باقی بمانند. دو رویکرد در مورد بیان پایایی یک مجموعه از اندازه‌گیریها وجود دارد. ۱- یک رویکرد به ما می‌گوید که در مجموعه‌ای از اندازه‌گیریهای مکرر از یک فرد، چه مقدار تغییرپذیری را می‌توانیم انتظار داشته باشیم. اگر امکان داشته باشد که فردی را بر روی ۲۰۰ ترازو وزن کنیم، یک توزیع فراوانی از ارقامی خواهیم داشت که وزن او را نشان می‌دهند. این توزیع فراوانی یک مقدار متوسط خواهد داشت که می‌توانیم آن را وزن «واقعی» محسوب کنیم. همچنین، این توزیع یک انحراف معیار خواهد داشت که پراکندگی را نشان می‌دهد. از آنجا که این

انحراف معیار، انحراف معیار «خطاهای» اندازه‌گیری وزن آن فرد است، آن را خطای معیار اندازه‌گیری می‌نامند. ما اغلب در مورد داده‌های روان‌شناختی یا آموزشی، بر روی یک فرد اندازه‌گیریهای مکرر انجام نمی‌دهیم. زمان به ما اجازه اندازه‌گیریهای متعدد بر روی هر فرد را نمی‌دهد. علاوه بر آن، اثرات تمرین و خستگی، که اندازه‌گیریهای متعدد به همراه دارد، بر روی نمره‌ها تأثیر خواهند گذاشت. بنابراین، به جای آن که یک فرد را بارها و بارها اندازه‌گیری کنیم، یک گروه بزرگ و متنوع را در دو زمان مختلف اندازه‌گیری می‌کنیم. سپس با استفاده از دو نمره‌ای که برای هر فرد به دست آورده‌ایم، می‌توانیم برآورد کنیم که پراکندگی نمره‌ها برای فرد متوسط، که به دفعات مورد اندازه‌گیری قرار گرفته است، چه خواهد بود.

۲- همچنین، پایایی اندازه‌گیری به ما می‌گوید هر فرد، تا چه حد رتبه نسبی خود را در داخل گروه حفظ می‌کند. فردی که امروز در یک آزمون بالاترین نمره را کسب می‌کند، بار دیگر که در آن آزمون شرکت می‌کند، باید جزو افرادی باشد که بالاترین نمره‌ها را به دست می‌آورند. در گروه، هر فرد باید در همان رتبه باقی بماند. ما برای تعیین این که افراد تا چه حد رتبه نسبی خود را حفظ کرده‌اند، می‌توانیم ضریب همبستگی بین دو اجرای یک آزمون یکسان را محاسبه کنیم. این ضریب را ضریب پایایی ( $r_{xx}$ ) می‌نامند. بنابراین، پایایی یک اندازه با کم بودن خطای معیار اندازه‌گیری یا زیاد بودن ضریب پایایی مشخص می‌شود.

### // شاخصهای پایایی

پایایی را می‌توان با محاسبه همبستگی نمره‌هایی که هر فرد در دو زمان مختلف، از یک آزمون یا از دو آزمون متفاوت هم ارزش کسب کرده است، برآورد کرد. این شیوه‌ها نیازمند دوبار آزمون کردن هستند. شیوه‌های دیگر، ثبات درونی آزمون را بررسی می‌کنند.

### / پایایی بازآزمایی

یک شیوه بدیهی برای برآورد پایایی یک آزمون، اجرای آن با یک گروه یکسان در دو موقعیت زمانی مختلف و محاسبه همبستگی برای نمره‌های زوج شده است. ضریب همبستگی که از این طریق به دست می‌آید، ضریب پایایی بازآزمایی نامیده می‌شود. برای مثال، امکان دارد که افراد یک کلاس در آزمون آمادگی جسمانی شرکت داده شوند و یک هفته بعد دوباره آن آزمون را تکرار کنند. اگر آزمون پایا باشد، رتبه نسبی هر فرد در اجرای دوم آزمون، به رتبه او در اجرای

اول نزدیک خواهد بود. ضریب پایایی ( $r_{xx}$ ) نزدیک ۱ خواهد بود. هر تغییری در رتبه نسبی افراد، خطا محسوب خواهد شد. اگر آزمون حاوی خطای قابل ملاحظه‌ای باشد،  $r_{xx}$  به صفر نزدیک خواهد بود. همان طور که قبلاً توضیح داده شد، یک ضریب پایایی، برآوردی از نسبت واریانس مشاهده شده در نمره‌های آزمون است که واریانس واقعی باشد. تفاوت بین مقدار ضریب پایایی و ۱، برآورد نااریبی از نسبت واریانس خطا در یک آزمون است. برای مثال، ضریب پایایی بازآزمایی ۰/۸ در یک آزمون آمادگی جسمانی حاکی است که بهترین برآورد ما آن است که ۸۰ درصد واریانس مشاهده شده، واریانس واقعی و ۲۰ درصد، واریانس خطا باشد. ضریب پایایی بازآزمایی به لحاظ آن که نشانگر ثبات نمره‌های آزمودنیها در طول زمان است، گاهی ضریب استواری نیز نامیده می‌شود. این ضریب به ما می‌گوید که آیا می‌توانیم نمره‌ای را که یک فرد در اجرای اول آزمون کسب کرده است، به اجرای دوم همان آزمون تعمیم دهیم یا خیر.

یک ضریب بازآزمایی فرض می‌کند که خصیصه‌ای که آزمون اندازه‌گیری می‌کند، در طول زمان ثابت است، پس هر تغییری در نمره‌ها، در دو موقعیت زمانی مختلف، ناشی از خطای تصادفی است. امکان دارد خطا ناشی از شرایط خود آزمودنیها یا شرایط انجام آزمون باشد. همچنین، ضریب بازآزمایی فرض می‌کند که اثر تمرین یا اثر حافظه وجود ندارد. برای مثال، ممکن است دانش‌آموزان از شرکت در آزمون چیزی یاد بگیرند و بنابراین در اجرای دوم آن به گونه‌ای متفاوت پاسخ بگویند. این اثر تمرینی حاصل از شرکت در آزمون اول، برای همه دانش‌آموزان یکسان نخواهد بود، بنابراین برآورد پایایی را کاهش می‌دهد. اگر فاصله زمانی بین دو آزمون کوتاه باشد، شاید اثر حافظه نیز وجود داشته باشد، یعنی دانش‌آموزان به یاد آورند که بار اول چگونه پاسخ داده‌اند و بار دوم نیز همان را تکرار کنند. این اثر حافظه، برآورد پایایی را افزایش می‌دهد. اثر حافظه را می‌توان با افزایش زمان بین آزمون اول و دوم به نحوی کنترل کرد. از سوی دیگر، اگر زمان بین دو آزمون خیلی طولانی باشد، شاید مشکل یادگیری افتراقی بروز کند، یعنی دانش‌آموزان در فاصله بین دو آزمون، همه به یک اندازه یاد نخواهند گرفت، و این خود بر ضریب پایایی اثر خواهد گذارد.

به دلیل مشکلاتی که ذکر شد، شیوه بازآزمایی معمولاً برای آزمونهای حیطة شناختی مناسب نیست. استفاده از این شیوه در مدارس تا حد زیادی به اندازه‌گیری آمادگی جسمانی و مهارت ورزشی محدود شده است.

## / پایایی «صورت‌های هم ارز»

روش «صورت‌های هم ارز» برآورد پایایی، که همچنین روش «شکل‌های جانشین» یا «شکل‌های هم‌تا» نیز نامیده می‌شود، هنگامی مورد استفاده قرار می‌گیرد که احتمال یادآوری پاسخ‌های قبلی به پرسش‌های آزمون وجود داشته باشد. در اینجا به جای تعیین همبستگی نتایج دو اجرای یک آزمون با گروه یکسان، پژوهشگر نتایج صورت‌های هم ارز آزمون را که با یک گروه اجرا شده است، همبسته می‌کند. اگر دو شکل آزمون اساساً در یک زمان (به طور متوالی) اجرا شود، ضریب پایایی حاصله ضریب هم ارزی نامیده می‌شود. این اندازه، بازتابی از تغییرات عملکردی از یک مجموعه پرسش‌های خاص به یک مجموعه دیگر است.

اگر آزمودنی‌ها با یک شکل آزمون در یک موقعیت و با شکل دیگر آن در موقعیتی دیگر آزمون شوند و نمره‌های آنها از دو شکل آزمون با هم همبسته گردند، ضریب حاصله، ضریب استواری و هم ارزی نامیده می‌شود. این ضریب، دو جنبه از پایایی آزمون را منعکس می‌کند: تغییرات در عملکرد از یک زمان به زمان دیگر و همچنین تغییرات از یک شکل آزمون به شکل دیگر آن. این ضریب پرزحمت‌ترین و موثکافانه‌ترین اندازه موجود برای تعیین پایایی یک آزمون است.

مسئله اساسی در این روش برآورد پایایی، اشکال در ساختن دو شکل واقعاً هم ارز از یک آزمون است. دو شکل آزمون باید هم ارز باشند وگرنه تغییر نمره‌ها از یک شکل آزمون به شکل دیگر آن نمی‌تواند واریانس خطا محسوب شود. صورت‌های هم ارز یک آزمون، آزمونهاهایی هستند که به طور مستقل ساخته می‌شوند و باید مشخصات یکسانی داشته باشند، یعنی از نظر تعداد پرسش‌ها، شکل، دستورالعمل‌ها، محدودیت زمانی، قطع و اندازه، محتوا، دامنه و سطح دشواری همسان باشند، اما پرسش‌های واقعی آنها یکسان نباشد.

روش صورت‌های هم ارز هنگامی توصیه می‌شود که کسی بخواهد از مشکل یادآوری یا اثر تمرین دوری بجوید و همچنین در مواردی که تعداد زیادی پرسش آزمون در دسترس داشته باشد که بتواند از میان آنها دو نمونه هم‌ارز را انتخاب کند. باور کلی بر این است که شیوه صورت‌های هم ارز بهترین برآورد اندازه‌های تحصیلی و روان‌شناختی را فراهم می‌آورد.

## // اندازه‌های ثبات درونی پایایی

شیوه‌های دیگر پایایی برای آن طراحی شده‌اند که تعیین کنند آیا همه سؤالات آزمون یک چیز واحد را اندازه‌گیری می‌کنند یا خیر. اینها را شیوه‌های ثبات درونی می‌نامند و تنها به یکبار اجرای آزمون نیاز دارند.

## / پایایی دو نیمه سازی

اولین شیوه از این شیوه‌ها، که به نام دو نیمه سازی شناخته شده است، آزمون را به طور مصنوعی به دو نیمه تقسیم و همبستگی بین نمره‌های هر فرد را در آن دو نیمه محاسبه می‌کند. آزمون با یک گروه از افراد اجرا می‌شود و سپس سؤالات آزمون به دو نیمه تقسیم می‌شوند. نمره‌های هر فرد در هر یک از دو نیمه به دست می‌آید و یک ضریب همبستگی محاسبه می‌شود. این ضریب پایایی دو نیمه سازی شبیه ضریب هم‌ارزی است؛ زیرا نوسانهای یک نمونه از سؤالات را به نمونه دیگر منعکس می‌کند. اگر هر آزمودنی رتبه کاملاً مشابهی را در دو نیمه آزمون به دست آورد، آزمون، پایایی زیادی دارد. اگر همسانی رتبه‌ها کم باشد، پایایی نیز کم است. این روش تنها به یک شکل از آزمون نیاز دارد، به گذشت زمان احتیاجی نیست و آزمودنی به هنگام پاسخگویی به دو نیمه آزمون، تحت تأثیر عوامل جسمی و ذهنی یکسانی عمل می‌کند.

اشکال این شیوه در تقسیم آزمون به دو نیمه قابل مقایسه است. اگر سطح دشواری هر پرسش از طریق تحلیل پرسشها تعیین شود، می‌توان پرسشها را بر مبنای دشواری یکسان و شباهت محتوا در دو گروه جای داد. به هر ترتیب، رایج‌ترین روش، تعیین همبستگی بین نمره پرسشهای فرد و نمره پرسشهای زوج آزمون است.

ضریب همبستگی که بین دو نیمه آزمون محاسبه می‌شود، به طور نظام‌مند یک برآورد نقصانی از پایایی کل آزمون به دست می‌دهد. در شرایط مساوی، آزمونها طویل‌تر پایاتر از آزمونها کوتاه‌تر هستند. بنابراین، در یک آزمون ۱۰۰ پرسشی، همبستگی بین ۵۰ پرسش فرد و ۵۰ پرسش زوج، برآورد پایایی برای یک آزمون ۵۰ پرسشی است، نه یک آزمون ۱۰۰ پرسشی. برای تبدیل همبستگی دو نیمه سازی به یک برآورد پایایی مناسب برای کل آزمون، فرمول پیشگویی اسپیرمن - براون به کار می‌رود:

$$r_{xx} = \frac{2r_{1/2 \ 1/2}}{1 + r_{1/2 \ 1/2}} \quad (۸-۵)$$

$r_{xx}$  برآورد پایایی کل آزمون =

همبستگی ۲ پیرسون بین دو نیمه آزمون =  $r_{1/2 \ 1/2}$

برای مثال، اگر بین دو نیمه یک آزمون ضریب همبستگی ۰/۶۵ را بدست آوریم، با استفاده از فرمول اسپیرمن - براون، پایایی برآورد شده کل آزمون چنین خواهد بود:

$$r_{xx} = \frac{(2)(0/65)}{1 + 0/65} = 0/79$$

شیوه اسپیرمن - براون بر این مفروضه مبتنی است که دو نیمه آزمون همتا هستند. از آنجا که این مفروضه به ندرت کاملاً درست از آب در می آید، در عمل روش دو نیمه سازی در مقایسه با روشهای بازآزمایی و صورتهای هم ارز، به برآورد اضافی پایایی گرایش دارد. به هنگام ارزشیابی پایایی آزمونهای رقیب، باید این نکته را به خاطر داشت.

پایایی دو نیمه سازی، هنگامی که باید از نوسانهای زمانی اجتناب شود و موقعی که آزمون نسبتاً طویل باشد، روش مناسبی است. برای آزمونهای کوتاه، روشهای دیگر نظیر بازآزمایی و صورتهای هم ارز، مناسب تر است.

روش دو نیمه سازی برای استفاده در آزمونهای سرعت مناسب نیست؛ زیرا در چنین آزمونهایی، به طور نادرست، ضریب هم ارزی بالایی به بار می آورد. آزمون سرعت، آزمونی است که تعمداً پرسشهای آسانی دارد و نمره ها بیشتر به سرعتی که آزمودنی می تواند پرسشها را بخواند و پاسخ خود را ارائه دهد، بستگی دارد. اشتباهات جزئی هستند و اکثر پاسخها، تا آنجا که مدت آزمون به آزمودنی اجازه پاسخگویی می دهد، درست می باشند. اگر دانش آموزی به ۵۰ پرسش پاسخ گفته باشد، نمره دو نیمه سازی او احتمالاً ۲۵-۲۵ خواهد بود، اگر دانش آموز دیگری ۶۰ پرسش را علامت زده باشد، نمره دو نیمه سازی او احتمالاً ۳۰-۳۰ خواهد بود و الی آخر. از آنجا که نمره های افراد در پرسشهای فرد و زوج بسیار همانند هستند، تغییرات درون فردی به حداقل می رسد و همبستگی بین نیمه ها تقریباً کامل خواهد بود. برای آزمونهای سرعت، استفاده از روشهای دیگر توصیه می شود.

## اندازه‌های همگنی

سایر اندازه‌های ثبات درونی پایایی نیازی به دو نیمه سازی آزمون و نمره دهی هر یک از دو نیمه، به تنهایی، ندارند. این شیوه‌ها ثبات بین سؤالی، یا همگنی سؤالیهای پرسشنامه را می‌سنجند. این شیوه‌ها دو منبع خطا را منعکس می‌کنند: (۱) نمونه‌گیری محتوایی، همانند دو نیمه‌سازی و (۲) ناهمگنی حیطه رفتاری، که از آن نمونه‌گیری شده است. هر چه این حیطه ناهمگن‌تر باشد، ثبات بین سؤالی کمتر است، برعکس، هر چه این حیطه همگن‌تر باشد، ثبات بین سؤالی بیشتر است.

روشهای کودر - ریچاردسون احتمالاً مشهورترین شاخص همگنی، فرمول ۲۰ کودر - ریچاردسون است، که بر نسبت پاسخهای درست و نادرست به سؤالات آزمون پی‌ریزی شده است.<sup>(۱)</sup>

$$r_{xx} = \frac{K}{K-1} \left( \frac{S_x^2 - \sum pq}{S_x^2} \right) \quad K - R_{20} \quad (۸-۶)$$

که در آن

$K$  = تعداد سؤالات آزمون

$S_x^2$  = واریانس نمره‌ها در کل آزمون (مربع انحراف معیار)

$p$  = نسبت پاسخهای درست در یک سؤال

$q$  = نسبت پاسخهای نادرست در همان سؤال

حاصلضرب  $pq$  برای هر سؤال محاسبه می‌شود و حاصلضربهای همه سؤالات با هم جمع می‌شوند تا  $\sum pq$  به دست آید. فرمول ۲۰ کودر - ریچاردسون در مورد آزمونهایی که سؤالیهای آنان به صورت دو مقوله‌ای، یعنی درست و نادرست، نمره‌گذاری می‌شوند، کاربرد دارد. بسیاری از روشهای ماشینی نمره‌دهی آزمون‌ها به طور معمول ضریب  $K - R_{20}$  را به همراه ضریب دو نیمه سازی ارائه می‌کنند.

یک فرمول دیگر (کودر - ریچاردسون ۲۱) از لحاظ محاسباتی ساده‌تر است، اما فرض آن بر این است که همه سؤالات آزمون از نظر دشواری یکسان هستند. این فرض اغلب غیر واقعی است.



$$r_{xx} = \frac{K S_x \bar{X} (K - \bar{X})}{S_x (K - 1)} \quad K-R \quad 21(8-7)$$

$r_{xx}$  = پایایی کل آزمون

$K$  = تعداد پرسشهای یک آزمون

$S_x^2$  = واریانس نمره‌ها

$\bar{X}$  = میانگین نمره‌ها

در میان همه روشهای برآورد پایایی، این روش کمترین زمان را صرف می‌کند. روش کودر - ریچاردسون تنها به یک اجرای آزمون نیاز دارد و اطلاعات سهل الوصول را به کار می‌گیرد. بدین گونه، می‌توان آن را برای استفاده کلاسی به آموزگاران توصیه کرد. برای مثال، فرض کنید آموزگاری یک آزمون ۵۰ پرسشی را به دانش آموزان کلاس خود داده و میانگین آن را ۴۰ و انحراف معیار آن را ۶ محاسبه کرده است. او با کاربرد فرمول ۸-۲ می‌تواند پایایی را این چنین برآورد کند:

$$r_{xx} = \frac{(50)6^2 - 40(50 - 40)}{6^2(50 - 1)} = \frac{1800 - 400}{1764} = 0.79$$

از آنجا که روشهای کودر - ریچاردسون بر هم ارزی همه پرسشهای آزمون تأکید دارد، بویژه هنگامی مناسب هستند که هدف آزمون، اندازه‌گیری یک ویژگی منفرد باشد. زمانی که یک آزمون قصد اندازه‌گیری ویژگیهای متعددی را داشته باشد، معمولاً برآورد پایایی کودر - ریچاردسون پایین‌تر از پایایی برآورد شده بر مبنای روشهای همبستگی خواهد بود.

از راه استدلال قیاسی نشان داده شده است که پایایی کودر - ریچاردسون برای هر آزمودنی برابر است با میانگین برآوردهای پایایی دو نیمه‌سازی، که برای همه گونه‌های ممکن دو نیمه‌سازی آن آزمون محاسبه شده باشد. این امر به توصیف رابطه بین دو روش کمک می‌کند. اگر دشواری آزمونی یکنواخت باشد و یک ویژگی منفرد را اندازه بگیرد، به هر شکلی که آزمون را دو نیمه کنیم، احتمالاً نیمه‌نمره‌های مشابهی را به دست خواهیم آورد. بنابراین، روشهای اسپیرمن - براون و کودر - ریچاردسون برآوردهای مشابهی حاصل خواهند کرد. اگر پرسشهای یک آزمون از نظر دشواری با هم متفاوت باشند و آزمون ویژگیهای متعددی را اندازه بگیرد،

انتظار می‌رود برآورد کودر - ریچاردسون پایین‌تر از برآورد دو نیمه‌سازی باشد. برای مثال، فرض کنید یک آزمون مهارت‌های منشی‌گری شامل مواد ماشین‌نویسی، تندنویسی، درست نویسی و دستور زبان انگلیسی است. با کاربرد روش دو نیمه‌سازی، سازنده آزمون، به هر نیمه آن به تعداد مساوی پرسش تخصیص خواهد داد. اگر آزمون ترکیب این مهارت‌ها را به خوبی اندازه بگیرد، پایایی دو نیمه‌سازی آن بالا خواهد بود. روش کودر - ریچاردسون، که میزان هم‌ارزی پرسش‌ها را با یکدیگر می‌سنجد، به طور قابل ملاحظه‌ای برآورد پایایی پایین‌تری حاصل خواهد کرد.

ضریب آلفا یک اندازه‌پراستفاده همگنی، ضریب آلفای کرونباخ<sup>(۱)</sup> است. فرمول ضریب آلفا عبارت است از:

$$\alpha \text{ یا } r_{xx} = \left( \frac{K}{K-1} \right) \left( \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{K}}{\sum x_i^2} \right) \quad (8-8)$$

K = تعداد سؤالات آزمون

$\sum x_i^2$  = مجموع واریانس نمره سؤال‌ها

$\bar{x}^2$  = واریانس نمره آزمون (تمام K سؤال)

فرمول آلفا شبیه  $R^2 - K$  است، به جز آن‌که به جای  $\sum pq$  مجموع واریانس‌های نمره‌های سؤال‌ها، یعنی  $\sum x_i^2$  قرار می‌گیرد. برای محاسبه آن، ابتدا واریانس همه نمره‌ها برای هر سؤال تعیین می‌شود و سپس این واریانس‌ها با هم جمع می‌شوند تا  $\sum x_i^2$  به دست آید. آلفا و  $R^2 - K$  برای سؤالاتی که به صورت دو مقوله‌ای نمره‌گذاری شده‌اند، برابر هستند.

از آلفای کرونباخ هنگامی استفاده می‌شود که آزمون ما دارای سؤالاتی باشد که نمره آنها چند گانه است؛ نظیر مقیاس‌های نگرش یا آزمون‌های انشایی. برای مثال، در یک مقیاس نگرش لیکرت، آزمودنی ممکن است برحسب گزینه‌ای که انتخاب می‌کند، نمره‌ای از ۱ تا ۵ به دست آورد. همچنین در آزمون‌های انشایی امکان دارد نمره‌های متفاوتی به هر پاسخ اختصاص یابد. بسیاری از برنامه‌های کامپیوتری پایایی، نظیر برنامه‌ای که در SPSS موجود است، ضریب آلفا را به عنوان شاخص پایایی ارائه می‌کنند.

اگر سؤالات آزمون ناهمگن باشند، یعنی بیش از یک صفت یا ویژگی را اندازه‌گیری کنند، شاخص پایایی کم خواهد شد، چه از طریق ضریب آلفا و چه از طریق  $K - R_{20}$  محاسبه شود. علاوه بر آن، این فرمولها برای آزمونهای زمان‌دار مناسب نیستند؛ زیرا واریانس‌های سؤالاها، تنها در صورتی دقیق خواهند بود که همه آزمودنیها به سؤال پاسخ گفته باشند.

جدول ۸-۲ خلاصه‌ای از انواع مختلف ضرایب پایایی را که بر حسب تعداد صورتهای آزمون و تعداد اجراهای لازم آزمون مرتب شده‌اند، ارائه می‌کند.

جدول ۸-۲ خلاصه ضرایب پایایی

تعداد اجراهای لازم آزمون	تعداد صورتهای لازم آزمون	
	یک	دو
یک	دو نیمه سازی / $K - R_{20}$ آلفای کرونباخ	صورتهای هم ارز (بدون گذشت زمان)
دو	بازآزمایی	صورتهای هم ارز (باگذشت زمان)

### / تفسیر ضریب پایایی

اکثر سازندگان آزمون و محققان اگر ضریب پایایی  $0/9$  یا بیشتر را به دست آورند، احساس رضایت می‌کنند، اما با ضریب کمتر از  $0/7$  ناراضی می‌شوند. به هر ترتیب، تفسیر این ضریب باید بر مبنای ملاحظات صحت پذیرد. عوامل مشخصی وجود دارند که بر ضریب پایایی تأثیر می‌گذارند و این عوامل باید در نظر گرفته شوند و گرنه هر تفسیری از پایایی سطحی خواهد بود. پایایی یک آزمون تا حدی تابعی از طول آزمون است. هر چه آزمون طولی‌تر باشد، پایایی آن بیشتر است. معمولاً یک آزمون شامل تعدادی پرسش نمونه است که به طور نظریه‌ای از یک مجموعه پرسشهای آزمون بیرون کشیده شده‌اند. براساس مطالعات خود درباره نمونه‌گیری می‌دانیم که هر چه اندازه نمونه بزرگتر باشد، انتظار می‌رود که آن نمونه بیشتر معرف جامعه آماری خود باشد. این امر در مورد آزمونها نیز صدق می‌کند. اگر امکان استفاده از تمام مجموعه

پرسشها وجود داشته باشد، نمره فردی که در آزمون شرکت می‌کند، نمره واقعی او خواهد بود. مجموعه پرسشهای نظری شامل تعداد نامحدودی سؤال است و کاربرد آن یک امکان عملی نیست. بنابراین، پژوهشگر آزمونی می‌سازد که نمونه‌ای از چنین مجموعه نظری است. طول این آزمون هر چه بیشتر باشد (یعنی، تعداد پرسشهایی که در آن گنجانده شده است، بیشتر باشد)، بیشتر معرف نمره واقعی فردی خواهد بود که در آزمون شرکت می‌کند. از آنجا که پایایی به آن مقداری که آزمون معرف نمره واقعی افراد باشد، اطلاق می‌شود، هر چه آزمون طویل‌تر باشد، پایایی آن بیشتر است، با این شرط که همه پرسشهای آزمون به مجموعه پرسشها تعلق داشته باشند.

پایایی تا حدی تابع ناهمگنی گروه است. ضریب پایایی با افزایش گستردگی یا ناهمگنی آزمودنی‌هایی که در آزمون شرکت می‌کنند، افزایش می‌یابد. برعکس، هر چه گروه، نسبت به ویژگی‌ای که اندازه‌گیری می‌شود، همگن‌تر باشد، ضریب پایایی کمتر خواهد بود. یکی از توصیفهای پایایی، میزان توانایی در تعیین جایگاه افراد نسبت به یکدیگر در یک گروه، بر مبنای ویژگیها یا کیفیتهای خاص است. چنین جایگزینی‌هایی، به هنگام سروکار داشتن با اندازه‌هایی که در یک دامنه وسیع قرار می‌گیرند، نسبت به اندازه‌هایی که در یک دامنه کوچک قرار می‌گیرند، آسان‌تر است. اگر بخواهیم جایگاه کودکانی را که دامنه سنی آنها از ۵ تا ۱۵ سال است، بر مبنای وزن در یک توزیع تعیین کنیم، نیاز به وسیله اندازه‌گیری حساس نداریم. در واقع این جایگزینی بدون استفاده از وسیله اندازه‌گیری با کمی دقت امکان پذیر است. به هر ترتیب، اگر همه افراد توزیع که باید با هم مقایسه و تعیین جایگاه شوند، پنج ساله باشند، برای انجام جایگزینی به یک وسیله اندازه‌گیری حساس نیاز داریم. بنابراین، ناهمگنی گروهی که وسیله اندازه‌گیری در مورد آنها به کار می‌رود، عاملی است که بر پایایی آن وسیله تأثیر می‌گذارد. هر چه گروه ناهمگن‌تر باشد، ضریب پایایی بالاتر خواهد بود.

به هنگام انتخاب یک آزمون استاندارد شده باید این واقعیت را در نظر داشت. امکان دارد ناشر آزمون بر مبنای نمونه‌ای با دامنه توانایی وسیع، ضریب پایایی بالایی را گزارش کند. به هر ترتیب، هنگامی که پژوهشگر این آزمون را با گروهی که دامنه توانایی بسیار محدودتری دارد، اجرا می‌کند، پایایی پایین‌تر خواهد بود.

پایایی یک آزمون تا حدی تابعی از توانایی افرادی است که در آن شرکت می‌کنند. شاید

آزمونی در یک سطح توانایی، پایا و در سطحی دیگر، ناپایا باشد. پرسشهای یک آزمون، ممکن است دشوار و فراتر از سطح توانایی افراد شرکت کننده باشد، یا ممکن است برای اکثر آزمودنیها آسان باشد. این سطح دشواری بر پایایی آزمون تأثیر می‌گذارد. زمانی که آزمونی دشوار است، آزمودنیها به اکثر پرسشها پاسخ حدسی می‌دهند و ضریب پایایی پایینی به دست خواهد آمد. هنگامی که پرسشها آسان باشد، همه آزمودنیها به اکثر آنها پاسخ درست می‌دهند و تنها چند پرسش دشوار، آزمودنیها را از یکدیگر متمایز می‌کنند. بنابراین پایایی پایینی را انتظار خواهیم داشت.

هیچ قانون ساده‌ای وجود ندارد که بتوان تعیین کرد یک آزمون چقدر باید دشوار یا آسان باشد. این امر به نوع آزمون، هدف و جامعه آماری مورد نظر بستگی دارد. پایایی تا حدی تابعی از روش خاصی است که برای برآورد آن به کار برده شده است. استفاده از روشهای متفاوت برای برآورد پایایی آزمونها، ضرایب پایایی متفاوتی را نتیجه می‌دهد. روش صورتهای هم ارز در مقایسه با روشهای بازآزمایی و دو نیمه سازی، برآورد پایایی کمتری به دست می‌دهد. این امر به دلیل آن است که در روش صورتهای هم ارز، نوسان شکل به شکل، همانند نوسان زمان به زمان، وجود دارد. از سوی دیگر، روش دو نیمه سازی، از روشهای جایگزین خود، ضریب پایایی بالاتری را نتیجه می‌دهد. دلیل این امر آن است که در هر آزمونی، سرعت تا حدی نقش دارد و ضریب پایایی به همان میزان بر آورد اضافی دارد. بنابراین، در ارزشیابی پایایی یک آزمون، اگر پایایی‌های گزارش شده مشابه باشند، ارجحیت به آزمونی داده می‌شود که ضریب پایایی آن با روش صورتهای هم ارز برآورد شده باشد. همین نتیجه‌گیری در مقایسه پایایی بازآزمایی با دو نیمه‌سازی صدق می‌کند. در میان دو ضریب مشابه که از روشهای بازآزمایی و دو نیمه سازی به دست آمده‌اند، ضریب حاصل از روش بازآزمایی رضایتبخش‌تر است.

### / پایایی ارزیابها

نوع دیگری از پایایی وجود دارد که در آن دسته از وسایل اندازه‌گیری که به رتبه‌بندی یا مشاهده آزمودنیها به وسیله افراد دیگر نیاز دارند، اهمیت پیدا می‌کند. پایایی ارزیابها یا مشاهده‌گران، شاخصی است از این که رتبه‌بندی‌ای که داوران/مشاهده‌گران مختلف از یک رفتار

یکسان ارائه می‌کنند، تا چه حد مشابهت دارند. محقق باید نشان دهد که رتبه‌بندیهای او تحت تأثیر ارزشها، نگرشها و ویژگیهای شخصیتی داوران قرار ندارد. شیوه سنجش این نوع پایایی آن است که دو یا چند رتبه‌بندی کننده، رفتار یکسانی را مشاهده یا رتبه‌بندی کنند. سپس برای تعیین میزان توافق داوران، همبستگی بین رتبه‌بندیهای آنان محاسبه می‌شود، ضریب همبستگی حاصله، شاخص پایایی ارزیابها است. با فرض این که رفتارهای مورد نظر کاملاً تعریف شده‌اند و داوران نیز بخوبی آموزش دیده‌اند، ضریب همبستگی (پایایی) باید مثبت و بسیار زیاد باشد.

### // خطای معیار اندازه‌گیری

همان گونه که قبلاً در این فصل توضیح داده شد، پایایی، استواری، یا قابلیت اطمینان یک آزمون را می‌توان برحسب خطای معیار اندازه‌گیری نیز بیان کرد. خطای معیار اندازه‌گیری، برآوردی را از دامنه تغییر پذیری در مجموعه‌ای از اندازه‌گیریهای مکرر یک ویژگی ارائه می‌کند. اگر به مثال پرتاب توپ بیس بال برگردیم، انتظار خواهیم داشت که در اجراهای متعدد آزمون، از روی اتفاق، نمره‌های متفاوتی برای یک فرد به دست آوریم. ما یک توزیع فراوانی از نمره‌ها خواهیم داشت. این توزیع فراوانی یک میانگین دارد که بهترین تقریب از نمره واقعی است. این توزیع، همچنین یک انحراف معیار دارد که میزان تغییرات نمره‌ها را نشان می‌دهد. از آنجا که این انحراف معیار، انحراف معیار خطاهای اندازه‌گیری است، خطای معیار اندازه‌گیری خوانده می‌شود. اگر یک نمودار چند ضلعی فراوانی که توزیع این نمره‌ها را نشان می‌دهد، بسازیم، شکل آن تقریباً شبیه منحنی نرمال خواهد شد. خطاهای اندازه‌گیری به صورت نرمال توزیع می‌شوند، ممکن است خطاهای کوچک زیادی در آن باشند، اما چند خطای بزرگ نیز دارد. انحراف معیار این توزیع خطا (خطای معیار اندازه‌گیری  $S_M$ ) به ما کمک می‌کند که به هنگام استفاده از آزمون، دفعات وقوع خطاها را برای یک اندازه معین نمونه برآورد کنیم.

در بسیاری از موقعیتهای اندازه‌گیریهای مکرر نخواهیم داشت، اما می‌توانیم برآوردی از خطای معیار اندازه‌گیری را با استفاده از ضریب پایایی به دست آوریم:

$$S_M = S_x \sqrt{1 - r_{xx}} \quad (۸-۹)$$

$S_M$  = خطای معیار اندازه‌گیری

$S_x$  = انحراف معیار نمره‌های آزمون

$r_{xx}$  = ضریب پایایی

اگر آزمون هوشی، ضریب پایایی  $0/96$  و انحراف معیار  $15$  داشته باشد، پس

$$S_M = 15\sqrt{1 - 0/96} = 15\sqrt{0/04} = 3$$

خطای معیار اندازه‌گیری شاخصی از تغییرات مورد انتظار نمره‌های حاصله، حول نمره واقعی است.  $S_M$  را می‌توان به عنوان انحراف معیار نمره خطا، که در نمره مشاهده شده وجود دارد و یا همانند هر انحراف معیاری تفسیر کرد. اگر ما نمره دانش‌آموزی را داشته باشیم، از  $S_M$  برای تعیین دامنه‌ای از مقادیر نمره‌ها، که با احتمال معین نمره واقعی را در برخواهند داشت، استفاده می‌کنیم. این دامنه نمره‌ها را فاصله اطمینان می‌نامیم. اگر بتوان فرض کرد که خطاهای اندازه‌گیری به صورت نرمال حول یک نمره معین و به طور مساوی در طول دامنه نمره‌ها توزیع می‌شوند، می‌توان  $68$  درصد اطمینان داشت که نمره واقعی در داخل یک  $S_M$  از دو طرف نمره مشاهده شده قرار خواهد داشت. برای مثال، اگر یک آزمودنی در آزمون هوشی که خطای معیار اندازه‌گیری آن  $3$  است، نمره مشاهده شده  $110$  را گرفته باشد، می‌توان استنتاج کرد که  $7/68$  احتمال وجود دارد که نمره واقعی او بین  $107$  و  $113$  باشد، یا ما در سطح  $95$  درصد اطمینان می‌توانیم بگوییم که نمره واقعی در  $S_M \pm 1/96$  (بین  $104$  و  $116$ ) قرار خواهد گرفت. ما همچنین برای تعیین این امر که در بازآزمایی یک فرد چه مقدار تغییرپذیری را می‌توانیم انتظار داشته باشیم، از خطای معیار اندازه‌گیری استفاده می‌کنیم. اگر این امکان وجود داشته باشد که یک آزمودنی را با یک آزمون هوش، چندین بار آزمون کنیم، می‌توانیم انتظار داشته باشیم که حدود دو سوم نمره‌های او در داخل یک دامنه  $6$  نمره‌ای و  $95$  درصد نمره‌های او در یک دامنه  $12$  نمره‌ای قرار بگیرد.

خطای معیار اندازه‌گیری و ضریب پایایی، روش‌های جایگزین‌های برای بیان پایایی آزمون هستند. با مقدار خطای معیار اندازه‌گیری و ضریب پایایی نشان داده می‌شود که برآورد نمره واقعی هر نمره مشاهده شده‌ای چقدر دقیق است. محقق به عنوان نشانه‌ای از پایایی یک آزمون، در پی خطای معیار اندازه‌گیری کم، یا ضریب پایایی زیاد است.

توصیه می‌کنیم که از خطای معیار اندازه‌گیری برای تفسیر نمره‌های افراد و از ضریب پایایی برای مقایسه همسانی آزمونهای مختلف استفاده شود.

## / پایایی آزمونهای ملاکی

ابداع روشهای رضایتبخش برای سنجش پایایی آزمونهای ملاکی دشوارتر بوده است. به خاطر دارید که آزمونهای ملاکی برای تعیین موقعیت فرد، با توجه به مجموعه کاملاً تعریف شده‌ای از هدفهای محتوایی، به کار می‌رود. پایایی این نوع آزمون به میزان ثباتی که در برآورد موقعیت فرد وجود دارد، مربوط می‌شود. آیا فرد در دو اجرای آزمون همان سطح مهارت را دارد؟ شیوه‌های سنتی پایایی، نظیر همبستگی و  $R_{20-K}$ ، برای آزمونهای ملاکی مناسب تشخیص داده نمی‌شوند؛ زیرا این شیوه‌ها به تغییرپذیری افراد وابسته هستند. در آزمونهای ملاکی، به طور معمول هیچ تغییرپذیری در نمره‌ها وجود ندارد یا بسیار کم وجود دارد؛ زیرا آموزش تا آنجا ادامه پیدا می‌کند که مهارت مورد نظر به طور کامل، یا تقریباً به طور کامل، فراگرفته شود. تغییرپذیری یا پراکندگی محدود نمره‌ها منجر به برآوردهای پایین یا نزدیک به صفر پایایی خواهد شد؛ حتی اگر آزمون ثبات درونی داشته باشد یا بسیار پایا باشد.

برای برآورد پایایی آزمونهای ملاکی، شیوه‌های متعددی پیشنهاد شده است. یک شیوه نسبتاً ساده آن است که دو صورت هم ارز از آزمونی را با آزمودنیها اجرا کنیم یا آزمون یکسانی را در دو موقعیت زمانی مختلف با آزمودنیها اجرا کنیم و سپس همسانی تصمیماتی را که بر مبنای نتایج آنها گرفته شده است، با هم بسنجیم. همسانی نتایج را از طریق یافتن درصد افرادی که در دو اجرای آزمون در مورد آنها تصمیم (مهارت یا عدم مهارت) یکسانی گرفته شده است، تعیین می‌کنند. از این شاخص با عنوان ضریب توافق ( $\rho_0$ ) نام می‌برند.

برای مثال، فرض کنید دو صورت هم ارز از یک آزمون ملاکی با یک نمونه ۱۰۰ نفری از دانش‌آموزان اجرا شده است. در هر دو اجرای آزمون، ۷۰ دانش‌آموز به عنوان افراد ماهر و ۱۴ دانش‌آموز به عنوان افراد غیرماهر طبقه‌بندی شدند. ضریب توافق، درصد کل افرادی است که در هر دو اجرای آزمون به یک نحو طبقه‌بندی شده‌اند.

$$\rho_0 = \frac{a + d}{N} = \frac{70 + 14}{100} = \frac{84}{100} = 0.84 \quad (8-10)$$

$a$  = تعداد افرادی که در هر دو اجرای آزمون به عنوان افراد ماهر طبقه‌بندی شده‌اند

$d$  = تعداد افرادی که در هر دو اجرای آزمون به عنوان افراد غیر ماهر طبقه‌بندی شده‌اند

$N$  = تعداد کل آزمودنیها

$\rho_0$  = ضریب توافق



بنابراین، ۸۴ درصد از آزمودنیها در هر دو اجرای آزمون به یک نحو طبقه‌بندی شده‌اند و شاخص پایایی این آزمون ۸۴٪ است. اگر طبقه‌بندی افراد ماهر و غیرماهر برای کل آزمودنیها در هر دو اجرای آزمون یکسان بود، ضریب توافق به حداکثر مقدار خود، یعنی ۱، می‌رسید. انتظار می‌رود که برخی از توافقات در طبقه‌بندی افراد به عنوان ماهر و غیرماهر صرفاً از روی اتفاق صورت گرفته باشد. بنابراین، برخی از مؤلفان استفاده از ضریب کاپا (K) (۱۱) را پیشنهاد کرده‌اند، که آماره‌ای است که توافقی اتفاقی را در نظر می‌گیرد. ضریب کاپا به نسبتی از طبقه‌بندیهای همسان اطلاق می‌شود که حاصل تصادف و اتفاق نیست.

منطق ضریب کاپا صریح و روشن است. محقق ابتدا درصد مواردی را که انتظار می‌رود طبقه‌بندی همسانی داشته باشند، محاسبه می‌کند؛ حتی اگر هیچ رابطه آماری بین آزمونها وجود نداشته باشد و به عبارتی اگر طبقه‌بندیهای دو اجرای آزمون کاملاً مستقل از هم باشند. این شاخص را توافقی اتفاقی مورد انتظار ( $\rho_c$ ) می‌نامند. توافق اتفاقی مورد انتظار را از توافق مشاهده شده کم می‌کنند ( $\rho_o - \rho_c$ ) تا فزونی واقعی بر همسانی اتفاقی معین شود. سپس این مقدار را بر  $1 - \rho_c$ ، یعنی حداکثر افزایش ممکن در همسانی تصمیم‌گیرها که رای اتفاق است، تقسیم می‌کنند تا ضریب کاپا به دست آید.

$$\rho_c = \frac{(a + b)(a + c) + (c + d)(b + d)}{N^2} \quad (۸-۱۱)$$

$\rho_c$  = درصد توافقی که انتظار می‌رود برحسب اتفاق باشد

$$K = \frac{\rho_o - \rho_c}{1 - \rho_c} \quad (۸-۱۲)$$

$K$  = درصد توافقی که حاصل اتفاق نیست

$\rho_o$  = ضریب توافق مشاهده شده

$\rho_c$  = توافق اتفاقی

اگر از داده‌های مثال قبل استفاده کنیم:

$$\rho_c = \frac{(80)(76) + (20)(24)}{100^2}$$

$$= \frac{6080 + 480}{10000}$$

$$= 0/66$$

$$K = \frac{0/84 - 0/66}{1 - 0/66}$$

$$= \frac{0/18}{0/34}$$

$$= 0/53$$

روشن است که ضریب کاپا (۰/۵۳) در مقایسه با ضریب توافق (۰/۸۴)، برآورد پایین تری از پایایی را ارائه می‌کند. اغلب هر دوی این ضرایب محاسبه می‌شوند؛ زیرا اطلاعاتی که این دو شاخص فراهم می‌آورند، با هم تفاوت دارند. برای بحث بیشتر در مورد عواملی که بر این اندازه‌های پایایی آزمونهای ملاکی تأثیر می‌گذارند، خواننده را به کتاب کراکر و آلگینا<sup>(۱)</sup> ارجاع می‌دهیم.

ضریب توافق و ضریب کاپا به دو اجرا از یک آزمون نیاز دارند. هیون و سوبکویاک روشهایی را برای برآورد این ضرایب از تنها یک بار اجرای آزمون ابداع کرده‌اند. خواننده را برای کسب اطلاعات بیشتر به کتابهای این مؤلفان ارجاع می‌دهیم.<sup>(۲)(۳)</sup>

به هر حال آسانترین روش، استفاده از جدولهایی است که سوبکویاک تهیه کرده است و به محقق اجازه می‌دهد که مقدار تقریبی ضریب توافق یا ضریب کاپا را بعد از یک بار اجرای آزمون مستقیماً از روی جدول بخواند. این جداول در مقاله‌ای که سوبکویاک نوشته است، ارائه شده‌اند.<sup>(۴)</sup>

### // مقایسه اعتبار و پایایی

اعتبار یک ویژگی مهمتر و جامعتر از پایایی است، همچنین تعیین آن نیز دشوارتر است. مطالعات تحقیقاتی منتشره، اغلب داده‌های پایایی را بسیار بیشتر از داده‌های اعتباری گزارش

1- Crocker, L., and J. Algina (1986). *Introduction to Classical and Modern Test Theory* (pp. 197 - 203). New York: Holt, Rinehart and Winston.

2- Huynh, H. (1976). On the Reliability of Decisions In Domain - Referenced Testing. *Journal of Educational Measurement*, 13, 253 - 264.

3- Subckovlak, M.J. (1978). Estimating Reliability From a Single Administration of a Mastery Test. *Journal of Educational Measurement*, 13, 265 - 276.

4- Subckovlak, M.J. (1988). A Practitioner's Guide to Computation and Interpretation of Reliability Indices for Mastery Tests. *Journal of Educational Measurement*, 25, 47 - 55.

می‌کنند. این امر به دلیل آن است که اعتبار را، همانند پایایی، نمی‌توان به طور مستقیم به دست آورد. برای سنجش اعتبار یک وسیله اندازه‌گیری، به قضاوت‌های ذهنی بسیار بیشتری نیاز است. محقق باید به سؤالاتی درباره تناسب محتوای آزمون، شایستگی ملاکها، تعریف صفات انسانی، مشخصات حیطة رفتاری، نظریه پشتوانه محتوای آزمون و غیره پاسخ گوید. تمام این مسائل به قضاوت و گردآوری داده‌ها از منابع مختلف نیاز دارند.

از سوی دیگر، پایایی را می‌توان به طور مستقیم از طریق داده‌های آزمون مورد بررسی قرار داد و برای این کار به داده‌های بیرونی، یعنی داده‌های نامرتبط با اطلاعات آزمون، نیازی نیست. مقوله‌های اصلی پایایی بسادگی مورد تجزیه و تحلیل ریاضی قرار می‌گیرند و نتایج منطقی درباره مقدار خطا را می‌توان به زبان ریاضی بیان کرد.

اگر یک اندازه‌داری اعتبار باشد، محقق باید در اندازه‌گیریهای مختلف نتایج یکسانی را به دست آورد، یعنی اندازه‌ای که دارای اعتبار است، پایایی هم دارد، هر چند عکس این گفته لزوماً صادق نیست. به خاطر داشته باشید که یک اندازه می‌تواند پایایی داشته باشد، بدون آن که دارای اعتبار باشد، یعنی می‌تواند به طور مداوم یک چیز نادرست را اندازه بگیرد. فلت و برنان با این بیان که «اگر اندازه‌ای که داده‌های پایایی در مورد آن اعمال می‌شوند نامرتبط یا بی‌ثمر باشد، به رغم ظرافت روشهایی که برای تحلیل داده‌های پایایی مورد استفاده قرار می‌گیرند، هیچ یک از آنها چندان ارزشی نخواهند داشت»<sup>(۱)</sup>، بر الویت اعتبار در ارزیابی شایستگی یک اندازه آموزشی تأکید می‌کنند.

### /// خلاصه

تنوع وسایل اندازه‌گیری موجود، استفاده از معیارهایی را برای ارزیابی این وسایل ضروری می‌سازد. دو معیار مهم برای ارزیابی وسایل اندازه‌گیری، اعتبار و پایایی هستند. اعتبار به تناسب، معناداری و سودمندی استنتاجهای مشخصی که از نمره‌های آزمون به عمل می‌آید، اطلاق می‌شود. در فرآیند اعتباریابی یک آزمون، شواهدی گردآوری می‌شود تا این استنتاجها تایید شوند.

1- Feldt, L. S., and R. L. Brennan (1989). Reliability. In R. Linn (ed). *Educational Measurement* (p. 143). New York: American Council on Education and Macmillan.

برای گردآوری شواهد مربوط به اعتبار آزمون، از شیوه‌های متعددی استفاده می‌شود. شواهد مرتبط با محتوا نشان می‌دهند که آیا محتوای آزمون از آن حیظه موضوعی، که باید در مورد آن نتایجی به دست آید، بخوبی نمونه‌گیری کرده است یا خیر. شواهد مرتبط با ملاک نشان می‌دهند که نمره‌های آزمون چقدر با ملاکهای مناسب همبستگی دارند. شواهد مرتبط با سازه نشان می‌دهند که آزمون کدام سازه‌های روان‌شناختی یا آموزشی را اندازه می‌گیرد.

آگاهی از این که نمره‌های آزمون چه کاربردی خواهند داشت، از اهمیت بسزایی برخوردار است. امکان دارد آزمونی برای انجام یک کار، معتبر و در موارد دیگر، نامعتبر باشد. آزمونی که برای اندازه‌گیری استعداد تحصیلی طراحی شده است، ممکن است اندازه معتبری برای استعداد تحصیلی باشد، اما برای اندازه‌گیری استعداد هنری اعتبار نداشته باشد.

همچنین، ما باید این سؤال را مطرح کنیم که یک آزمون یا چه ثباتی اندازه‌گیری می‌کند. این مسئله پایایی است. هیچ آزمونی نمی‌تواند اعتبار داشته باشد، مگر آن که با دقت و ثبات اندازه‌گیری کند، یعنی پایا باشد. پایایی به میزان ثبات یک آزمون در اندازه‌گیری هر آنچه که اندازه می‌گیرد، اطلاق می‌شود. به طور اخص، پایایی یک آزمون که با ضریب پایایی زیاد، یا با خطای معیار کم نشان داده می‌شود، به معنای آن است که در اندازه‌گیریهای متعددی که با آن آزمون در مورد یک فرد انجام می‌شود، نتایج تقریباً یکسانی به دست می‌آید. ضرایب پایایی را با توجه به این که چه منبع خطایی در نظر گرفته می‌شود، می‌توان از راههای مختلفی محاسبه کرد. ضریب پایایی نشان می‌دهد که خطاهای تصادفی اندازه‌گیری تا چه حد بر نمره‌های آزمون تأثیر می‌گذارند. خطای معیار اندازه‌گیری به ما امکان می‌دهد که از متحنی نرمال برای برآورد حدودی که انتظار می‌رود نمره واقعی آزمودنی در داخل آن قرار گیرد، استفاده کنیم.

برای آزمونهای هنجاری و ملاکی انواع مختلفی از ضرایب پایایی مورد نیاز است. تعیین پایایی آزمونهای ملاکی پیچیده‌تر است، اما دو شیوه‌ای که در مورد این آزمونها به کار می‌روند، در این فصل ارائه شدند.

## // مفاهیم اصلی

اعتبار	پایایی
پایایی بازآزمایی	پایایی ثبات درونی
پایایی دو نیمه سازی	پایایی صورتهای هم ارز

خطای معیار اندازه‌گیری	پایایی ارزیابیها
شواهد محتوایی برای اعتبار آزمون	شواهد سازه‌ای برای اعتبار آزمون
ضریب توافق	شواهد مرتبط با ملاک برای اعتبار آزمون
	فرمول اسپیرمن - براون

### /// تمرینها

- ۱- با توجه به موارد زیر، اعتبار و پایایی را مقایسه کنید:
  - الف - معنای هر یک از مفاهیم
  - ب - اهمیت نسبی هر یک از مفاهیم
  - ج - میزان وابستگی یکی به دیگری
- ۲- در مورد این جمله توضیح دهید: یک وسیله اندازه‌گیری می‌تواند پایا باشد، بی‌آنکه اعتبار داشته باشد، اما نمی‌تواند اعتبار داشته باشد و در عین حال پایا نباشد.
- ۳- در مورد اعتباریابی یک آزمون استعداد تحصیلی، که به منظور سنجش دانش‌آموزان سال آخر دبیرستان تهیه شده است، چه پیشنهادی دارید؟
- ۴- از شما خواسته شده که یک وسیله را که برای اندازه‌گیری خود پنداره تحصیلی دانش‌آموزان طراحی شده است، اعتباریابی کنید. منظور از خود پنداره تحصیلی این است که دانش‌آموزان چه درکی از خود به عنوان یک دانش‌آموز دارند. شما برای تعیین اعتبار این وسیله اندازه‌گیری چه کار خواهید کرد؟
- ۵- هر یک از موقعیتهای زیر بر کدام یک از سه نوع شواهد اعتباری دلالت دارد:
  - الف - دو گروه از دانشجویان دانشگاه را در نظر بگیرید که افراد گروه اول ترک تحصیل کرده‌اند و افراد گروه دوم به تحصیل ادامه می‌دهند. نمره‌های آزمون دبیرستانی تبحر در زبان آنان با هم مقایسه می‌شود تا تعیین شود که آیا داده‌های آزمون با وضعیت تحصیلی آنها همبستگی دارد یا خیر.
  - ب - مشخص می‌شود که یک آزمون جدید استعداد تحصیلی با SAT، که برای پیش‌بینی موفقیت در تحصیل دانشگاهی به کار رفته است، همبستگی ۰/۹۳ دارد.
  - ج - یک آزمون هوش جدید تهیه شده است. مؤلف آزمون استدلال می‌کند که فرآیندهای ذهنی که آزمون می‌طلبد با نظریه Z هوش همخوانی دارد. بعلاوه، با هر سال افزایش سن، نمره متوسط آزمون‌دنیها در این آزمون افزایش می‌یابد.
  - د - آموزگاری یک آزمون پیشرفت تحصیلی استاندارد شده را به دقت بررسی می‌کند تا ببیند

آیا اطلاعات و مهارتهایی را که در کلاس بر آنها تأکید شده است، در برمی‌گیرد یا نه.

۵- مشخص شده است که پس از شرکت در مقیاس لیبرالیسم، تفاوت میانگین بین رتبه‌بندیهای اعضای گروه کوکلوکس کلان و اعضای گروه اقدام دموکراتیک، کاملاً معنادار است. و - گروهی از آموزگاران در مورد یک آزمون ریاضی داوری می‌کنند تا مشخص کنند که آیا یک نمونه کافی و معرف از مطالب درس ریاضی هست یا خیر.

۶- در هر یک از موارد زیر چه روشی برای برآورد پایایی به کار گرفته شده است:

الف - آزمون یکسانی دو بار به یک گروه معین داده شد. همبستگی بین نمره‌های دو اجرای آزمون ۰/۹ بود.

ب - همبستگی نمره آزمودنیها در سؤالات فرد آزمون با نمره آنها در سؤالات زوج همان آزمون محاسبه شد:  $r_{xx} = 0/95$ .

ج - صورتهای هم ارز یک آزمون به فاصله یک ماه اجرا شد و همبستگی نتایج دو اجرای آزمون محاسبه شد:  $r_{xx} = 0/85$ .

د - واریانس، میانگین و تعداد سؤلهای آزمون برای برآورد پایایی مورد استفاده قرار گرفت.

۷- با فرض این که در سؤال ۶، یک گروه در دو اجرای آزمون شرکت کند، شما تفاوتی ضرایب پایایی را چگونه توجیه خواهید کرد؟

۸- شما یک آزمون آمادگی خواندن را چگونه اعتباریابی خواهید کرد؟

۹- به هنگام ساختن یک آزمون، برای افزایش پایایی آن چه کار می‌توان کرد؟

۱۰- نوع شواهدی را که برای اعتباریابی انواع آزمونهای زیر بیشترین اهمیت را خواهد داشت، مشخص کنید:

الف - یک آزمون هجی کردن کلاسی

ب - وسیله‌ای برای اندازه‌گیری انگیزه پیشرفت

ج - یک اندازه که برای شناسایی ترک تحصیل کنندگان بالقوه طراحی شده است.

۱۱- توضیح دهید که چگونه یک آزمون ریاضی می‌تواند در یک کلاس ریاضی، اعتبار زیاد

و در کلاس ریاضی دیگر، اعتبار کمی داشته باشد.

۱۲- این جمله را نقد کنید: پایایی آزمون هوش ۰/۹ است. بنابراین، می‌توان فرض کرد که این

آزمون واقعاً هوش را اندازه می‌گیرد.

۱۳- برای آزمونی که انحراف معیار آن ۱۶ و ضریب پایایی آن  $r_{xx} = 0/84$  است، خطای معیار

- اندازه‌گیری را تعیین کنید. این خطای معیار اندازه‌گیری را چگونه تفسیر خواهید کرد؟
- ۱۴- یک آزمون پیشرفت تحصیلی استاندارد شده را که می‌توانید در یک مطالعه تحقیقاتی به کار برید، انتخاب کنید و داده‌های اعتباری لازم را در مورد این آزمون به دست آورید (شما می‌توانید از سالنامه بوروس و کتابچه راهنمای آزمون استفاده کنید).
- ۱۵- به کتابچه راهنمای آزمون پیشرفت تحصیلی که در مدرسه شما مورد استفاده قرار می‌گیرد، مراجعه کنید. چه نوع داده‌های پایایی در آن گزارش می‌شود؟
- ۱۶- هنگامی که دو صورت هم ارز یک آزمون ملاکی در ریاضیات به گروهی از کودکان دبستانی داده شد، داده‌های زیر به دست آمد. در هر آزمون ۵۰ سؤال وجود داشت. یک دانش‌آموز در هر آزمون برای قبول شدن باید ۸۰ درصد پاسخ صحیح می‌داد. پایایی این آزمون را برحسب ضریب توافق ( $P$ ) و ضریب کاپا ( $K$ ) بیان کنید.

آزمودنی	صورت ۱	صورت ۲
۱	۴۵	۴۷
۲	۴۳	۴۸
۳	۴۵	۳۱
۴	۳۹	۳۹
۵	۳۹	۴۸
۶	۳۴	۳۷
۷	۴۶	۴۶
۸	۴۸	۴۹
۹	۴۳	۳۸
۱۰	۳۶	۴۶
۱۱	۴۵	۴۸
۱۲	۳۸	۳۹
۱۳	۴۴	۴۵
۱۴	۳۱	۳۴
۱۵	۴۲	۴۸

۱۷- شیوه‌های اعتباریابی زیر را نقد کنید:

الف - یک آموزگار زبان انگلیسی دبیرستان، آزمونی نگارشی برای شناسایی دانش‌آموزان

مستعد تهیه و آن را با دانش‌آموزان کلاس انگلیسی سال آخر دبیرستان اجرا کرد. به دانش‌آموزانی که نمره بالا کسب کردند، اجازه داده شد که در کلاس زبان انگلیسی دانشگاه محلی ثبت نام کنند. این آموزگار در پایان نیمسال، همبستگی نمره‌های آزمون اولیه را با نمره‌هایی که دانش‌آموزان در کلاس زبان دانشگاه به دست آوردند، محاسبه کرد. آموزگار از این که بین این نمرات همبستگی ناچیزی پیدا کرد، متعجب شد. مشکل چیست؟

ب - یک مشاور مدرسه برای اندازه‌گیری نیاز به پیشرفت تحصیلی در کودکان دبستانی، مقیاسی تهیه کرد. این مقیاس در دو کلاس یک دبستان اجرا شد و نتایج آن در اختیار آموزگارهای آن دو کلاس گذاشته شد. از آموزگاران خواسته شد طی نیمسال تحصیلی این کودکان را به دقت مورد مشاهده قرار دهند، سپس از آنها خواسته شد این کودکان را برحسب نیاز به پیشرفت رتبه‌بندی کنند. پس از آن همبستگی رتبه‌بندی آموزگاران با نمره‌های کودکان در مقیاس مذکور محاسبه شد. همبستگی کاملاً بالا بود، بنابراین مشاور تحصیلی نتیجه گرفت که مقیاس او برای اندازه‌گیری نیاز به پیشرفت اعتبار زیادی داشته است.

آیا شما با نتیجه‌گیری مشاور تحصیلی موافق هستید؟

۱۸- فرض کنید که شما می‌خواهید «فرسودگی» معلمان را بررسی کنید. برخی از شاخصهای این سازه را که می‌توانید برای اندازه‌گیری به کار ببرید، پیشنهاد کنید.

۱۹- یک آزمون درست - نادرست به ده دانش‌آموز داده شد. عدد ۱ پاسخ درست و عدد صفر

پاسخ نادرست را نشان می‌دهد. نتایج به قرار زیر است:

### پرسشها

دانش‌آموزان	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	کل
الف	۱	۱	۱	۱	۱	۱	۱	۱	۰	۰	۸
ب	۱	۱	۱	۱	۰	۱	۰	۱	۰	۰	۶
ج	۱	۱	۱	۱	۱	۰	۰	۰	۰	۰	۵
د	۱	۱	۱	۰	۰	۰	۰	۱	۱	۰	۵
ه	۱	۱	۱	۱	۱	۱	۱	۱	۰	۰	۸
و	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱۰
ز	۱	۱	۱	۱	۱	۱	۱	۱	۱	۰	۹
ح	۱	۱	۱	۰	۰	۰	۰	۰	۰	۰	۳
ط	۱	۱	۱	۱	۰	۰	۰	۰	۰	۰	۴
ی	۱	۱	۱	۱	۰	۱	۱	۱	۱	۱	۹



الف - ضریب پایایی دو نیمه سازی این آزمون را با تقسیم آن به دو آزمون فرعی پرسشهای فرد و پرسشهای زوج محاسبه کنید.

ب - پایایی آزمون کامل چیست؟

۲۰- برای وسایل اندازه گیری زیر چه نوع برآورد پایایی مناسبتر خواهد بود؟

الف - یک آزمون پیشرفت تحصیلی چند گزینه‌ای به عنوان متغیر وابسته در یک بررسی آزمایشی مورد استفاده قرار خواهد گرفت.

ب - یک محقق تغییر نگرش را مطالعه خواهد کرد و یک صورت از مقیاس نگرش را هم به عنوان پیش آزمون و هم به عنوان پس آزمون اجرا خواهد کرد.

ج - محقق دو صورت از یک آزمون پیشرفت تحصیلی را در دست دارد. او یک صورت را در آغاز بررسی و صورت دیگر را در پایان بررسی به آزمودنیها داد. او می خواهد پایایی آزمون را تعیین کند.

۲۱- یک آزمون ۱۰۰ سؤالی به دو نیمه تقسیم و مشخص شد که ضریب همبستگی دو نیمه سازی آن  $0/6$  است. ضریب پایایی را برای کل آزمون محاسبه کنید.

### /// پاسخها

۱- اعتبار به حد و اندازه‌ای اطلاق می شود که یک وسیله اندازه گیری آن چیزی را که برای آن طراحی شده است، اندازه می گیرد. پایایی به میزان ثبات اندازه گیری گفته می شود. اعتبار جنبه مهم تری از پایایی است؛ زیرا فقدان اعتبار، فقدان معنا است؛ هرچند یک وسیله اندازه گیری ابتدا باید پایا باشد تا اعتبار داشته باشد.

۲- یک اندازه می تواند در کاربردهای مکرر آن با گروهی معین از آزمودنیها نمره‌های باثباتی به ما بدهد؛ هر چند شاید هیچ رابطه‌ای با سایر اندازه‌های پذیرفته شده سازه نداشته باشد یا نتواند رفتاری را که با سازه پیوند خورده است، پیش بینی کند. نمره‌های آزمونی که پایایی آن صفر است، کاملاً تصادفی هستند و بنابراین نمی توانند با هیچ ملاکی همبستگی داشته باشند.

۳- برای تعیین اعتبار سازه، ابتدا باید تعریف کنیم که منظور ما از استعداد چیست. اگر بخواهیم توانایی تحصیلی عمومی را اندازه بگیریم، پس اعتبار محتوا را می توانیم با بررسی سؤالات آزمون از لحاظ معرف بودن آنها، تعیین کنیم. آیا این سؤالات مهارتهای تحصیلی

اساسی نظیر خواندن، هجی کردن، ریاضیات و غیره را می‌سنجند؟ اعتبار ملاکی با ضریب همبستگی بین نمره‌های آزمون، معدل سال آخر دبیرستان، معدل سال اول دانشگاه و دیگر ملاکها سنجیده خواهد شد. همبستگی با دیگر آزمونهای استعداد، که اعتبار آنها مشخص شده است نیز می‌تواند مدنظر باشد.

۴. سوالات مقیاس یا پرسشنامه می‌باید جنبه‌های رفتاری دانش‌آموزان را که منطقاً بخشی از سازه خود پنداره تحصیلی خواهند بود، در بگیرند (برای مثال، من قصد دارم وارد دانشکده شوم). اندازه‌های ملاکی می‌توانند مصاحبه فردی یا دانش‌آموزان یا سنجش مستقل به وسیله آموزگاران باشند. با فرض این که خود پنداره تحصیلی با پیشرفت تحصیلی مرتبط است، می‌توان همبستگی نمره‌های خود پنداره را با معدل و یا نمره‌های آزمون پیشرفت تحصیلی محاسبه کرد.

۵. الف - مرتبط با ملاک

ب - مرتبط با ملاک

ج - مرتبط با سازه

د - مرتبط با محتوا

ه - مرتبط با سازه

ز - مرتبط با محتوا

۶. الف - پایایی بازآزمایی

ب - پایایی دو نیمه سازی

ج - پایایی صورتهای هم ارز

د - هم ارزی منطقی (فرمول ۲۱ کودر - ریچاردسون)

۷. پایایی‌های دو نیمه سازی معمولاً از پایایی‌های بازآزمایی بیشتر هستند؛ زیرا تغییرپذیری آزمودنیها که ناشی از پختگی، افزایش مهارتهای شرکت در آزمون و سایر عوامل تصادفی باشد، کمتر است. پایایی صورتهای هم ارز کمتر از پایایی آزمونی است که هر دو صورت آن یکی باشند؛ زیرا (الف) ساختن صورتهای دقیقاً هم ارز از یک آزمون ناممکن است و (ب) وقتی از صورتهای غیریکسان آزمون استفاده می‌شود، یک منبع تغییرپذیری اضافی نیز دخالت می‌کند. اگر آزمون همگن نباشد، پایایی هم ارزی منطقی کاهش خواهد یافت.

۸. ابتدا باید دید کدام مهارتهای مشخص (برای مثال، بازشناسی حروف، جهت‌یابی چپ به

راست) شامل آمادگی خواندن است و سپس تعیین کرد که آیا آزمون این مهارتها را به نسبت مناسب در خود دارد یا نه. هنگامی که آزمودنیهایی که در آزمون شرکت کرده بودند، برنامه‌های خواندن خود را آغاز کردند، تعیین خواهیم کرد که نمره‌های آزمون و آزمونهای فرعی با نمره‌های آزمون خواندن، رتبه‌بندی آموزگاران و سایر ملاکها، چقدر همبستگی دارند.

۹- بازنویسی سؤالات مبهم و تشریح دستورالعملها پایایی را افزایش خواهند داد. طولانی‌تر کردن آزمون با گنجانیدن سؤالات اضافی که از همان حیطه محتوایی استخراج شده باشند، پایایی را افزایش می‌دهد؛ همان طور که آزمون کردن یک گروه ناهمگن‌تر نیز موجب آن می‌شود.

۱۰- الف - محتوا

ب - سازه

ج - مرتبط با ملاک

۱۱- یک آزمون ریاضی که فقط محاسبه را مورد نظر قرار داده است، در کلاسی که تأکید آن بر مفاهیم و استدلال است، اعتبار کمی خواهد داشت. اگر محتوا و تأکید یک کلاس دیگر یا محتوا و تأکید آزمون همخوانی داشته باشد، این آزمون در آن کلاس اعتبار زیادی خواهد داشت.

۱۲- یک آزمون می‌تواند پایا باشد و در عین حال چیزی را اندازه بگیرد که قصد اندازه‌گیری آن را ندارد. برای تعیین اعتبار، لازم است که محتوا، سازه‌ها و روابط، با سایر اندازه‌های همان سازه را مدنظر قرار دهیم و نیز روابط با اندازه‌های رفتاری را، که به فرض با آن سازه همبستگی دارند، در نظر بگیریم.

۱۳- برحسب فرمول ۹ - ۸

$$\begin{aligned} S_M &= S_x \sqrt{1 - r_{xx}} \\ &= ۱۶ \sqrt{1 - ۰/۸۴} \\ &= (۱۶)(۰/۴) \text{ پ} \\ &= ۶/۴ \end{aligned}$$

خطای معیار اندازه‌گیری را مانند انحراف معیار تفسیر می‌کنند. بنابراین، می‌توان گفت از هر سه بار، دو بار این احتمال وجود دارد که نمره واقعی فرد در دامنه (نمره مشاهده شده  $\pm ۶/۴$ ) قرار گیرد.

۱۴- پاسخها متفاوت خواهند بود.

۱۵- پاسخها متفاوت خواهند بود.

۱۶- نمره ۴۰، معرف تسلط بر مطلب است ( $40 = 0.8 \times 50$ ).

صورت ۱

		مسلط	غیرمسلط	
صورت ۲	مسلط	(الف) ۷	(ب) ۲	۹
	غیرمسلط	(ج) ۲	(د) ۴	۶
		۹	۶	۱۵

$$p_o = \frac{7+4}{15} = \frac{11}{15} = 0.73 \text{ (۷۳٪ دانش آموزان به یک نحو طبقه بندی شدند)}$$

$$p_c = \frac{(9)(9) + (6)(6)}{15^2} = \frac{81 + 36}{225} = \frac{117}{225} = 0.52$$

$$K = \frac{0.73 - 0.52}{1 - 0.52} = \frac{0.21}{0.48} = 0.44$$

۱۷- الف - آزموننی که آموزگار تهیه کرده بود، برای آنچه که او در نظر داشت، مناسب نبود.  
ب - در واقع ناخالصی ملاک وجود داشته است. چون به آموزگاران اجازه داده شد که نتایج اندازه گیری اولیه نیاز به پیشرفت را ببینند، این امر بر رتبه بندی دانش آموزان در زمینه نیاز به پیشرفت تأثیر گذاشت و آن را ناخالص کرد.

۱۸- شاخصهای احتمالی متعددی در مورد فرسودگی معلم وجود دارند. در این زمینه می توان به میزان غیبت از مدرسه، ارزیابی پایین مدیران، بروز خصومت نسبت به دانش آموزان یا مدیران و ابراز مخالفت با سیاستهای مدرسه اشاره کرد. همچنین می توان مقیاسی برای

اندازه‌گیری نگرش معلمان نسبت به کار خود تهیه و از پاسخهای معلمان وجود فرسودگی را استنتاج کرد.

۱۹-

دانش آموز	فرد	زوج	کل
الف	۴	۴	۸
ب	۲	۴	۶
ج	۳	۲	۵
د	۳	۲	۵
ه	۴	۴	۸
و	۵	۵	۱۰
ز	۵	۴	۹
ح	۲	۱	۳
ط	۲	۲	۴
ی	۴	۵	۹

الف - ضریب پایایی دو نیمه سازی  $r = 0/72$  است (ضریب همبستگی،  $r$  را با استفاده از فرمول نمره‌های خام با  $N$  در مخارج محاسبه کنید).

ب - برآورد پایایی برای کل آزمون  $0/84$  است.

۲۰- الف - محقق به ثبات درونی این آزمون علاقه‌مند خواهد بود. ضریب پایایی دو

نیمه‌سازی مناسب است.

ب - با به کار بردن یک صورت از آزمون، هم به عنوان پیش‌آزمون و هم به عنوان پس

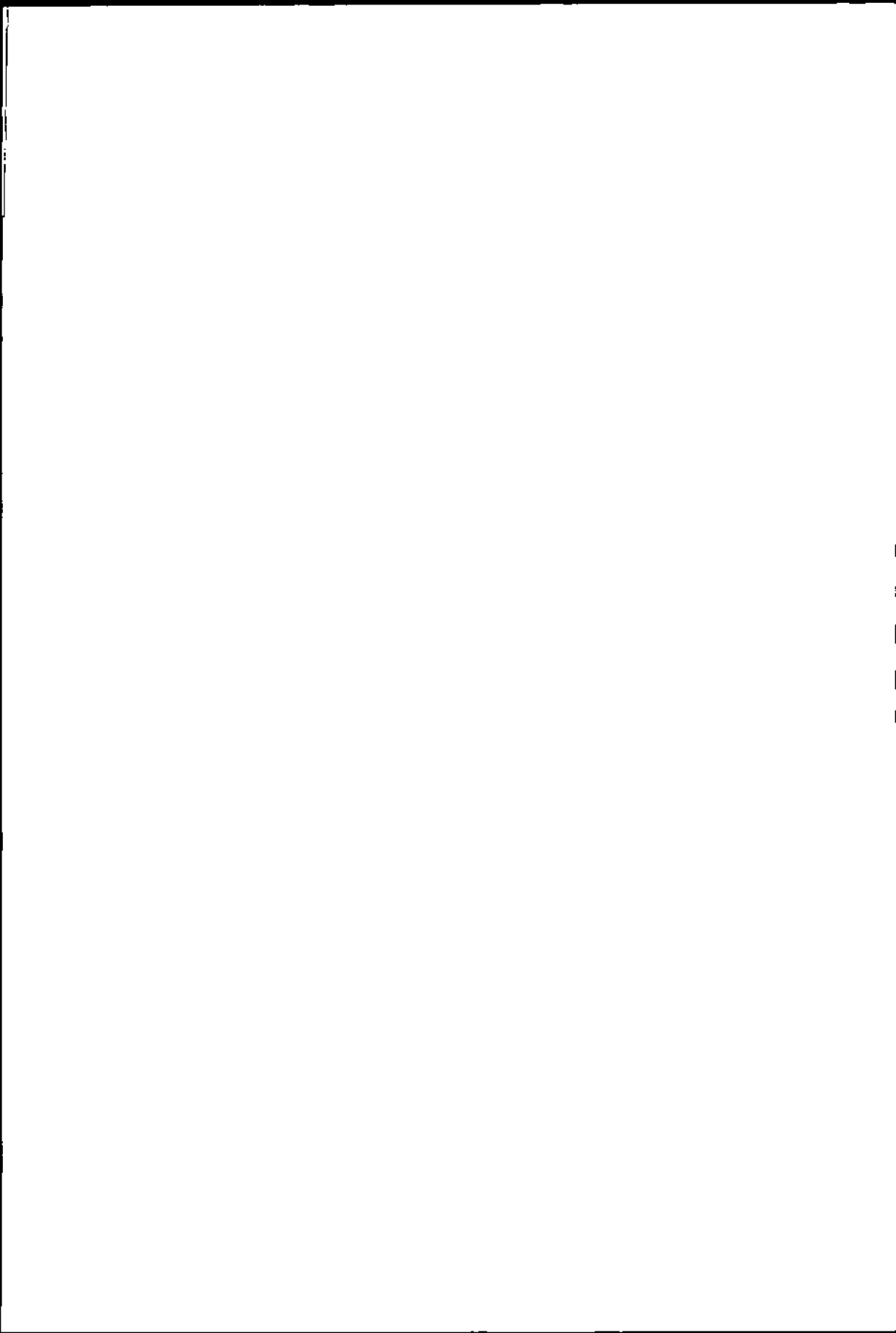
آزمون، محقق ضریب استواری را محاسبه خواهد کرد.

ج - با دو صورت از یک آزمون و دو اجراء، محقق ضریب استواری و هم ارزی را محاسبه

خواهد کرد.

۲۱-

$$r_{xx} = \frac{2(0/60)}{1 + 0/60} = 0/75$$



# بخش پنجم

## روشهای تحقیق

### فصل نهم

#### تحقیق آزمایشی در تعلیم و تربیت

##### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- ویژگیهای تحقیق آزمایشی را توصیف کند.
- ۲- اعتبار داخلی و خارجی را تعریف کند.
- ۳- عوامل احتمالی مخدوش کننده اعتبار داخلی و راهبردهایی را برای اجتناب از آنها و یا به حداقل رساندن آنها، بشناسد.
- ۴- مشکلات مربوط به اعتبار داخلی را در طرحهای پیشنهادی و گزارشهای تحقیقاتی بشناسد.
- ۵- عوامل احتمالی مخدوش کننده اعتبار خارجی و راهبردهای اجتناب از آنها را بشناسد.
- ۶- مشکلات مربوط به اعتبار خارجی را در طرحهای پیشنهادی و گزارشهای تحقیقاتی بشناسد.
- ۷- تفاوت بین نمونه گیری تصادفی و تخصیص تصادفی را توصیف کند و نشان دهد که ارتباط بین آنها و اعتبارهای داخلی و خارجی چگونه است.
- ۸- تحقیقات آزمایشی تک آزمودنی و موارد استفاده از آنها را توصیف کند.

یک آزمایش عبارت است از یک بررسی علمی که در آن آزمایشگر یک یا چند متغیر مستقل را کنترل و اثرات این دستکاری را روی متغیر یا متغیرهای وابسته، مشاهده می‌کند. محقق از پیش یک یا چند فرضیه را در مورد ماهیت رابطه مورد انتظار تبیین می‌کند. آزمایش با این هدف طرح ریزی و اجرا می‌شود که محقق بتواند شواهدی در مورد فرضیه‌هایش گردآوری کند. آزمایشگر، آگاهانه و نظام‌مند تغییراتی در پدیده‌های طبیعی به وجود می‌آورد و سپس پیامدهای این تغییرات را مشاهده می‌کند. فرضیه‌ها بیانگر انتظارات محقق، در قالب یافته‌هایی است که خود حاصل تغییراتی هستند که او ایجاد کرده است. در اجرای یک آزمایش، محقق در مورد دستکاری و کنترل متغیرها و نیز مشاهده و اندازه‌گیری نتایج، دقت زیادی صرف می‌کند. از طریق چنین روش تحقیقی است که اقتناع‌کننده‌ترین شواهد، در مورد اثر یک متغیر بر دیگری، می‌تواند به دست آید.

دانشمندان قدیم، به ارزش مشاهده در مطالعه محیط پی برده بودند، ولی خیلی زود دریافته‌اند که پیچیدگی طبیعت را نمی‌توان تنها با مشاهده ساده رویدادهای بی‌شمار آن درک کرد. آنها دریافته‌اند که رویدادهایی که در حالت طبیعی خود اتفاق می‌افتند، اغلب آنچنان به وسیله عوامل نامرتبط دستخوش پیچیدگی می‌شوند که فعالیت عامل مورد مطالعه آنها، مجهول باقی می‌ماند. این مشکل، با کنترل شرایطی که رویدادها تحت آن رخ می‌دهند و با حذف متغیرهای نامرتبط، حل شد. بنابراین، دانشمندان توانستند آگاهانه متغیرهای مستقل را دستکاری کنند و تغییرات حاصل از متغیرهای وابسته را، که منتج از تغییر در متغیرهای مستقل هستند، اندازه‌گیری کنند. به سخن دیگر، دانشمندان آزمایش را آغاز کردند.

از آنجا که کاربرد روشهای آزمایشی در بررسی جهان فیزیکی مشمر ثمر بود، این روشها در دیگر رشته‌ها نیز به کار گرفته شدند. قرن نوزدهم شاهد ورود این روشها به علوم زیستی بود، پیشرفتهای زیادی در جانورشناسی، فیزیولوژی و طب حاصل شد. در اواخر قرن نوزدهم، دانشمندان همین روشها را در مسائل روان‌شناسی به کار بردند و از همین جا رشته روان‌شناسی آزمایشی به وجود آمد. در دهه ۱۹۸۰، روش آزمایشی را در بررسی یک مسئله آموزشی به کار بردند. بررسی رایس<sup>(۱)</sup> در مورد پیشرفت هجی کردن در دبستان، به عنوان اولین قدم در کاربرد



آزمایش در رشته تعلیم و تربیت محسوب می‌شود. ثورندایک<sup>(۱)</sup> و دیگر محققان هم عصر او کاربرد روش آزمایشی را در تعلیم و تربیت توسعه دادند.

یک آزمایش در ساده ترین وجه خود، دارای سه مشخصه است:

۱- یک متغیر مستقل دستکاری می‌شود.

۲- کلیه متغیرهای دیگر، به استثنای متغیر مستقل، ثابت نگاه داشته می‌شوند.

۳- تأثیر دستکاری متغیر مستقل بر متغیر وابسته مشاهده می‌شود.

در نتیجه، در یک آزمایش، دو متغیری که به طور خاص مورد توجه هستند، متغیر مستقل و متغیر وابسته می‌باشند. متغیر مستقل به وسیله آزمایشگر دستکاری می‌شود (یا تغییر داده می‌شود). متغیری که تأثیر تغییرات بر روی آن مشاهده می‌شود، ولی دستکاری نمی‌شود، متغیر وابسته خوانده می‌شود. دلیل خوانده شدن متغیر وابسته به این نام، این است که فرض می‌شود مقدار آن وابسته به مقدار متغیر مستقل است و با تغییر آن، تغییر می‌کند. برای مثال، ممکن است آزمایشگری بخواهد اثرات روشهای مختلف تدریس را در میزان پیشرفت خواندن دانش آموزان، بررسی کند. در این مورد، او به منظور تعیین تأثیر روش تدریس بر روی میزان پیشرفت، که همانا متغیر وابسته است، روش تدریس را که همان متغیر مستقل است، تغییر می‌دهد.

### /// ویژگیهای تحقیق آزمایشی

در یک آزمایش، سه جزء اساسی وجود دارند که محقق فعالانه با آنها سر و کار دارد. این سه جزء عبارتند از: کنترل، دستکاری و مشاهده.

### // کنترل

اساس روش آزمایشی کنترل است. بدون کنترل، ارزشیابی بدون ابهام تأثیرات یک متغیر مستقل امکان پذیر نیست. حال به بررسی مختصر مفهوم کنترل در آزمایش می‌پردازیم.

اصولاً روش آزمایشی در علم، بر مبنای دو فرض درباره متغیرها استوار است: ۱- چنانچه دو وضعیت از هر نظر همسان باشند و تنها تفاوت آنها وجود و یا عدم یک عامل در یکی از وضعیتها باشد، تفاوت بین دو موقعیت را می‌توان ناشی از همان عامل دانست. این تبیین را

«قانون متغیر تنها»<sup>(۱)</sup> می خوانند. چنانچه دو وضعیت همسو نباشند، ولی بتوان نشان داد که هیچ یک از متغیرها در ایجاد پدیده مورد مطالعه تأثیری ندارند یا متغیرهای مهم با هم همسان هستند، هر تغییری را که بین این دو موقعیت، بعد از دخالت دادن یک متغیر جدید در یکی از سیستمها، به وجود آید، می توان به متغیر جدید نسبت داد. این تبیین را «قانون تنها متغیر با اهمیت»<sup>(۲)</sup> می خوانند.

هدف از کنترل در یک آزمایش، ایجاد وضعیتی است که تحت آن بتوان اثرات متغیرها را بررسی کرد. شرایطی که قانون متغیر تنها بر آن استوار است، در علوم تجربی بیش از تعلیم و تربیت کاربرد دارد. از آنجا که تحقیقات در زمینه تعلیم و تربیت با نوع بشر سروکار دارد، همیشه متغیرهای بسیار زیادی دخالت دارند. کوشش در تقلیل مسائل تعلیم و تربیت در جهت کاربرد یک متغیر، نه تنها غیر واقعی، بلکه شاید حتی غیر ممکن باشد. خوشبختانه چنین کنترل دقیقی واقعاً ضروری نیست، زیرا بسیاری از جنبه هایی که در آن موقعیتها با هم تفاوت دارند، با هدف بررسی ارتباطی ندارند و در نتیجه می توان آنها را نادیده گرفت. در نظر داشتن قانون تنها متغیر مستقل با اهمیت، کافی است. به عنوان مثال، در بررسی تأثیرات افتراقی دو روش متفاوت تدریس حساب، محقق به دو گروه کودک نیاز دارد که از کلیه جوانب، به استثنای روش یادگیری حساب، یکسان باشند. از آنجا که دستیابی به دو گروه کودک که به طور مطلق یکسان باشند، غیر ممکن است، محقق به دنبال تشکیل دو گروه می رود که تا حد امکان از نظر متغیرهایی که در امر پیشرفت در یادگیری حساب دخالت دارند، مشابه باشند؛ مانند: توانایی خواندن، انگیزش، هوش عمومی و موارد نظیر آن. دیگر متغیرهایی که به احتمال بسیار با یادگیری حساب نامرتبط هستند، مانند توانایی ورزشی، قد و رنگ مو، نادیده گرفته می شوند. با وجود آنکه قانون متغیر تنها، به صورت مطلق غیر عملی است، اما آزمایشگر سعی می کند تا آنجا که امکان دارد، آن را به متغیرهای مربوط نزدیک کند. در نتیجه، در بررسیهای آزمایشی در تعلیم و تربیت به روشهای نیاز داریم که به ما امکان مقایسه گروهها را بر مبنای متغیرهای با اهمیت می دهد. روشهای کنترل جندی جهت امکان پذیر ساختن این مقایسه، به وجود آمده است.

فرض کنیم می خواهیم این فرضیه را آزمون کنیم: «کودکانی که با روش استقرایی آموزش

1- Law of the Single Variabl

2- Mill J.S. (1846). A system of Logic (P.224) New York, Harper Brothers

می‌بینند (گروه A)، در فراگیری مفاهیم علمی نسبت به کودکانی که با روش قیاسی آموزش می‌بینند (گروه B)، برتری خواهند داشت. به عبارت دیگر، می‌خواهیم رابطه بین روش تدریس (متغیر مستقل) را با یادگیری مفاهیم علمی (متغیر وابسته)، مورد بررسی قرار دهیم. به منظور دستیابی به یک نتیجه‌گیری در مورد رابطه بین متغیر مستقل و متغیر وابسته، می‌باید تأثیرات کلیه متغیرهای نامربوط را کنترل کنیم. متغیر نامربوط، متغیری است که به هدف بررسی ربطی ندارد، ولی بر روی متغیر وابسته اثر می‌گذارد. واژه کنترل به روشهایی اطلاق می‌شود که آزمایشگر برای حذف تأثیرات افتراقی تمام متغیرهایی که با هدف بررسی نامربوط هستند، به کار می‌برد. چنانچه متغیری با متغیر وابسته نامرتب باشد، دیگر تأثیری بر روی آن ندارد و به کنترل آن از نظر تأثیر گذاری نیازی نداریم.

در تجربه یادشده، هوش مسلماً عاملی است که بر روی یادگیری مفاهیم علمی اثر می‌گذارد، بنابراین به عنوان متغیر نامربوط تلقی می‌شود و باید کنترل شود. در غیر این صورت، چنانچه کودکان گروه A با هوشتر از کودکان گروه B باشند، یادگیری بیشتر آنها را می‌توانیم به متغیر هوش نسبت دهیم و در نتیجه به طور کامل نمی‌توانیم اثرات روش آموزش را در یادگیری ارزشیابی کنیم. به عبارت دیگر، هوش رابطه بین متغیرهای مورد نظر ما را درهم می‌آمیزد.

واژه «درآمیختگی» به «مخلوط شدن متغیرهای نامربوط با مسئله تحقیق، با متغیرهای مستقل مورد بررسی» بر می‌گردد، به نحوی که تأثیرات آنها تفکیک ناپذیر می‌شود. به طور روشن نمی‌توانیم ادعا کنیم که آیا رابطه به دست آمده، ۱- بین متغیرهای مستقل و وابسته مورد بررسی ۲- بین متغیرهای نامربوط و متغیر وابسته و یا ۳- ترکیبی از موارد ۱ و ۲ است. عامل درآمیختگی را می‌توان با کنترل اثر متغیرهای نامربوط حذف کرد.

از سوی دیگر، از آنجا که به عنوان مثال رابطه‌ای بین اندازه کفش دانش آموزان و توانایی آنها در یادگیری مفاهیم علمی دیده نشده است، آزمایشگر نیازی ندارد که خود را ملزم سازد تا اندازه کفش را کنترل کند.

در درجه اول باید سعی خود را به کنترل هرگونه تفاوت‌های از قبل موجود بین آزمودنیها، معطوف سازیم. تنها از این طریق است که می‌توان تا حدود قابل قبولی مطمئن شد که هر تفاوتی که بعد از آزمایش مشاهده می‌شود، به دلیل شرایط آزمایش، و نه تفاوت‌های از قبل موجود بین آزمودنیها است. معمولاً برای افزایش همسانی گروههایی که در معرض موقعیتهای

گونگون آزمایش قرار می‌گیرند، ۵ روش اصلی به کار برده می‌شود. ۵ روشی که برای کنترل تفاوت‌های بین آزمودنیها به کار برده می‌شوند، عبارتند از: ۱- گمارش تصادفی ۲- هم‌تاسازی تصادفی ۳- انتخاب همگن ۴- تحلیل کوواریانس و ۵- استفاده از آزمودنیها به عنوان کنترل خود.

### // کنترل تفاوت‌های بین آزمودنیها

#### / گمارش تصادفی (تخصیص تصادفی)

به کار آزمایشگر نگاهی می‌کنیم. او تعدادی آزمودنی در دسترس دارد که می‌باید آنها را به دو گروه تقسیم کند. این گروهها، در موقعیتهای متفاوت قرار می‌گیرند و سپس با یکدیگر مقایسه می‌شوند. او برای تعیین اینکه چه کسی در چه گروهی قرار گیرد، نیاز به سیستمی دارد که به صورت مستقل از قضاوت شخصی او و نیز ویژگیهای خود آزمودنیها، عمل کند. برای مثال، او نباید کسانی را که می‌داند نمرات بالایی دارند، در گروه A و آنهایی را که نمرات پایینی دارند، در گروه B قرار دهد. سیستمی که این ضرورت را برآورده می‌کند، گمارش تصادفی است. گمارش تصادفی عبارت است از تخصیص آزمودنیها به گروهها، به نحوی که از نظر قرار گرفتن در هریک از گروهها، برای هریک از اعضای جامعه احتمال مساوی وجود داشته باشد. واژه تصادفی کردن معمولاً مترادف واژه گمارش تصادفی به کار برده می‌شود.

باید توجه نمود که گمارش تصادفی همان انتخاب تصادفی نیست. انتخاب تصادفی (مراجعه کنید به فصل ۶) استفاده از شیوه تصادفی برای انتخاب یک نمونه از یک جامعه آماری است. گمارش تصادفی استفاده از شیوه تصادفی برای تخصیص آزمودنیها به تیمار است. برای مثال، خانم براون<sup>(۱)</sup>، برای آزمایشی، ۱۰۰۰ آزمودنی در اختیار دارد، ولی امکانات او برای آزمایش ۱۰۰ نفر، کافی است. او از طریق انتخاب تصادفی، ۱۰۰ نفر را برای شرکت در آزمون بر می‌گزیند تا از ایجاد تفاوت‌های نظام‌مند بین آزمودنیهای آزمایش و جامعه‌ای که آزمودنیها به آن تعلق دارند، اجتناب کند. همچنین از طریق گمارش تصادفی ۵۰ نفر از این ۱۰۰ نفر برای تیمار A و ۵۰ نفر بقیه برای تیمار B، از تفاوت‌های نظام‌مند قبل از تیمار، بین دو گروه حذر می‌کند.

به منظور دستیابی به نمونه‌های تصادفی، آزمایشگر می‌تواند به تمام آزمودنیها شماره بدهد

و تعداد مورد نیاز را از جدول اعداد تصادفی بیرون بیاورد. سپس با پرتاب یک سکه به صورت شیر یا خط می‌تواند تصمیم بگیرد که کدام گروه نمونه در معرض تیمار A و کدام در معرض تیمار B قرار گیرند.

چنانچه آزمودنیها برای گروههای آزمایشی به صورت تصادفی انتخاب شده باشند، گروهها را می‌توان از نظر آماری، هم ارز فرض نمود. هم ارزی آماری بدان معنی نیست که گروهها کاملاً یکسانند، بلکه بدان معناست که هرگونه تفاوت موجود بین گروهها تابع احتمال است و ناشی از قضاوت مغرضانه آزمایشگر، انتخاب خود آزمودنیها و یا عامل دیگری نیست. یک آزمودنی با هوشبهر بالا بخت یکسانی برای قرار گرفتن در معرض تیمارهای A یا B دارد. همین امر در مورد آزمودنی با هوشبهر پایین نیز صادق است. حال اگر کل نمونه را در نظر بگیریم، اثر هوش بر روی متغیر وابسته به سوی تعادل می‌رود و یا در اثر انتخاب تصادفی، از بین می‌رود. به همین نحو، تفاوت‌های آزمودنیها از نظر دیدگاههای سیاسی، منش، انگیزش و دیگر ویژگیها نیز تقریباً به صورت همسان در دو گروه توزیع می‌شوند.

چنانچه شیوه گمارش تصادفی را به کار بگیریم، تفاوت‌های از قبل موجود بین گروهها، تنها تابع احتمال خواهد بود. در چنین حالتی، می‌توان از آمار استنباطی برای تعیین این امر که تا چه حد احتمال دارد تفاوت‌های بعد از تیمار ناشی از عامل تصادف باشد، استفاده کرد. لازم است خاطر نشان کنیم که می‌توان انتظار داشت نه تنها متغیرهای خارجی شناخته شده، بلکه دیگر متغیرهای خارجی اثرگذار ناشناخته برای آزمایشگر نیز احتمالاً به طور تصادفی پخش شوند.

### / همتا کردن تصادفی

روش جایگزینه‌ای برای تخصیص آزمودنیها به گروههای آزمایشی، عبارت است از همتا کردن آزمودنیها به تعداد متغیرهای نامربوطی که محقق فکر می‌کند احتمالاً بر متغیر وابسته تأثیر خواهند گذاشت. سپس در مرحله بعد، برای تخصیص هر عضو زوجهای همتا شده به یکی از گروههای آزمایشی، از یک روش تصادفی استفاده می‌شود. چنانچه از نظر این متغیرها، گروهها به صورت یکنواخت همتا شده باشند، می‌توان مطمئن بود که هر تفاوتی که بعد از آزمون مشاهده شود، در اثر تیمار آزمایشی است.

هرچند همتا کردن، روشی مؤثر برای کنترل تفاوت‌های بین آزمودنیها است، اما محقق احتمالاً

با پاره‌ای از مشکلات برخورد خواهد کرد. اولین مشکل، تصمیم‌گیری بر روی انتخاب متغیر یا متغیرهایی است که باید برای همتا کردن از آنها استفاده کرد. متغیرهایی مانند هوشبهر، سن تقویمی، موقعیت اجتماعی، جنس، سن، نمره خواندن و نمره پیش‌آزمون می‌تواند مورد استفاده قرار گیرد. متغیرهایی که بر مبنای آن آزمودنیها همتا می‌شوند، باید با متغیر وابسته همبستگی داشته باشند، در غیر این صورت همتا کردن بی‌مورد است. به عنوان یک اصل کلی، پیشنهاد می‌کنیم چنانچه ضریب همبستگی متغیری با متغیر وابسته  $0/50$  یا بالاتر نباشد، نباید آن را برای روش همتا کردن به کار برد، زیرا کمک قابل ملاحظه‌ای به دقت عمل بررسی نمی‌کند. به صورت ایده‌آل، دو یا چند متغیر را که با متغیر وابسته همبستگی خوبی دارند، ولی با یکدیگر همبستگی معنی‌داری ندارند، همتا خواهیم کرد. به هر حال، زمانی که می‌خواهیم بیش از دو متغیر را برای همتا کردن در نظر داشته باشیم، یافتن آزمودنیهایی که بر مبنای چند متغیر، بخوبی همتا شوند، تقریباً غیر ممکن می‌شود. بنابر این، به دلیل عدم امکان یافتن همتهای مناسب، تعدادی از آزمودنیها را از دست خواهیم داد.

سؤال دیگری که در این زمینه مطرح است، این است که آزمودنیها باید تا چه حد از نظر متغیر یا متغیرها به یکدیگر نزدیک باشند. همتا کردن نزدیک، به افزایش دقت روش کمک می‌کند، ولی در عین حال تعداد آزمودنیهایی را نیز که نمی‌توانند همتا شوند، زیاد می‌کند. این امر، مسلماً از حجم نمونه می‌کاهد و نمونه‌گیری بررسی ما را اریب می‌سازد.

**روشهای همتا کردن** محقق باید در مورد اینکه در موقعیت خاص بررسی او، کدام روش همتا سازی عملی است، تصمیم‌گیری کند. روش متداول، روش فرد به فرد است، که بر طبق آن سعی می‌شود دو فردی را که نمره هایشان در بین آزمودنیهای قابل دسترس در محدوده‌ای قبل تعیین شده قرار می‌گیرد، در دو گروه جای دهند. به عنوان مثال، چنانچه متغیر همتا سازی، هوشبهر باشد، محقق دو فردی را که مثلاً نمره هوشبهر آنها ۵ نمره تفاوت داشته باشد، انتخاب می‌کند و به صورت تصادفی هر کدام را در یکی از گروههای تیمار A و B قرار می‌دهد. اگر بهره هوشی، تنها متغیر همتا سازی باشد، کار محقق خیلی مشکل نیست، ولی اگر جنس و طبقه اجتماعی هم متغیرهای مربوط باشند، کار همتا کردن آزمودنیها بر طبق سه متغیر، بسیار مشکل خواهد بود. محقق در حقیقت آزمودنیهایی را که یافتن همتهای برای آنها غیر ممکن باشد، از دست می‌دهد.

به خاطر داشته باشید که برای همتا کردن تصادفی، باید نخست آزمودنیها را همتا کرد و سپس به صورت تصادفی در معرض تیمار قرار داد. بررسی ای را که در آن دو گروه از آزمودنیها، که قبلاً در معرض دو تیمار متفاوت قرار داشته‌اند، با هم همتا می‌شوند، نمی‌توان به عنوان یک بررسی آزمایشی تلقی کرد. چنین بررسیهایی (در فصل ۱۰ در مورد آنها بحث شده است) که در آن همتا کردن صورت می‌گیرد، لکن گمارش تصادفی در گروهها انجام نمی‌شود، می‌تواند محقق را به نتیجه گیریهای اشتباه هدایت کند.

### / انتخاب همگن

روش دیگری که می‌تواند گروهها را از نظر متغیرهای خارجی قابل مقایسه سازد، انتخاب نمونه‌ها به گونه‌ای است که تا حد امکان همگن باشند. اگر آزمایشگر گمان برد که متغیر سن بر متغیر وابسته تأثیر می‌گذارد، می‌تواند کودکان هم سن را انتخاب کند. برای مثال، او با انتخاب بچه‌های ۶ ساله می‌تواند تأثیر سن را به عنوان یک متغیر مستقل خارجی کنترل کند، یا به همین ترتیب وقتی احتمال بود که عامل هوش بر روی متغیر وابسته یک بررسی تأثیر می‌گذارد، محقق می‌تواند کودکان را در یک محدوده هوشبهر معین، به عنوان مثال ۱۱۰-۱۰۰، انتخاب کند. با این روش، تأثیرات عامل هوشبهر کنترل می‌شود، سپس آزمایشگر می‌تواند افراد همگن را به صورت تصادفی در گروهها توزیع کند و مطمئن باشد که از نظر هوشبهر قابل قیاس هستند. آغاز کار با گروهی که از نظر یک متغیر مرتبط همگن باشند، موجب می‌شود که مشکل همتا کردن آزمودنیها بر مبنای آن متغیر، برطرف شود.

با وجود آنکه روش انتخاب همگن روشی مؤثر در کنترل متغیرهای خارجی است، اما دارای این نقیصه نیز هست که میزان تعمیم را به موقعیتهای دیگر کاهش می‌دهد. چنانچه آزمایشگری در مورد میزان تأثیر یک روش آموزش بر روی یک نمونه همگن، برای مثال کودکان با هوشبهر متوسط، بررسی کرده باشد، نمی‌تواند یافته‌هایش را به کودکان با هوشبهر متفاوت تعمیم دهد. او نمی‌تواند در مورد میزان مؤثر بودن این روش بر روی کودکان با هوشبهر پایین یا بالا اطلاعی داشته باشد. آزمایش باید با آزمودنیهایی با هوشبهرهای متفاوت تکرار شود.

همانند همتا سازی، یک آزمایش واقعی نیازمند آن است که ابتدا آزمودنیها انتخاب شوند و سپس به صورت تصادفی برای قرار گرفتن در معرض تیمارها، تخصیص داده شوند.

## / تحلیل کوواریانس

تحلیل کوواریانس (ANCOVA) روشی آماری است، که دقت طرح تحقیق را با در نظر گرفتن متغیرهای از قبل موجود، که با متغیر وابسته مربوطند، افزایش می‌دهد.<sup>(۱)</sup> برای مثال، آزمایشی را که اثرات دو روش آموزش خواندن را مورد بررسی قرار می‌دهد، در نظر بگیرید. در ابتدای نیمسال، یک پیش‌آزمون انجام می‌شود و سپس نیمی از آزمودنیها برای آموزش دیدن با روش الف و نیمی دیگر برای آموزش دیدن با روش ب تخصیص داده می‌شوند. متغیر مستقل، روش آموزش خواندن و متغیر وابسته، نمره پس‌آزمون خواندن است. پیش‌آزمون خواندن، متغیر کمکی و متغیری است که به متغیر وابسته مربوط است و به منظور افزایش دقت طرح، مورد استفاده قرار می‌گیرد.

انتظار می‌رود که دانش آموزانی که از توانایی خواندن بالایی برخوردارند، در هر دو پیش‌آزمون و پس‌آزمون نمرات بالایی بگیرند و آنهایی که از این جهت ضعف دارند، نمرات پیش‌آزمون و پس‌آزمونشان هر دو پایین باشد، در نتیجه همبستگی مثبتی بین نمرات پیش‌آزمون و پس‌آزمون وجود خواهد داشت. روش ANCOVA آن بخش از نمره پس‌آزمون آزمودنی را که با نمره پیش‌آزمون او مشترک است، حذف می‌کند.

با استفاده از این روش آماری، تنها نمره پس‌آزمون آزمودنی را در نظر نمی‌گیریم، بلکه با در نظر گرفتن نمره پیش‌آزمون و همبستگی پس‌آزمون و پیش‌آزمون، قادر خواهیم بود که تفاوت بین نمرات پس‌آزمون و آنچه را که انتظار داریم نمره پس‌آزمون باشد، تحلیل کنیم.

از طریق حذف آن بخش از واریانس نمره متغیر وابسته که به صورت نظام‌مند با واریانس پیش‌آزمون مرتبط است، دقت عمل آزمایش افزایش می‌یابد. اگر آن بخش واریانس نمرات پس‌آزمون که ناشی از تیمار نیست، حذف شود، تفاوتی که ناشی از تیمار است، به صورت بارزتر مشخص می‌شود. استفاده از یک متغیر کمکی، که به متغیر وابسته مربوط است، موجب می‌شود که احتمال خطای نوع دوم کاهش یابد.

علاوه بر نمره پیش‌آزمون، اندازه‌های دیگری نیز می‌توانند به عنوان متغیر کمکی مورد استفاده قرار گیرند. در مثال ما، نمرات آزمون هوش و یا معدل کلاسی می‌تواند به کار گرفته شود.



متغیر کمکی برای آنکه کاربرد داشته باشد، باید با متغیر وابسته همبستگی داشته باشد. در مثال فوق، نمرهٔ پرتاب توپ بیس بال و یا اندازهٔ کفش نمی تواند به عنوان متغیر کمکی تلقی شود، زیرا انتظار می رود این اندازه ها همبستگی قابل اغماضی با نمرات خواندن داشته باشند.

### / استفاده از آزمودنیها برای کنترل خود

روش دیگر، مستلزم تخصیص گروهی از آزمودنیها به تمام شرایط آزمایشی است. در این روش، ابتدا آزمودنیها را تحت یک تیمار آزمایشی قرار می دهند و اندازه گیریهای لازم را به عمل می آورند و سپس همان آزمودنیها را تحت تیمار دیگر قرار می دهند و اندازه گیریها را تکرار می کنند. به طور کلی، تیمار آزمایشی شامل مقادیر انتخابی یک متغیر مستقل است. برای مثال، از آزمودنیها خواسته می شود که دو فهرست هجاهای بی معنا و یک فهرست با میزان تداعی بالا و یک فهرست با میزان تداعی پایین را یاد بگیرند. تفاوت زمان مورد نیاز برای یادگیری دو فهرست برای هر آزمودنی و متوسط تفاوتهای زمان یادگیری برای تمام آزمودنیها به دست می آید و آزمون معنی داری تفاوتها صورت می گیرد.

چنانچه امکان کاربرد این روش وجود داشته باشد، روشی مؤثر برای کنترل خواهد بود، ولی در مواردی نمی توانیم از این روش استفاده کنیم. در پاره ای از بررسیها، نمی توان آزمودنیها را پس از قرار دادن در یک موقعیت آزمایشی در موقعیت دیگر قرار داد. نمی توانیم به کودکان، تقسیم کردن متعارفی را با یک روش آموزش دهیم و سپس آنچه را که یاد گرفته اند، از حافظه شان پاک کنیم و با روش دیگری آن را آموزش دهیم.

در آزمایش یاد شده، که در آن از آزمودنیها برای بررسی میزان آسانی نسبی یادگیری هجاهای بی معنا با تداعی در حد بالا و پایین استفاده می شد، می توانست یک تأثیر «یادگرفتن یادگیری» وجود داشته باشد و در نتیجه هر فهرستی که در نوبت دوم مورد یادگیری قرار می گرفت، از امتیازی نسبت به اولی برخوردار می شد و بالعکس، تأثیر عامل خستگی یا تداخل هم می تواند در عملکرد ضعیف تر در مورد فهرست دوم مؤثر باشد. در هر دو مورد، نمی توانیم اثر متغیر مستقل را از اثر ترتیب انجام آزمون بر روی متغیر وابسته جدا سازیم. روش مؤثر می تواند تقسیم تصادفی آزمودنیها به دو گروه باشد: گروه اول، هجاهای با تداعی بالا و گروه دوم، هجاهای با تداعی پایین را یاد می گیرند. با این روش، اثرات یادگرفتن یادگیری و یا خستگی، «متعادلی» می شود.

به هر حال، چنانچه یادگیری هجاهای با تداعی بالا در مرحله اول، به یادگیری هجاهای با تداعی پایین در مرحله بعد کمک کند، اما عکس این عمل صادق نباشد، این امر می تواند تفسیر نتایج را مخدوش سازد.

### // کنترل تفاوت‌های وضعیت

علاوه بر تفاوت‌های فردی، هرگونه متغیر نامربوط نیز که ممکن است بر روی موقعیت آزمایشی اثر بگذارد، باید کنترل شود. اگر متغیرهای مربوط به موقعیت در یک آزمایش کنترل نشوند، نمی توان مطمئن شد که آیا این متغیر مستقل است که باعث به وجود آمدن تفاوت در متغیر وابسته می شود یا تفاوت‌های اتفاقی موجود در بین افراد گروه.

برای مثال، فرض کنیم محققی می خواهد بر روی تأثیر یک فیلم در تغییر نگرش نسبت به موضوع خاصی مطالعه کند. او گروهی را به صورت تصادفی از بین کودکان کلاس انتخاب می کند و برای دیدن فیلم می فرستد و گروه مقایسه‌ای در کلاس باقی می ماند. احتمال دارد که او به صورت ناخودآگاه نیروهای بی شماری را دخالت داده باشد. گروه گواه ممکن است برنجند، یا احساس طردشدگی پیدا کنند و یا نسبت به دیگران احساس حقارت بنمایند. هر یک از این عوامل می تواند در نتایج بررسی تأثیر بگذارد. تغییری را که او می خواهد به فیلم نسبت بدهد، ممکن است در حقیقت در اثر یکی از این حالات اتفاقی به وجود آمده باشد. در این مورد، محقق باید تدابیری را اتخاذ کند تا آزمودنی‌های گروه گواه هم، نظیر گروه آزمایشی، فیلمی را تماشا کنند؛ و یا اعضای هر دو گروه بدانند که در یک آزمایش شرکت کرده‌اند، یا هیچیک از دو گروه از این امر آگاه نباشند. این اقدام از جهت کنترل تأثیری موسوم به هاتورن، مؤثر است. این اصطلاح به پدیده‌ای اطلاق می شود که اولین بار در کارخانه هاتورن، وابسته به شرکت وسترن الکتریک<sup>(۱)</sup>، مشاهده شد و نشان داد که تقریباً هر تغییری، هرگونه توجه اضافی، هر دستکاری آزمایشی و حتی اطلاع از اینکه آزمایشی در حال انجام است، کافی است که در آزمودنیها تغییر ایجاد کند. به طور خلاصه، هرگونه توجهی می تواند در آزمودنیها نوعی پاسخ برانگیزد.

در مطالعه‌ای که بر روی تأثیر یک دارو بر روی نحوه عمل مهارت یدی صورت می گیرد، کلیه آزمودنیها باید به گونه‌ای تصور کنند که دارو خورده‌اند. در نتیجه محقق باید به تمام

آزمودنیها یک نوع داروی مشابه، از نظر کمیت، مزه و رنگ، بدهد. این ماده، در گروهی همان داروی مورد مطالعه و در گروه دیگر، یک ماده شیمیایی بی اثر است که به آن دارونما<sup>(۱)</sup> می‌گویند. در غیر این صورت، تنها علم به اینکه به آنها دارو داده شده است، می‌تواند بر روی آزمودنیها تأثیر بگذارد و یا آنها را بیش از حد محتاط و یا برعکس کاملاً بی قید کند. در نتیجه محقق نمی‌تواند بفهمد که آیا نتایج به دست آمده، ناشی از تأثیر دارو و یا نگرش آزمودنی و یا هر دو بوده است.

معمولاً سه روش متداول برای کنترل متغیرهایی که در پاره‌ای شرایط به صورت بالقوه تحت تأثیر متغیرهای دیگر هستند، وجود دارد.<sup>(۲)</sup> محقق می‌تواند: ۱- آنها را ثابت نگاه دارد، ۲- آنها را به طور تصادفی توزیع کند و ۳- آنها را به صورت نظام‌مند و جداگانه از متغیر مستقل اصلی، دستکاری کند.

ثابت نگاه داشتن متغیرهای نامربوط بدین معنا است که با گروههای مختلف از همه نظر، به استثنای قرار گرفتن در معرض تیمار مستقل، یکسان رفتار شود. به عنوان مثال، در تحقیقی در مورد خواندن، لازم است که تعداد افراد هر گروه، کنترل شود، زیرا ثابت شده است که بزرگی گروه، عاملی مؤثر در پیشرفت خواندن است. باید مطمئن شد که تعداد افراد گروه آزمایشی و گروه گواه برابر هستند. متغیر معلم نیز باید کنترل شود، زیرا توانایی و جدیت معلم عواملی هستند که می‌توانند در نتایج تحقیق بر روی خواندن، اثر بگذارند. بنابراین، یک معلم باید با دو روش مورد مقایسه آموزش بدهد.

در یک آزمایش، دستیاران مختلف باید روش یکسانی را از نظر آموزش، وسایل و آزمونها، دنبال کنند. سعی کنند نگرش یکسانی را نسبت به تمام گروهها داشته باشند. تمام گروهها باید در ساعت مشابهی از روز و در یک اتاق واحد گرد هم آیند. نباید گروه آزمایشی در ساعت اول مدرسه آموزش ببینند و آزمون شوند و گروه گواه در ساعت آخر. به همین نحو، شرایط محیط، درجه حرارت، شدت نور، رطوبت، مبلمان اطاق، وجود یا نبود صداهای اختلالگر باید برای همه گروهها یکسان باشد.

چنانچه ثابت نگاه داشتن شرایط امکان پذیر نباشد، محقق باید در جهت تصادفی کردن و یا متعادل کردن متغیرهای موقعیتی ویژه، کوشش کند. به عنوان مثال، چنانچه امکان نداشته باشد

که یک معلم دو روش را آموزش دهد، محقق می‌تواند دو گروه اصلی را به دو گروه کوچکتر تقسیم کند و به طور تصادفی نیمی از گروه الف و نیمی از گروه ب را به هریک از معلمان واگذار کند. می‌توان در مورد دیگر شرایط آزمایشی، مانند وسایل، نیز به همین نحو عمل کرد. از این طریق، متغیرهای وابسته به وضعیت، به صورت تصادفی توزیع می‌شوند. شرایط نامربوط گوناگونی وجود دارند، ولی به آنها اجازه تأثیر گذاری نظام‌مند بر روی متغیر وابسته داده نمی‌شود.

امکان کنترل متغیرهای موقعیتی نامربوط از طریق دستکاری نظام‌مند آنها وجود دارد. در بسیاری آزمایشهای آموزشی، لازم است که به منظور کنترل تأثیرات افزایشده، مانند تمرین و خستگی، از یک توالی شرایط آزمایشی و کنترل استفاده شود. این امر با کنترل ترتیبی که در آن شرایط آزمایشی به طور متعادل ارائه می‌شود، امکان پذیر است. نیمی از آزمودنیها ممکن است با ترتیب الف - ب و نیم دیگر با ترتیب ب - الف مورد آزمایش قرار گیرند. در این مورد، یک متغیر نامربوط به صورت نظام‌مند دستکاری شده است. این روش نه تنها تأثیر بالقوه مخرب ترتیب را کنترل می‌کند، بلکه می‌تواند برآوردی از اندازه تأثیر ترتیب، با تعیین این که آیا متوسط مقادیر الف و ب حاصله در دو توالی متفاوت هستند یا خیر، به دست دهد.

باید خاطر نشان کرد که دیگر انواع متغیرها را می‌توان مانند آنهایی که با خود آزمودنیها همبسته هستند، در طرح تحقیق آزمایشی در نظر گرفت و در نتیجه کنترل کرد. برای مثال، چنانچه کنترل عامل جنسیت در تحقیقی لازم شود و امکان کاربرد هیچیک از روشهای کنترل تفاوت‌های بین فردی وجود نداشته باشد، می‌توان جنسیت را به عنوان یک متغیر مستقل در نظر گرفت و برای تحلیل از روش تحلیل واریانس دو عاملی استفاده کرد. با این عمل نه تنها متغیرهای نامربوط کنترل می‌شوند، بلکه اطلاعاتی در مورد اثرات این متغیر بر روی متغیر وابسته و همچنین تعامل احتمالی آن با دیگر متغیرهای مستقل به دست می‌آید.

این روش کنترل متغیرهای نامربوط، همان افزودن متغیرهای مستقل بیشتر به آزمایش است. اگرچه با این عمل، به پیچیدگی آزمایش افزوده می‌شود، اما این عمل از مزیت به دست آوردن اطلاعات اضافی در مورد اثرات متغیرهای مربوط بر روی متغیر وابسته و تعاملهای آنها با متغیرهای مستقل اصلی برخوردار است. استفاده از این روش کنترل، با ورود کامپیوترهای الکترونیکی برای انجام تحلیل اطلاعات بررسیهای پیچیده، رو به افزایش نهاده است. بهره‌گیری

از طرحهای تحلیل واریانس دو عاملی و یا طرحهای پیچیده تر تحلیل واریانس، بررسی همزمان تعدادی متغیر را به صورت جداگانه و همچنین در تعامل با یکدیگر، فراهم می سازد. تحلیل تعامل متغیرهای یک بررسی، معمولاً مهمترین بخش بررسی را تشکیل می دهد.

### // دستکاری آزمایشی

ویژگی متمایز کننده تحقیق آزمایشی، همانا دستکاری متغیرهاست. دستکاری متغیر، به عمل بخردانه ای که توسط محقق صورت می گیرد، اطلاق می شود. در تحقیقات آموزشی و دیگر علوم رفتاری، دستکاری متغیرها شکل ویژه ای می یابند که در آن آزمایشگر، آزمودنیها را در معرض تعدادی شرایط گوناگون از قبل تعیین شده قرار می دهد. این شرایط گوناگون از قبل تعیین شده را متغیر مستقل، متغیر آزمایشی و متغیر تیمار می خوانیم. برای ارائه دو یا چند مقدار از متغیرهای مستقل، شرایط متفاوتی طراحی می شوند. این تفاوتها می توانند در میزان یا نوع متغیر مستقل باشند. ما می توانیم یک متغیر را تغییر دهیم و یا به صورت همزمان در چند متغیر، تغییر ایجاد کنیم.

### // مشاهده

در آزمایشها اثر دستکاری متغیر مستقل بر روی متغیر وابسته مورد توجه است. مشاهدهها با در نظر گرفتن پاره ای از ویژگیهای رفتاری آزمودنیهای تحقیق صورت می گیرد. متغیر وابسته، مشاهداتی هستند که در صورت امکان از نظر ماهیت کمی می باشند. متغیر وابسته در تحقیق آموزشی، اغلب گونه ای از پیشرفت تحصیلی، مانند «یادگیری» است. معمولاً ما می خواهیم پیشرفت تحصیلی را توصیف یا پیش بینی کنیم. به خاطر داشته باشید که ما مستقیماً نمی توانیم یادگیری را اندازه گیری کنیم. ما تنها می توانیم از طریق اندازه هایی چون نمرات یک آزمون یادگیری را برآورد کنیم. بنابر این، چنانچه بخواهیم دقیق تر بگوئیم، متغیرهای وابسته نمره ها و مشاهدات هستند و نه پیشرفت تحصیلی به خودی خود.

### /// مقایسه آزمایشی

در ساده ترین نوع آزمایش، به دو گروه آزمودنی نیاز است. این دو گروه به نام گروه آزمایشی

و گروه گواه خوانده می شوند. یکی از این گروهها تحت یک نوع تیمار قرار می گیرند، در حالی که گروه دیگر در معرض هیچگونه تیماری قرار نمی گیرد. در بررسی اثر یک دارو، گروه آزمایشی دارو را دریافت می کنند، در حالی که گروه گواه یا اصلاً دارویی دریافت نمی کنند و یا نوعی از دارونما را می گیرند. استفاده از گروه گواه این امکان را به محقق می دهد که بسیاری از تبیینهای جایگزین را حذف کند. برای مثال، محصول فراوان مزرعه ای که کود داده شده است، می تواند نتیجه هوای معتدل و یا عوامل دیگر باشد. حال چنانچه مزرعه ای مشابه این مزرعه و در همسایگی آن، که کود داده نشده است، محصول کمتری بدهد، می توان اثر کود را در میزان محصول قابل ملاحظه دانست.

برای مثال، مقایسه پیشرفت هجی کردن یک گروه آزمایشی که با روش الف آموزش دیده اند با گروه گواه، که اصلاً آموزش هجی کردن نداشته اند، امری بیهوده است و به جای آن، بین گروههایی که در معرض تیمار روش الف و ب هستند، مقایسه صورت می گیرد. مقایسه گروههایی که دو تیمار مختلف دریافت می کنند، همان کنترلی را در مورد تبیینهای جایگزینهای فراهم می آورد که وقتی دو گروه را، که یکی تیمار را دریافت می کند و دیگری دریافت نمی کند، با هم مقایسه کنیم. به منظور تسهیل مباحث بعدی، واژه گروه گواه را برای هر دو گروهی که یا تحت تیماری نبوده اند، و یا تحت تیمارهای دیگری بوده اند، به کار می بریم. دربررسیهای علمی، مقایسه امری ضروری است. از طریق مقایسه گروهی که در معرض تیمار هستند با گروه مشابهی که تحت تیماری نیستند یا گروه یا گروههایی که تحت تیمار دیگری هستند، می توان به نتیجه گیریهای متقن دست یافت.

گروههای آزمایشی و گواه از نظر کلیه عواملی که می توانند بر روی متغیر وابسته اثر داشته باشند، باید هم ارز باشند و تنها از جهت قرار گرفتن در معرض متغیر مستقل می توانند تفاوت داشته باشند. محقق پس از قرار دادن آزمودنیهایش در موقعیتهای مختلف، بر روی هر یک از آنها، از نظر متغیر وابسته، اندازه گیری صورت می دهد.

قدم بعدی، ارزشیابی است. آیا تفاوتی بین دو گروه وجود دارد؟ آیا اثر تیمار الف با اثر تیمار ب متفاوت است؟ این سؤال مقایسه اندازه های وابسته را در یک گروه با اندازه های پاسخ در گروه دیگر، ضروری می سازد. مقایسه باید برای آزمایشگر روشن کند که آیا تفاوتی مشاهده شده در متغیر وابسته، با تفاوتی مستقل که به صورت دو موقعیت مختلف الف و ب ارائه شده، همبسته است؟

## // آزمایشهای آزمایشگاهی و آزمایشهای میدانی

آزمایش در زمینه تعلیم و تربیت می‌تواند در هر دو شرایط آزمایشگاهی و یا میدانی اجرا شود. کنترل متغیر نامرتبط، که در روش آزمایشی امری حیاتی است، با بالاترین دقت در محیط آزمایشگاه امکان پذیر است. آزمایشگر می‌تواند در شرایط آزمایشگاهی، محیط را به گونه‌ای کنترل کند که متغیرهای مستقل جدا شوند. در نتیجه این شیوه از نظر ارائه تعاریف علمی متغیر، بسیار دقیق است. به همین دلایل، امکان اجرای دوباره آزمایشهای آزمایشگاهی با دقت عمل بالا، وجود دارد.

آزمایشهای میدانی می‌تواند در کلاسهای درس، زمینهای بازی، گردهمایی‌های باشگاهها و یا دیگر موقعیتهای طبیعی اجرا شود. آزمایشگر ضمن دستکاری متغیرهای مستقل، تا آنجا که ممکن باشد متغیرهای بیرونی را کنترل می‌کند، اما در آزمایش میدانی، ناگزیر، کنترل ناقص‌تر خواهد بود. به هر حال آزمایشهای میدانی نیز از مزایای خاصی بهره‌مند هستند. اول آنکه، متغیرهای آزمایش می‌تواند در آزمایشهای میدانی به مراتب قوی‌تر از آزمایشهای آزمایشگاهی باشند. قرار دادن آزمودنی تحت یک تیمار در شرایط آزمایشگاهی برای بیش از مدتی کوتاه، مشکل است، در حالی که آزمایش میدانی می‌تواند جلسات روزانه را برای تمام سال تحصیلی، در بر بگیرد. دوم آنکه، به دلیل اجرای آزمایشهای میدانی در شرایط واقعی‌تر، احتمال ارائه راه‌حلهایی برای مشکلات روزمره عملی دست‌اندرکاران تعلیم و تربیت بیشتر است.

از نظر مسائل نظری، آزمایشهای آزمایشگاهی و از دید مسائل علمی، آزمایشهای میدانی ارجحیت دارند. دو نوع کلی آزمایش تحت شرایط کلاسی انجام می‌شود: ۱- بررسی روشها، که در آنها دو یا چند راه انجام کار به شیوه‌ای بی طرفانه، مقایسه می‌شوند و ۲- تحقیق بنیادی، که هدف آن رسیدن به اصولی کلی است که در هر شرایطی کاربرد داشته باشند. به هر دو نوع تحقیق کلاسی، شدیداً نیاز است.

## /// طرح آزمایشی

طرح آزمایشی عبارت است از یک چهارچوب مفهومی که آزمایش در آن چهارچوب صورت می‌گیرد. یک طرح آزمایشی شامل دو عملکرد است: ۱- مقایسه‌هایی را که طبق فرضیه‌های تحقیق ضروری است، امکان پذیر می‌سازد. ۲- آزمایشگر را قادر می‌سازد تا از طریق تحلیل آماری یافته‌ها، توصیف معنی داری برای نتایج بررسی‌اش ارائه دهد. چنانچه طرحی با

هدف رسیدن به این عملکردها برنامه ریزی شده باشد، محقق باید آن را با در نظر گرفتن پاره‌ای معیارهای کلی ذهنی، انتخاب کند. توصیف دقیق‌تر انواع مختلف طرح تحقیق، پس از بحثی در مورد معیارهای کلی، در همین فصل خواهد آمد.

مهمترین معیاری که باید در نظر گرفته شود، مناسب بودن طرح برای آزمون فرضیه موردنظر بررسی است. نشانهٔ یک بررسی پیشرفته، پیچیدگی یا سادگی آن نیست، بلکه مناسب بودن آن است. طرحی صحیح است که کاری را که باید انجام دهد، انجام بدهد. بنابر این اولین قدم آزمایشگر انتخاب طرحی است که به بهترین وجه، شرایط آزمایش را آماده کند تا نیازهای مسئلهٔ آزمایش خاص او را برآورده سازد.

چنانچه فرضیهٔ تحقیق از نوع فرضیهٔ تعاملی باشد، آزمون آن تنها از طریق طرح عاملی امکان پذیر است. متأسفانه به کرات دیده می‌شود که محققان در زمینهٔ تعلیم و تربیت برای آزمون فرضیهٔ تعاملی خود، از دو یا چند بررسی جداگانه استفاده می‌کنند. این نوع طرح، در حقیقت برای آزمون فرضیه ناتوان است.

فرض کنیم محقق می‌خواهد در مورد اثرات آموزش با بهره‌گیری از کامپیوتر بر روی یادگیری مفاهیم علوم مقدماتی دبستان مطالعه کند. او بر این باور است که اثر افتراقی این روش، احتمالاً بر اندازهٔ کلاس (تعداد دانش‌آموزان) و سطح هوشی دانش‌آموزان مبتنی است. برای حل این مسئله باید از طرح عاملی استفاده کرد. او برای پاسخگویی به مسئلهٔ مورد بررسی، نمی‌تواند دو یا سه بررسی جداگانه را، هر یک با یک متغیر مستقل، انجام دهد.

نمونه‌ای دیگر از یک طرح نامناسب، سعی در استفاده از طرح هم‌تا کردن آزمودنیها، در مواردی است که برای محقق هم‌تا سازی آزمودنیها از نظر کلیهٔ متغیرهای خارجی مربوط غیر ممکن است. حتی اگر او در هم‌تا کردن آزمودنیها از نظر یک یا دو متغیر موفق باشد، نمی‌تواند فرض کند که گروههای مورد بررسی‌اش از نظر کلیهٔ متغیرهای مربوط، هم‌ارز هستند. در چنین مواردی یک طرح توزیع تصادفی آزمودنیها ارجحیت دارد.

معیار دوم آن است که طرح باید کنترل کافی داشته باشد تا بتوان اثرات متغیر مستقل را تا حد امکان بدون ابهام، ارزشیابی کرد. تا زمانی که طرحی متغیرهای نامرتبط را کنترل نکند، نمی‌توان هرگز از رابطهٔ بین متغیرهای بررسی‌مطمئن بود. چنانچه قبلاً ذکر کردیم، توزیع تصادفی، بهترین و تنها راه دستیابی به کنترل مورد لزوم است. بنابر این، بهترین پیشنهاد این است که طرحی انتخاب شود که توزیع تصادفی را در جنبه‌های هرچه بیشتر به کار برد.



## // اعتبار طرحهای تحقیق

کمپبل و استنلی<sup>(۱)</sup>، سهمی بسزا در ارزیابی طرحهای تحقیق دارند. آنها دو ضابطه کلی را برای طرحهای تحقیق پیشنهاد می‌کنند: اعتبار درونی و اعتبار بیرونی.

### / اعتبار درونی

اعتبار درونی مربوط می‌شود به سؤالی نظیر: آیا تیمار آزمایشی واقعاً تغییری در متغیر وابسته ایجاد کرد؟ آیا متغیر مستقل تفاوت معنی داری به وجود آورد؟ این سؤالی اعتبار درونی را نمی‌توان به صورت مثبت پاسخ داد، مگر اینکه طرح به گونه‌ای باشد که کنترل کافی بر روی متغیرهای نامرتب داشته باشد؛ بدین معنا که چنانچه طرح بتواند متغیرها را کنترل کند، آزمایشگر قادر خواهد بود تبیین‌های جایگزین نتیجه مشاهده شده را حذف و آن را به گونه‌ای تفسیر کند که نوعی رابطه ذاتی را بین متغیرها نشان دهد. اعتبار درونی، اساساً مسئله کنترل است.

طرح کنترل‌های مناسب، مسئله ساختن راههایی برای حذف متغیرهای نامرتب است، یا به عبارت دیگر متغیرهایی که می‌توانند به تفسیرهای جایگزین منتهی شوند. هر چیزی که به کنترل یک طرح مربوط شود، به اعتبار درونی آن مربوط می‌شود.

کمپبل و استنلی هشت متغیر نامرتب را مشخص کرده‌اند که معمولاً اعتبار درونی یک طرح تحقیق را به خطر می‌اندازند. این متغیرها چنانچه کنترل نشوند، به احتمال زیاد می‌توانند تأثیری بگذارند که با تأثیر آزمایشی اشتباه گرفته خواهد شد.

۱- تاریخچه: حوادث ویژه بجز تیمار آزمایشی، ممکن است در فاصله زمانی اندازه‌گیری اول و دوم آزمودنیها اتفاق بیفتند و در متغیر وابسته تغییر ایجاد کند. برای مثال، در طول آزمایشی که به اندازه‌گیری میزان تأثیر گذاری یک واحد درسی بر چگونگی آشنایی با بازار بورس می‌پردازد، قیمت‌ها در بازار بورس سریعاً پایین می‌آید و در نتیجه تلویزیون و رادیو توجه خاصی را به بازار بورس معطوف می‌سازند.

محقق نمی‌تواند تعیین کند که آیا افزایش سطح معلومات دانشجو، نتیجه گذراندن آن واحد

1- Campbell, D.T., and J.C. Stanley (1966). *Experimental and Quasi-experimental Designs for Research* (p.5).

مؤلفان برای واژه‌شناسی و طراحی مورد استفاده در این قسمت از فصل، مدیون کتاب کمپبل و استنلی می‌باشند.

درسی است، یا حاصل مطالعه رسانه‌های گروهی. اثرات واحد درسی و قرارگرفتن دانشجویان در معرض رسانه‌های گروهی، متداخل و به صورت لاینحل مخلوط شده‌اند و در این مورد که یادگیری دانشجو تا چه حد در اثر گذراندن واحد درسی بوده و تاریخچه تا چه حد به عنوان یک مسئله اعتبار داخلی، در این امر مؤثر بوده، غیر ممکن است که بتوانیم نتیجه‌گیری کنیم.

۲- بلوغ: فرآیندهایی که در آزمودنیها تنها در اثر گذشت زمان به وجود می‌آید، می‌تواند موجب اثراتی شود که اشتباهاً به متغیر آزمایش نسبت داده می‌شود. آزمودنیها ممکن است تنها به علت مسن تر، گرسنه تر، خسته تر و یا کم‌انگیزه تر شدن از زمانی که برای اولین بار مورد آزمون قرار گرفتند، در اندازه‌گیریهای متغیر وابسته به صورت متفاوتی عمل کنند. برای مثال، اندازه‌گیری اثرات درمان مشکلات گفتاری در بین کودکان پیش دبستانی، دشوار است، زیرا کودکان خردسال به صورت طبیعی در اثر رشد بر این مشکل فائق می‌آیند.

۳- پیش آزمون: در معرض آزمون قرار گرفتن می‌تواند بر روی نحوه عمل آزمودنی در آزمون دوم، بدون توجه به تیمار آزمایشی اثر بگذارد. آزمودنیها می‌توانند موضوع بررسی را از یک پیش آزمون فراگیرند.

ممکن است عملکرد بهتر آزمودنیها در پس آزمون به دلیل آشنا شدن آنها با چهارچوب و فضاهای آزمایش باشد. احتمال دارد راهبردی برای انجام بهتر آزمون اتخاذ کرده باشند و یا در دومین آزمون، اضطراب کمتری داشته باشند. این اثرات تمرینی می‌تواند باعث شود که نمره پس آزمون، در حالی که متغیر مستقل هیچ تغییری نکرده است، بالا رود. انجام پیش آزمون در پرسشنامه‌های نگرش و شخصیت، می‌تواند آزمودنیها را برانگیزد تا به سوالات و مطالب پیش آزمون فکر کنند و در پس آزمون، پاسخهای متفاوتی بدهند.

۴- وسایل اندازه‌گیری: تغییر در وسایل اندازه‌گیری، یا در نمره دهندگان یا در مشاهده‌گران، ممکن است تغییراتی در نمره به دست آمده، پدید آورد. عواملی چون مشکلتر بودن پس آزمون و یا اندازه‌گیری قبل و بعد توسط دو مشاهده‌گر مختلف، ممکن است تغییراتی در دو نمره ایجاد کنند. اگر مشاهده‌گرانی که یک متغیر وابسته را ثبت می‌کنند، بدانند کدام آزمودنیها تیمار را دریافت کرده‌اند و کدام دریافت نکرده‌اند، ممکن است که آنها ناخودآگاه عملکرد گروه دریافت‌کننده تیمار را بیش از واقعیت و عملکرد گروه گواه را کمتر از واقعیت، برآورد کنند.

۵- رگرسیون آماری: چنانچه گروهها بر مبنای نمره‌های انتهایی انتخاب شده باشند،

رگرسیون آماری ممکن است تأثیری بگذارد که به عنوان اثر آزمایشی تفسیر شود. مقصود از تأثیر رگرسیون، گرایش نمره‌های انتهایی به برگشت یا حرکت به سوی میانگین مشترک در اندازه گیریهای بعدی است.

برای مثال، فرض می‌کنیم  $\frac{1}{4}$  از دارندگان نمره پایین در آزمون توانایی زبان انگلیسی، برای شرکت در یک برنامه آزمایشی ویژه زبان انگلیسی انتخاب می‌شوند. میانگین این گروه در آزمون دوم، چه تیمار آزمایشی اعمال شود و چه اعمال نشود، به سوی میانگین جامعه آماری گرایش پیدا خواهد کرد. به همین گونه، میانگینهای اولیه بالا در آزمون دوم، به پایین آمدن به سوی میانگین جامعه آماری، گرایش خواهند داشت.

حال رگرسیونی را در یک نمودار پراکنندگی (شکل ۱ - ۹) می‌بینیم که نشان دهنده الگویی است که اگر همبستگی نمرات آزمون خواندن کلاس چهارم و پنجم برابر  $r = 0.7$  باشد، آن را به دست خواهیم آورد. هر نقطه، نمایشگر دو نمره  $Z$  برای هر فرد است. چنانچه افرادی را با نمره  $Z$  خاصی ( $X$ ) در کلاس چهارم انتخاب و به نمرات کلاس پنجم آنها رجوع می‌کنیم، در می‌یابیم که به ندرت می‌توان افرادی را یافت که نمره  $Z$  آنها در هر دو آزمون یکی باشد. چنانچه نمره  $Z$  کلاس چهارم این گروه خاص بالای میانگین باشد، در خواهیم یافت که تعداد قلیلی از این دانش آموزان در کلاس پنجم میانگین نمره‌شان بالاتر از کلاس چهارم خواهد بود، لکن اکثریت آنها نمره  $Z$  نزدیکتر به میانگین و حتی زیر میانگین خواهند داشت.

برای زیرگروهی که نمره  $Z$  آنها در  $X$  به اندازه یک انحراف استاندارد از میانگین بالاتر است، میانگین آنها در  $Y$  برابر است با  $0.7$ ؛ برای زیرگروه دیگری با نمرات  $Z$  دو انحراف استاندارد زیر میانگین در  $X$ ، میانگین  $Y$  برابر  $0.1$  و برای گروهی که نمرات  $Z$  در  $X$  آنها  $+2$  است، میانگین نمره  $Z$  در  $Y$  برابر می‌شود با  $0.1$  و الی آخر. خط اریب در نمودار پراکنندگی، میانگینهای  $Y$  برای نمرات مختلف  $X$  را به هم متصل می‌کند. توجه داشته باشید که برای هر گروه با نمرات بالای میانگین در آزمون خواندن کلاس چهارم، نمره میانگین  $Y$  در آزمون خواندن کلاس پنجم پایین تر خواهد بود. در مورد هر گروهی که نمرات آنها زیر میانگین آزمون خواندن کلاس چهارم است، نمره میانگین در آزمون کلاس پنجم بالاتر خواهد بود. به عبارت دیگر، متوسط نمرات برای هر گروه به میانگین نزدیکتر می‌شود و این همان نتیجه رگرسیون است.

بدیهی است در صورتی رگرسیون پیش می‌آید که همبستگی بین دو متغیر کمتر از همبستگی

کامل باشد. از آنجا که عملاً هیچ یک از متغیرهای مورد نظر در تعلیم و تربیت کاملاً همبستگی ندارند، در طرح تحقیق خود همواره باید اثر رگرسیون را در نظر داشته باشیم. یک وجه رگرسیون را در مثلثهای قدیمی زیر مشاهده می‌کنیم: «وقتی در انتها قرار گرفته‌اید، چاره‌ای به جز بالا رفتن ندارید.» و «وقتی که در اوج قرار دارید، راهی به جز پایین رفتن ندارید.»

برای مثال، مدرسه‌ای یک برنامه جبرانی ریاضیات را برای دانش آموزان دوره راهنمایی، که دو سال نمره‌شان در آزمون استاندارد ریاضی پایین تر از سطح دوره راهنمایی بوده است، برگزار می‌کند. پس از یک نیم سال، به دانش آموزان آزمون استاندارد شده معادلی داده می‌شود. اکثریت آنان نمره‌ای نزدیکتر به سطح دوره راهنمایی می‌گیرند و به نظر می‌رسد که از برنامه جبرانی استفاده کرده‌اند. قبل از آنکه یادگیری آنها را به چیزی نسبت دهیم باید به خاطر داشته باشیم که وقتی در انتها قرار داریم، چاره‌ای به جز بالا رفتن نداریم.

پایین‌ترین نمرات در پیش آزمون شامل کسانی می‌شود که نمراتشان به علت بیماری موقتی، مشکلات عاطفی، عدم دقت و یا مشکلات دیگر پایین بوده است. همچنین شامل کسانی می‌شود که نمراتشان به علت اشتباه اندازه‌گیری پایین آمده بود. از اکثریت این افراد انتظار می‌رود که در آزمون بعدی عملکرد بهتری داشته باشند.

برعکس، چنانچه دانش آموزانی که بالاترین نمرات را در یک آزمون داشته‌اند، در معرض تیمار خاصی چون یک برنامه تقویتی قرار گیرند، احتمال دارد اندازه‌گیری بعد از تیمار نشان دهد که تیمار اثری منفی داشته است. آنهایی که در اوج قرار دارند، راهی به جز پایین رفتن ندارند.

۶- انتخاب افتراقی آزمودنیها: ممکن است حتی قبل از به کار بردن تیمار آزمایشی، بین گروهها تفاوت‌های مهم وجود داشته باشد. چنانچه گروه آزمایشی در یک آزمایش یادگیری باهوشتر از گروه گواه باشد، گروه آزمایشی می‌تواند در اندازه‌گیری متغیر وابسته، حتی اگر تیمار آزمایشی را هم دریافت نکرده باشد، بهتر عمل کند.

۷- میرایی آزمایشی: امکان دارد در طول بررسی، پاسخگویی را از گروههای مقایسه‌ای از دست بدهیم. چنانچه یک گروه آزمودنی با ویژگی خاص، طی بررسی از آزمایش خارج شود، این فقدان افتراقی می‌تواند در نتیجه بررسی اثر بگذارد. چنانچه برای مثال، چند نفر از دارندگان پایین‌ترین نمره‌ها در پیش آزمون به تدریج از آزمایش خارج شوند، میانگین این گروه در پس آزمون بالا خواهد رفت و این تغییر به علت تیمار آزمایشی نیست، بلکه دلیل آن عدم حضور

ضعیفترین آزمودنیها در پس آزمون است.

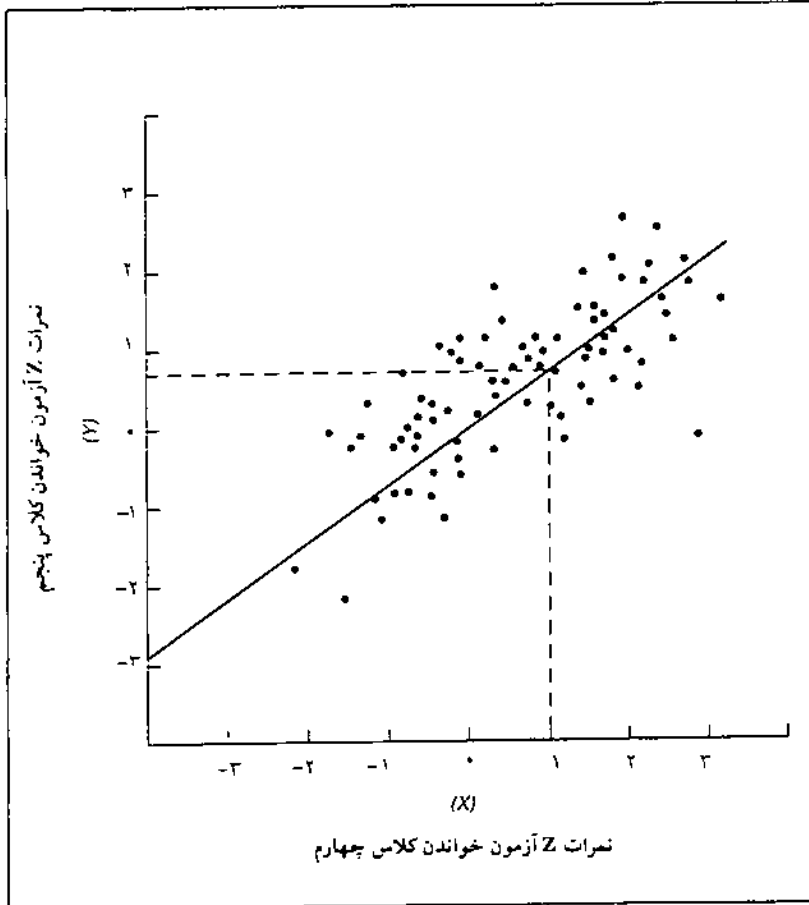
برای مثال، مشاور کلیسایی می‌خواست اثر دو روش مختلف مشاوره در امور ازدواج را مقایسه کند. او در یک پیش آزمون، میزان سازگاری در زندگی زناشویی زوجهایی را که در برنامه‌اش شرکت داشتند، اندازه‌گیری کرد و سپس آنها را به صورت تصادفی در معرض روش الف و روش ب قرار داد. در روش الف که میزان قابل توجهی برخوردارها و خودکامی را می‌طلبید، تنها ۱۱ تا ۲۰ زوج در برنامه باقی ماندند و برای اندازه‌گیری پس آزمون در دسترس بودند. در برنامه ب که از آزمون بیشتری برخوردار و آسانتر بود، ۱۸ تا ۲۰ زوج باقی ماندند و در مرحله پس آزمون شرکت داشتند. میانگین نمره پس آزمون روش الف بالاتر از میانگین گروه متعلق به روش ب بود. بالاتر بودن میانگین نمره روش الف نسبت به روش ب، می‌تواند به عنوان دلیلی برای مؤثرتر بودن روش الف تلقی شود، لکن دلیل آن می‌تواند این واقعیت باشد که تنها زوجهایی در این برنامه باقی ماندند که از انگیزه بیشتری برای بهبود وضع زندگی زناشوییشان برخوردار بودند، در حالی که نمرات پس آزمون افراد گروه ب متعلق به زوجهایی بود که انگیزه کمتری داشتند.

۸- تعامل بین بلوغ و انتخاب: این نوع تعامل در یک طرح شبه آزمایشی دیده می‌شود. در این نوع طرحها، گروههای آزمایشی و گواه به صورت تصادفی انتخاب نمی‌شوند، بلکه گروهها به صورت از قبل موجود، مانند کلاسهای درس، شکل گرفته‌اند. حتی چنانچه یک پیش آزمون نشان دهد که گروهها مشابه هستند، امکان دارد گروه آزمایشی از بلوغ بیشتری برخوردار باشد و همین بلوغ بیشتر عامل اثر مشاهده شده باشد. دانش آموزانی که بلوغ زودرس تری دارند، برای گروه آزمایشی انتخاب می‌شوند و این تعامل بین بلوغ و انتخاب است که اشتباهاً به جای اثر متغیر آزمایشی تلقی می‌شود.

برای مثال، گروهی که در سال اول دانشگاه برای گذراندن یک کلاس انگلیسی تخصصی فوق برنامه انتخاب شده‌اند، ممکن است در مقایسه با گروهی که یک کلاس انگلیسی عادی را می‌گذرانند، از نظر گنجینه لغات رشد بیشتری نشان دهند؛ زیرا گنجینه لغات آنان هم قبل از شرکت در کلاس و هم در طول سال تحصیلی، در مرتبه بالاتری قرار داشته است.

مقایسه تفاوت‌های پیش آزمون و پس آزمون کلاسهای تخصصی و عادی می‌تواند به این نتیجه‌گیری اشتباه منجر شود که متغیر مستقل مسبب تفاوت به دست آمده است، در حالی که در

شکل ۱-۹ پراکندگی برای نمرات خواندن کلاس چهارم و پنجم



واقع عامل این پیشرفت تعامل انتخاب و بلوغ است. این مشکل معمولاً زمانی پیش می‌آید که نمرات افراد داوطلب با افراد غیر داوطلب مقایسه می‌شود. داوطلبان ممکن است برای به دست آوردن نمره بالاتر در متغیر وابسته، نسبت به افراد غیر داوطلب، از انگیزه بیشتری برخوردار باشند و این تفاوت انگیزه پیشرفت به اشتباه به متغیر مستقل نسبت داده شود. احتمال دارد چنین امری حتی در زمانی که دو گروه در پیش آزمون یکسان بوده‌اند، باز هم اتفاق بیفتد. کلیه روشهای کنترل که قبلاً در همین فصل مورد بحث قرار گرفت، برای کنترل آن دسته از متغیرهای نامرتبط که اعتبار درونی یک طرح را خدشه دار می‌سازند، طراحی شده‌اند.

### اعتبار بیرونی

اعتبار بیرونی به قابلیت تعمیم و یا معرف بودن یافته ها بر می‌گردد. محقق این سؤال را مطرح می‌کند: این یافته‌ها را به چه جامعه آماری، موقعیتی، متغیرهای آزمایشی و متغیرهای اندازه‌گیری می‌توان تعمیم داد؟

هر بررسی با بهره‌گیری از گروه خاصی از آزمودنیها و وسایل اندازه‌گیری منتخب و تحت شرایطی یگانه، صورت می‌گیرد، ولی محقق می‌خواهد نتایج بررسی‌اش حوزه وسیعتری از آزمودنیها، موقعیتها و عملیاتی را که به طور واقعی مورد بررسی قرار گرفته است، در بر گیرد. برای آنکه بتوان مشاهده شده‌ای را به مشاهده نشده‌ای تعمیم داد، اطمینان از این که نمونه رویدادهایی که واقعاً مورد مطالعه قرار گرفته است، نماینده جامعه بزرگتری است که می‌خواهیم یافته‌ها را به آنها تعمیم بدهیم، ضروری است. میزان اعتبار بیرونی یک آزمایش، بستگی دارد به اینکه تا چه حد بتوانیم یافته‌های یک آزمایش را به آزمودنیها، موقعیتها و وسایل اندازه‌گیری دیگر تعمیم دهیم.

اسمیت و گلاس<sup>(۱)</sup> سه نوع اعتبار بیرونی را مشخص کرده‌اند: اعتبار بیرونی جامعه آماری، اعتبار بیرونی بوم شناختی و اعتبار بیرونی عملیات.

اعتبار بیرونی جامعه آماری به شناسایی جامعه‌ای که نتایج آزمایش به آنها تعمیم داده می‌شود، مربوط می‌شود. اعتبار بیرونی جامعه آماری این سؤال را مطرح می‌سازد که از کدام جامعه آماری آزمودنیها می‌توان انتظار همان رفتاری را داشت که از آزمودنیهای گروه نمونه سر زده است. اعتبار بیرونی بوم شناختی با تعمیم تأثیرات آزمایش به شرایط دیگر محیطی سر و کار دارد. در اینجا این سؤال مطرح می‌شود که تحت چه شرایطی می‌توان انتظار داشت که همین نتایج به دست آید. اعتبار بیرونی عملیات، با این امر سر و کار دارد که تعاریف کاربردی تا چه حد معرف سازه‌های مورد بررسی هستند. آیا چنانچه محقق دیگری همین مسئله را با عملیات متفاوتی مورد بررسی قرار می‌داد، همین روابط به دست می‌آمد؟

اعتبار بیرونی جامعه      آرزوی محقق آن است که بتواند یافته‌های خود را از گروه

1- Smith and G. V. Glass (1987) Research and Evaluation in Education PP.(143-153), and the Social Sciences Englewood Cliffs, N.J. Printice - Hall.

آزمودنیهایش به یک جامعه بزرگتر انسانی که هنوز مورد مطالعه قرار نگرفته‌اند، تعمیم دهد. برای مثال، فرض کنیم که محقق به یک روش جدید مؤثر در آموزش خواندن در مورد یک نمونه کلاس اولیها دست یافته باشد. مسلماً او مایل است نتیجه بگیرد که این روش برای دیگر گروههای کلاس اولی، حتی شاید تمام کلاس اولیهای ایالات متحده، نسبت به دیگر روشها ارجحیت دارد. برای آنکه بتواند از یافته‌های آزمایشی خود، نتیجه معتبرتری برای جامعه‌های بزرگتر بگیرد، باید به طور صحیح جامعه‌هایی را که این یافته‌ها به آنها تعمیم‌پذیر هستند، بشناسد. این بحث به تمایز کمپ‌نورن<sup>(۱)</sup> مابین جامعه قابل دستیابی آزمایشی<sup>(۲)</sup> و جامعه هدف<sup>(۳)</sup> مربوط است که در فصل ۶ به آن پرداختیم. جامعه اول به جامعه آزمودنیهای اطلاق می‌شود که جهت بررسی در دسترس و یا قابل دستیابی هستند. جامعه هدف، کل گروه آزمودنیها است که محقق می‌خواهد نتایج یافته‌هایش را در مورد آنها به کار برد. در مثال فوق، جامعه قابل دستیابی آزمایشی احتمالاً تمام کلاس اولیهای مدارس منطقه هستند. جامعه هدف، کلیه دانش آموزان کلاس اول در ایالات متحده خواهند بود. تعمیمهای محقق در دو مرحله اتفاق می‌افتد:

۱- از نمونه، به جامعه قابل دستیابی آزمایشی ۲- از جامعه قابل دستیابی، به جامعه هدف.

چنانچه محقق نمونه‌اش را به صورت تصادفی از بین جامعه قابل دستیابی آزمایشی (تمام کلاس اولیهای مدارس منطقه) انتخاب کرده باشد، می‌تواند یافته‌هایش را بدون هیچ اشکالی به گروه بزرگتر تعمیم دهد. آمار استنباطی احتمال این امر را نشان می‌دهد که آنچه در مورد نمونه صادق است، در مورد جامعه‌ای که نمونه از آن انتخاب شده است نیز می‌تواند صادق باشد.

در مرحله دوم، محقق می‌خواهد یافته‌هایش را در مورد جامعه قابل دستیابی به جامعه هدف (کلیه کلاس اولیها در ایالات متحده) تعمیم دهد. این تعمیم را نمی‌توان با همان درجه اطمینان تعمیم نوع اول انجام داد. به منظور رسیدن به این نتیجه، باید آگاهی کاملی از ویژگیهای هر دو جامعه داشته باشیم. هرچه جامعه قابل دستیابی و جامعه هدف از شباهت بیشتری برخوردار باشند، میزان اطمینان در امر تعمیم یکی به دیگری بالاتر است. تعمیم جامعه قابل دسترسی به جامعه هدف، با استفاده از روشهای آماری امکان‌پذیر نیست. این امر از طریق داوری

1- Kempthorne, O. (1961). The design and analysis of experiments with some reference to educational research. In R.D. Collier, Jr., and S.M. Elam (eds.), *Research Design and Analysis: Second Annual Phi Delta Kappa Symposium on Educational Research* (pp. 97 - 126). Bloomington, IN: Phi Delta Kappa.

2- Experimentally Accessible Population

3- Target Population



صورت می‌گیرد. محقق باید با دقت هرچه بیشتر جامعه قابل دستیابی را تعریف کند و نقاط تمایز این جامعه را با جامعه هدف، نشان دهد. از این طریق محقق و خواننده گزارش تحقیقاتی می‌توانند در مورد احتمال قابلیت تعمیم نتایج به جامعه هدف یا هر جامعه دیگری، داوری زیرکانه کنند.

جنبه دیگر اعتبار بیرونی جامعه آماری، احتمال وجود تعامل میان ویژگیهای آزمودنی و تیمار است. چنانچه دو جامعه قابل دستیابی آزمایشی، معرف جامعه هدف واحدی نباشند، بررسیهای به ظاهر مشابه می‌توانند به نتایج کاملاً متفاوت منجر شوند، یعنی رابطه میان تیمار و متغیر وابسته می‌تواند در مورد یک گروه وجود داشته باشد و در مورد گروه دیگری با ویژگیهای متفاوت وجود نداشته باشد، بنابراین تعمیم یافته‌های مربوط به یک گروه به گروه دیگر امکان پذیر نباشد. روش مشاوره الف می‌تواند در مدارس شهری نتایج بهتری از روش ب بدهد؛ در حالی که روش ب در بسیاری از مدارس حومه ثروتمندترین، ممکن است ارجحیت داشته باشد. بهترین روش تدریس برهان کمی برای کلاس دومها ممکن است بدترین روش برای کلاس هشتمی‌ها باشد، آنچنانکه مثلی قدیمی می‌گوید: «نان یک مرد می‌تواند زهر دیگری باشد.» بار دیگر خاطر نشان می‌سازیم که یک توصیف دقیق از جامعه قابل دستیابی به دست اندرکاران تعلیم و تربیت در امر داوری در این مورد که یک تیمار بخصوص برای جامعه مورد بررسی شان «نان است یا زهر»، کمک می‌کند.

معمولاً محقق می‌تواند اعتبار بیرونی یک بررسی را با تقسیم آزمودنیها بر حسب ویژگیهای مربوطشان و بکارگیری یک طرح عاملی (رجوع شود به فصل ۶)، قوت بخشد. برای مثال، در مورد یک جامعه آماری مشتمل بر مدرسه‌های راهنمایی قابل دستیابی، که به منظور بررسی مقایسه‌ای اثر دو روش تدریس انتخاب شده است، محقق می‌تواند یک طرح عاملی  $2 \times 3 \times 2$  (روشها در کلاس درجنس) انجام دهد. نتایج ارائه شده از این طریق نه تنها می‌تواند اثربخشی دو روش را مقایسه کند، بلکه میزان ثبات نتایج را در بین کلاس ششمی‌ها، هفتمی‌ها، هشتمی‌ها و نیز میان پسران و دختران نشان دهد.

اعتبار بیرونی بوم شناختی      آزمایشگران باید اعتبار بوم شناختی را نیز در نظر داشته باشند؛ بدین معنی که آنها می‌خواهند این قدرت را داشته باشند که بگویند نتایج در شرایط

محیطی دیگر نیز به دست خواهد آمد. برای دستیابی به اعتبار بوم شناختی، یک طرح تحقیق باید این اطمینان را که تأثیر آزمایشی، مستقل از محیط آزمایشی خاص است، فراهم آورد.

مسلماً اولین ضرورت دستیابی به اعتبار بوم شناختی این است که محقق توصیف کاملی را از شرایط آزمایشی به کار برده شده در بررسی ارائه دهد. تنها در چنین حالتی است که خواننده می‌تواند در مورد میزان تعمیم پذیری یافته‌ها به شرایط دیگر داوری کند. برای مثال، آیا می‌توان یافته‌های به دست آمده از یک کلاس وسیع و مجهز را به کلاس درسی شلوغ و فاقد تجهیزات، تعمیم داد؟ در اینجا نیز محقق، باید توصیفی کافی از شرایط را ارائه کند تا خواننده بتواند در مورد تعمیم یافته‌ها داوری منطقی کند. معمولاً آزمایشها در شرایطی صورت می‌گیرند که متغیرهایی نظیر نور، صدا و دیگر عوامل اخلاص کننده، کاملاً قابل کنترل هستند. اعمال چنین کنترلی از نظر غنی کردن اعتبار بیرونی مطلوب است، لکن اگر میزان این کنترل به حدی باشد که وضعیت آزمایش، دیگر معرف محیط طبیعی آزمودنیها نباشد، اعتبار بیرونی آزمایش کاهش خواهد یافت. شاید بهتر باشد ابتدا تیماری را در شرایطی با کنترل کافی امتحان کنیم تا به اعتبار درونی دست یابیم، سپس چنانچه تیمار کارآیی داشته باشد، دوباره آن را در شرایطی طبیعی تر امتحان کنیم.

احتمال دارد اثر واکنشی ناشی از ترتیبات آزمایشی وجود داشته باشد. آگاهی آزمودنیها از اینکه در یک بررسی شرکت می‌کنند، می‌تواند بر روی پاسخ آنها نسبت به تیمار اثر بگذارد. حضور مشاهده‌گران یا تجهیزات به هنگام انجام آزمایش، احتمال دارد پاسخهای طبیعی را به گونه‌ای تغییر دهد که محقق نتواند اثر متغیر آزمایشی را به افرادی که در شرایط غیر آزمایشی، در معرض آن متغیر قرار می‌گیرند، تعمیم دهد.

هنگامی که آزمودنیها از این که در یک آزمایش شرکت دارند، آگاه هستند، اغلب به جای آن که رفتار عادی خود را بروز دهند، رفتار خود را تغییر می‌دهند تا «آزمودنیهای خوبی» باشند.

یک مورد از اثر واکنشی به عنوان مسئله‌ای از اعتبار بیرونی اثر هاوورن<sup>(۱)</sup> است. این اثر بر مبنای بررسیهای میزان تولید در کارخانه وسترن الکتریک هاوورن در سیسرو<sup>(۲)</sup>، ایالت ایلینویز<sup>(۳)</sup>، نامگذاری شده است. ابتدا فرضیه پژوهشگران مبنی بر اینکه «با افزایش نور کارخانه

1- Hawthorne Effect

2- Cicero

3- Illinois

میزان تولید افزایش می‌یابد، ظاهراً تأیید شد، لکن بررسیهای بعدی نشان داد که هرگونه تغییری، حتی کم کردن نوره، تولید را افزایش می‌دهد. محققان چنین نتیجه‌گیری کردند که تولید کارگران کارخانه به این دلیل افزایش یافته است که آنها می‌دانستند که در یک بررسی شرکت دارند و به آنها توجه خاصی می‌شود.

احتمال دارد یک روش تدریس جدید به نظر موثرتر از روش قدیمی جلوه کند، در حالی که این دو روش از نظر اثر بخشی یکسان هستند و در واقع، چون معلمان و دانش‌آموزان برنامه جدید احساس می‌کنند که استثنایی هستند و به آنچه که انجام می‌دهند، اشتیاق پیدا کرده‌اند و از طرف دیگر آنهایی که در گروه گواه هستند، احساس طرد شدگی می‌کنند، بین این دو روش تدریس تفاوت ایجاد شده است.

از سوی دیگر احتمال دارد روش جدید غیر موثرتر از روش قدیمی به نظر رسد، بدین لحاظ که تا زمان اندازه‌گیری متغیر وابسته، معلمان و شاگردان، مطالب و روش کار با آن را فرا نگرفته باشند. برای مثال، در جمع بندی شون<sup>(۱)</sup> از تحقیقات قبلی در مورد مقایسه نتایج آموزش ریاضیات مدارس ابتدایی به صورت گروهی و فردی، این نتیجه به دست آمد که از ۱۷ بررسی تفاوت معنی دار بین دو روش و تأیید آموزش گروهی ملاحظه شد. نتایج ۱۱ بررسی، فرضیه صفر را رد نکرد و تنها نتایج ۸ بررسی تفاوت معنی دار بین دو شیوه و تأیید آموزش انفرادی را نشان داد، در حالی که اغلب محققان فرضیه‌شان را چنین تبیین کرده بودند که آموزش انفرادی نسبت به آموزش گروهی ارجحیت دارد. نتایج فوق می‌تواند دلیلی بر این واقعیت باشد که در مورد اکثر جوامع، آموزش گروهی موثرتر است، لکن همچنین احتمال دارد که علت عملکرد بهتر دانش‌آموزان در آموزش گروهی، شناخت قبلی آنان و معلمانشان از شیوه تدریس گروهی و چگونگی کنار آمدن با آن باشد؛ در حالی که در آموزش انفرادی، دانش‌آموزان علاوه بر یادگیری محتوای درس حساب، باید راه جدیدی نیز برای آموختن درس بیابند.

پاره‌ای از اثرات تعاملی می‌تواند عمل تعمیم‌پذیری یافته‌های بررسی را خدشه‌دار سازد. برای مثال، یک پیش‌آزمون ممکن است میزان حساسیت و پاسخگویی آزمودنیهای آزمایش را نسبت به متغیر مورد بررسی کاهش یا افزایش دهد و در نتیجه یافته‌های به دست آمده از این

1- Schoen, H.L. (1976). Self - Paced Mathematics Instruction. How Effective has it been? (The Arithmetic Teacher, 23, 90-96).

جامعه پیش آزمون شده، معرف اثرات متغیر آزمایشی برای جامعه پیش آزمون نشده، نباشد. در چنین موردی تنها می توان یافته ها را در مورد گروههای پیش آزمون شده، نه گروههای پیش آزمون نشده، تعمیم داد.

**اعتبار بیرونی عملیات** هر یک از محققان تحقیقاتشان را با تعاریف کاربردی خاصی از متغیرهای مستقل و وابسته و روشهای ویژه صورت می دهند. مسئله اعتبار بیرونی عملیات این سؤال را مطرح می سازد که آیا انتظار می رود از محققان مختلف با تعاریف کاربردی متفاوت و یا روشهای متفاوت، نتایج یکسان به دست آید؟

**اثر آزمایشگر** می تواند یک مسئله اعتبار بیرونی باشد. اجراکنندگان یک طرح آزمایشی نسبت به آنچه انجام می دهند، بسیار احساس تعهد می کنند و دچار اشتیاق می شوند. آزمودنیها می توانند درگیر این اشتیاق شوند و عملکردشان بهتر از آن باشد که در شرایط طبیعی می بود. در پاره‌ای موارد، آزمایشگران بدون آنکه آگاه باشند، کمک بیشتری از آنچه در تعریف کاربردی تیمارشان تعیین شده است به آزمودنیهایشان می کنند. همچنین انتظارات آزمایشگر از نتایج می تواند به صورت بسیار ظریف و ناخودآگاه، به آزمودنیها منتقل شود. ممکن است آزمودنیهای گروه آزمایش به علت انتظاراتی که از آنها می رود، بهتر و آزمودنیهای گروه گواه بدتر از شرایط طبیعی عمل کنند.

جنبه دیگر اعتبار بیرونی تعاریف کاربردی متغیرهای مستقل و وابسته است. در پاره‌ای از آزمایشها، اضطراب با شوک الکتریکی و در پاره‌ای دیگر به وسیله دستورالعمل های کلامی، ایجاد شده است. تا چه حد می توان نتایج به دست آمده در یک موقعیت را به موقعیت دیگر تعمیم داد؟ آیا ناکامی حاصل از محروم کردن کودکان از اسباب بازی مورد علاقه شان با ناکامی حاصل از ارائه مسائل غیر قابل حل به آنها، یکسان است؟

چنانچه برای اندازه گیری متغیر وابسته از یک آزمون عینی استفاده شود، آیا آزمایشگر می تواند ادعا کند که در صورت استفاده از یک آزمون انشایی به عنوان وسیله اندازه گیری، همان اثر مشاهده می شد؟ تا چه حد نمرات آزمون تفکر خلاق تورنس<sup>(۱)</sup> با سازه خلاقیت که توسط یک معلم انگلیسی یا هنر تعریف می شود، همخوانی دارد؟

خواننده گزارش تحقیقاتی این جنبه اعتبار را ابتدا با این سؤال ارزشیابی می‌کند که متغیرهای مستقل و وابسته تا چه حد روشن تعریف شده‌اند. گزارشهایی که تعاریف مبهم را ارائه می‌کنند، برای هر گونه کاربردی بی ارزش هستند. سپس خواننده باید اعتبار و پایایی این تعاریف را ارزشیابی و در مورد چگونگی تجانس تعاریف کاربردی با تعاریف سازه‌های معنی‌دار برای خود، داوری کند.

یک مشکل بالقوه دیگر اعتبار بیرونی عبارت است از حساس‌سازی پیش‌آزمون. اجرای یک پیش‌آزمون می‌تواند آزمودنیها را نسبت به تیمار به گونه‌ای متفاوت از شرایطی که پیش‌آزمون انجام نشده بود، حساس و هشیار کند. برای مثال، به گروهی از دانش‌آموزان کلاس هفتم پرسشنامه‌ای حاوی سؤالاتی در زمینه رژیم غذایی داده شد و سپس آنها به صورت تصادفی به دو گروه آزمایشی و گواه تقسیم شدند. به گروه آزمایشی چند فیلم در زمینه رژیمهای خوب غذایی نشان داده می‌شود، در حالی که گروه گواه فیلمهایی بهداشتی، که ارتباطی با رژیم غذایی ندارد، می‌بینند. متغیر وابسته از طریق مشاهده انتخاب غذا توسط کودکان، در موقعیتی که امکان انتخاب آزاد دارند، اندازه‌گیری می‌شود. چنانچه گروه آزمایشی به وضوح غذاهای سالم را ترجیح دهند، قاعدتاً محقق تمایل خواهد داشت که نتیجه‌گیری کند که فیلمها مؤثر بوده‌اند. محقق قبل از آنکه به نتیجه‌گیری بپردازد، باید این احتمال را در نظر گیرد که ممکن است پیش‌آزمون، موجب شده باشد دانش‌آموزان درباره عادات غذایی‌شان فکر کنند و آنها را در ارتباط با فیلمها تنظیم کنند. ممکن است این اثر در مورد یک گروه پیش‌آزمون نشده، مشاهده نشود. در بخشهای بعد در مورد راههای جداسازی و اجتناب از تعامل پیش‌آزمون تیمار بحث می‌کنیم (برای مثال انجام ندادن پیش‌آزمون و اعتماد به اینکه تقسیم تصادفی، گروههایی به دست می‌دهد که در یک پیش‌آزمون فرضی، تقریباً مساوی بوده‌اند). به خاطر داشته باشید که حساس شدن نسبت به پیش‌آزمون با انجام آزمون به عنوان تهدیدی برای اعتبار درونی متفاوت است. مورد قبل فی‌نفسه به خود آزمون بر می‌گردد که در متغیر وابسته تغییر ایجاد می‌کند، در صورتی که مورد اخیر بدین معناست که ترکیب پیش‌آزمونها و تیمارها اثری به وجود می‌آورند که تیمار به تنهایی آن اثر را ندارد.

اگرچه اعتبار درونی جزء لاینفک تحقیق است، لکن آزمایشگر می‌خواهد طرحی انتخاب کند که از نظر اعتبار درونی و بیرونی قوی باشد. به هر حال، در پاره‌ای موارد کسب یک نوع

اعتبار در نوع دیگر خدشه وارد می‌سازد. برای مثال، وقتی سعی می‌کنیم کنترل دقیق تری را در یک آزمایش در زمینه تعلیم و تربیت اعمال کنیم، وضعیت را بیشتر غیر طبیعی می‌کنیم و از میزان کاربرد نتایج در وضعیت عادی در کلاس درس می‌کاهیم. در عمل سعی می‌کنیم به سازشی بین اعتبار بیرونی و درونی برسیم. این بدین معنی است که باید طرحی انتخاب شود که میزان کنترل لازم را برای آنکه نتایج قابل تفسیر باشد، فراهم آورد و در عین حال از واقعیت دور نشود تا بتوان نتایج را به موقعیتهای مورد نظر تعمیم داد.

### /// طبقه بندی طرحهای آزمایشی

در بحث دربارهٔ طرحهای آزمایشی، که در پی می‌آید، این طرحها بر مبنای میزان کنترلی که اعمال می‌کنند، به سه گروه قبل آزمایشی، آزمایشی واقعی و شبه آزمایشی طبقه بندی شده‌اند. در مورد اعتبار بیرونی و درونی طرحها، به هنگام ارائهٔ طرحها بحث خواهد شد. قبل از شروع بحث طرحهای آزمایشی لازم است اصطلاحات و نمادهایی که مورد استفاده قرار خواهند گرفت، معرفی شوند.

۱- X نشان دهندهٔ متغیر مستقلی است که محقق می‌تواند آن را تغییر دهد. همچنین به آن متغیر آزمایشی یا تیمار نیز گفته می‌شود.

۲- Y به عنوان اندازهٔ متغیر وابسته به کار برده می‌شود. Y نشان دهندهٔ متغیر وابسته قبل از دستکاری متغیر مستقل X است که معمولاً نوعی پیش آزمون است که قبل از ارائهٔ تیمار آزمایشی صورت می‌گیرد. Y نمایانگر متغیر وابسته پس از دستکاری متغیر مستقل است که معمولاً پس آزمونی است که بعد از تیمار آزمایشی به آزمودنیها داده می‌شود.

۳- S معرف آزمودنی و یا پاسخگویی است که در آزمایش شرکت کرده است و جمع آن به صورت Ss می‌آید.

۴- گروه E به گروه آزمایشی و یا گروهی که در معرض تیمار متغیر مستقل قرار می‌گیرند، اطلاق می‌شود.

۵- گروه C به گروه کنترل و یا گروهی که تیمار آزمایشی دریافت نکرده‌اند، اطلاق می‌شود.

۶- R برای نشان دادن توزیع تصادفی آزمودنیها در گروههای آزمایشی و توزیع تصادفی تیمارها در گروهها آمده است.

M-۷ نشان می‌دهد که هم‌تاسازی برای یکسان ساختن گروه‌های مقایسه‌ای صورت گرفته است. در نمودار طرح‌های مختلف X ها و Y هادر یک ردیف واحد، مربوط به افراد یکسان است. جهت راست به چپ نشان دهنده ترتیب زمانی و Xها و Yهایی که عمود نسبت به یکدیگر هستند، به طور هم‌زمان ارائه می‌شوند. خط فاصله (-) برای گروه گواه نشان‌دهنده این امر است که آنها تیمار X را دریافت نکرده‌اند.

### // طرح‌های ماقبل آزمایشی

در این بحث، به دو طرح، که تحت عنوان «ماقبل آزمایشی» طبقه بندی شده‌اند، می‌پردازیم. دلیل این طبقه بندی، نبود کنترل و یا کنترل کم بر روی متغیرهای نامرتبط است. متأسفانه می‌بینیم که این گونه طرحها هنوز در تحقیقات آموزشی مورد استفاده قرار می‌گیرند. بحث در مورد این طرح‌های ضعیف از آن جهت مفید است که کاملاً چگونگی عملکرد متغیرهای نامرتبط را در به خطر انداختن اعتبار درونی یک طرح، نشان می‌دهد. چنانچه خوانندگان از نقاط ضعف این طرح‌ها آگاهی یابند، باید بتوانند از کاربرد آنها اجتناب کنند.

### / طرح ۱ - طرح یک گروهی پیش آزمون - پس آزمون

طرح یک گروهی معمولاً سه مرحله دارد: ۱- پیش آزمونی که متغیر وابسته را اندازه می‌گیرد، انجام می‌شود. ۲- تیمار آزمایشی X در مورد آزمودنیها به کار برده می‌شود. ۳- یک پس آزمون دوباره متغیر مستقل را در مورد آزمودنیها اندازه گیری می‌کند. تفاوت‌های منسوب به کاربرد تیمار آزمایشی از مقایسه نمرات پیش آزمون و پس آزمون، معین می‌شود.

### طرح ۱- طرح یک گروهی پیش آزمون - پس آزمون

پس آزمون	متغیر مستقل	پیش آزمون
$Y_2$	X	$Y_1$

برای نشان دادن چگونگی استفاده از این طرح، فرض کنیم که یک معلم ابتدایی می‌خواهد

میزان تأثیر روش جدیدی از تدریس علوم اجتماعی را در مورد دانش آموزان کلاس چهارم، ارزیابی کند. در آغاز سال تحصیلی، او یک آزمون استاندارد شده را که ظاهراً اندازه خوبی از دستیابی به اهداف درس علوم اجتماعی کلاس چهارم است، اجرا می‌کند. سپس روش جدید آموزش را اعمال می‌کند و در پایان سال، همان آزمون استاندارد شده را دوباره به کار می‌برد. نمرات آزمون اول و دوم را به منظور یافتن تفاوتی که در اثر قرار گرفتن در معرض روش جدید در یادگیری دانش آموزان ایجاد شده است، مقایسه می‌کند.

از آنجا که طرح ۱ فقط شامل یک گروه آزمودنی و یک معلم است، به نظر می‌رسد که می‌تواند تفاوت‌های فردی و متغیرهای موقعیتی را کنترل کند. به هر حال، کنترل فقط ظاهری است. مهمترین محدودیتی که در طرح یک گروهی وجود دارد، نبود گروه گواه است. آزمایشگر نمی‌تواند فرض کند که تفاوت مشاهده شده بین نمرات پیش آزمون و پس آزمون، معلول تیمار آزمایشی است. همیشه این احتمال وجود دارد که متغیرهای نامرتبب تمام و یا قسمتی از تغییر را به وجود آورده باشند. در نتیجه، این طرح فاقد اعتبار درونی است.

برخی از متغیرهای نامرتبب که می‌توانند باعث به وجود آمدن تفاوت بین نمره های پیش آزمون و پس آزمون شوند، کدامها هستند؟ دو متغیر نامرتبب که در این طرح کنترل نشده‌اند، سابقه و بلوغ هستند. سابقه به عنوان منبعی از متغیر نامرتبب، به رویدادهای ویژه‌ای که در فاصله زمانی بین پیش آزمون و پس آزمون اتفاق افتاده است، بجز تیمار آزمایشی، اطلاق می‌شود. در مثال درس علوم اجتماعی، علاقه گسترده جامعه به امر انتخابات، تأکید روزافزون بر روی آموزش علوم اجتماعی در مدارس و یا ورود یک معلم مؤثر می‌تواند در میزان موفقیت تحصیلی دانش آموزان در این رشته مؤثر باشد. شیوع یک بیماری مسری، که باعث غیبت دانش آموزان شده باشد، می‌تواند از میزان پیشرفت تحصیلی آنان بکاهد. بلوغ به تغییراتی که در اثر گذشت زمان در خود آزمودنی به وجود می‌آید، اطلاق می‌شود. در فاصله زمانی بین پیش آزمون و پس آزمون کودکان از نظر بدنی و عقلانی رشد می‌کنند و تجربیاتی در امر یادگیری به دست می‌آورند که می‌تواند بر روی متغیر وابسته اثر بگذارد. سابقه و بلوغ بخصوص چنانچه فاصله زمانی بین  $Y_1$  و  $Y_2$  طولانی باشد، عوامل نامرتبب مؤثری می‌شوند.

نقیصه دیگر طرح ۱، نبودن راهی برای ارزشیابی اثر پیش آزمون  $Y_1$  به خودی خود است. زمانی که آزمودنیها در یک آزمون و یا حتی آزمونهای شبیه به آن شرکت می‌کنند، عامل تمرین



اثر دارد؛ بدین معنا که آزمودنیها در بار دوم، حتی بدون اینکه در این فاصله آموزشی دیده باشند و یا در مورد موضوع بحثی شده باشد، نمرهٔ بهتری می‌گیرند. این امر نه تنها در مورد آزمونهای هوش و پیشرفت تحصیلی، بلکه در مورد آزمونهای شخصیت هم صادق است. در مورد آزمونهای شخصیت، معمولاً روندی به سوی سازگاری بهتر مشاهده می‌شود.

اکتساب بازآزمایی جنبهٔ دیگری از مشکل بزرگتر واکنش پذیری وسایل اندازه‌گیری است. این امر به این واقعیت اطلاق می‌شود که اغلب واکنشی بین آزمودنی و پیش‌آزمون وجود دارد و این واکنش است، نه دستکاری  $X$ ، که تغییراتی در  $Y$  ایجاد می‌کند. اندازه‌هایی که باعث می‌شوند آزمودنی واکنش نشان دهد، اندازه‌های واکنشی خوانده می‌شوند. برای مثال، در یک بررسی تغییرات نگرش، مقیاسها خود می‌توانند به عنوان محرک عمل کنند؛ بدین معنا که آزمودنی می‌تواند در مقابل محتوای مقیاس، واکنش نشان دهد و همین واکنش است که تغییر مشاهده شده را در نگرش، حتی بدون هیچگونه تیمار آزمایشی ایجاد می‌کند. این تأثیر زمانی به بهترین وجه مشهود می‌شود که پیش‌آزمون دارای محتوای جدید و بحث‌انگیز است و یا تأثیرات انگیزشی خاصی بر روی آزمودنیها دارد.

طرح ۱، نکات مثبت زیادی برای ترغیب محقق، در جهت به کار بردنش ندارد. بدون وجود یک گروه کنترل که امر مقایسه را امکان‌پذیر سازد، نتایج به دست آمده از طرح یک گروهی، اصولاً غیر قابل تفسیر است.

### طرح ۲ - طرح مقایسهٔ گروهی ایستا

در طرح ۲ از دو گروه یا بیشتر استفاده می‌شود که تنها یکی از آنها در معرض تیمار آزمایشی قرار می‌گیرند. فرض می‌شود که این دو گروه از نظر کلیهٔ جوانب مربوط به بررسی، یکسان و تنها از نظر قرار گرفتن در معرض  $X$ ، متفاوت باشند. سپس اندازه‌های متغیر وابسته برای دو گروه مقایسه می‌شود تا بتوان در مورد اثر تیمار  $X$  نتیجه‌گیری کرد.

این طرح در غالب روشهای تحقیق در زمینهٔ تعلیم و تربیت به کار برده می‌شود. میزان پیشرفت تحصیلی دانش‌آموزانی که با یک روش جدید آموزش دیده‌اند و کلاس مشابهی که با روش سنتی آموزش دیده‌اند، مقایسه می‌شوند.

طرح ۲ دارای گروه یا گروههای گواه است که انجام مقایسه را که برای رعایت اصول علمی

ضروری است، امکان پذیر می‌سازد، اما به هر حال تقیصه‌ای اساسی در این طرح وجود دارد. از آنجا که نه توزیع تصادفی و نه هم‌تاسازی، در توزیع آزمودنیها در گروههای آزمایشی و گواه به کار برده نمی‌شوند، نمی‌توانیم فرض کنیم که دو گروه قبل از تیمار آزمایشی، هم ارز بوده‌اند. احتمال دارد دو گروه از نظر برخی متغیرهای مربوط متفاوت باشند و این تفاوتها، نه  $X$  عامل به وجود آمدن تغییر مشاهده شده باشد. از آنجا که نمی‌توانیم مطمئن باشیم که گروهها از نظر کلیه عوامل اثرگذار بر متغیر وابسته یکسان هستند، این طرح فاقد کنترل لازم است و باید در زمره طرحهای ماقبل آزمایشی طبقه بندی شود.

### طرح ۲ - طرح مقایسه گروهی ایستا

گروه	متغیر مستقل	پس آزمون
E	X	$Y_1$
C	—	$Y_2$

### // طرحهای آزمایشی واقعی

طرحهایی از این دست، به لحاظ کنترلی که فراهم می‌آورند، بیشتر از دیگر انواع، جهت تحقیقات در زمینه تعلیم و تربیت توصیه می‌شوند.

### / طرح ۳ - طرح گروه گواه، آزمودنیهای تصادفی، تنها پس آزمون

طرح ۳ یکی از ساده ترین و در عین حال محکمترین طرحهای آزمایشی است. در این طرح به دو گروه آزمودنی، که هر یک از آنها در معرض یک نوع موقعیت قرار می‌گیرند، نیاز داریم. پیش آزمون صورت نمی‌گیرد. توزیع تصادفی، کلیه متغیرهای نامربوط احتمالی را کنترل می‌کند و اطمینان می‌دهد که هرگونه تفاوت اولیه‌ای که بین گروهها مشاهده می‌شود، فقط تصادفی است و در نتیجه از قانون احتمالات تابعیت می‌کند.

پس از آنکه آزمودنیها به گروهها تخصیص یافتند، فقط گروه آزمایشی در معرض تیمار آزمایشی قرار می‌گیرد. در کلیه موارد دیگر با دو گروه به یک نحو رفتار می‌شود. سپس اعضای دو گروه از نظر متغیر وابسته  $Y_1$  مورد اندازه گیری قرار می‌گیرند. نمرات به منظور رسیدن به

چگونگی اثر X مقایسه می‌شوند. چنانچه تفاوت بین میانگینها معنی‌دار باشد (بدین معنا که این تفاوت بیشتر از آن چیزی باشد که تنها بر مبنای اتفاق انتظار می‌رود)، آزمایشگر می‌تواند تا حد معقولی مطمئن باشد که وضعیتهای آزمایشی عامل به وجود آورنده نتایج مشاهده شده بوده‌اند. مهمترین مزیت طرح ۳، توزیع تصادفی آن است که هم ارزی آماری گروهها را قبل از قرار گرفتن در معرض متغیر مستقل تضمین می‌کند. به خاطر داشته باشید که هرچه بر حجم نمونه افزوده شود، احتمال اینکه توزیع تصادفی، گروههای هم ارزی به وجود آورد، افزایش می‌یابد. طرح ۳ اثرات اصلی عوامل سابقه، بلوغ و پیش‌آزمون را کنترل می‌کند. از آنجا که هیچ پیش‌آزمونی مورد استفاده قرار نگرفته است، تعاملی از تأثیر پیش‌آزمون و X نمی‌تواند وجود داشته باشد. این طرح، بویژه در مورد موقعیتهایی که احتمال پیش‌آمد واکنش‌پذیری پیش‌آزمون وجود دارد، توصیه می‌شود. همچنین این طرح در بررسیهایی که در آنها اجرای پیش‌آزمون امکان‌پذیر نیست و یا اجرای آن اصولاً مناسب نیست، مفید است. برای مثال، در بررسیهای مربوط به کودکانها و کلاسهای ابتدایی، که چون یادگیری قبلی وجود ندارد، اجرای پیش‌آزمون امکان‌پذیر نیست، مفید است. مزیت دیگر آن، این است که در صورت لزوم می‌توان بیش از دو گروه را انتخاب کرد.

### طرح ۳ - طرح گروه گواه، آزمودنیهای تصادفی، تنها پس‌آزمون

پس‌آزمون	متغیر مستقل	گروه	
$Y_1$	X	E	(R)
$Y_2$	—	C	(R)

طرح ۳ به محقق امکان سنجش تغییر را نمی‌دهد. چنانچه چنین امری مورد نظر محقق باشد، می‌تواند از طرح دیگری مانند طرح ۵ که در آن از هر دو پیش‌آزمون و پس‌آزمون استفاده می‌شود، سود جوید.

/ طرح ۴ - طرح آزمودنیهای هم‌تا شده تصادفی، تنها پس‌آزمون با گروه گواه  
این طرح شبیه طرح ۳ است، با این تفاوت که به جای انتخاب تصادفی برای یافتن گروههای

هم ارز، در اینجا از روش هم‌تاسازی استفاده می‌شود. آزمودنی‌هایی که از نظر یک یا چند متغیر که می‌توان آنها را به راحتی اندازه‌گیری کرد، هم‌تا می‌شوند، مانند هوش‌بهر و نمره خواندن. زوجها با هم‌تا ساختن آزمودنی‌هایی که نمراتشان تا حد امکان نزدیک است، شکل می‌گیرند. البته، متغیرهایی که بر مبنای آنها عمل هم‌تاسازی انجام می‌شود، آنهایی هستند که احتمالاً دارای همبستگی معنی‌داری با متغیر وابسته هستند. اگرچه در طرح ۴ پیش‌آزمون انجام نمی‌شود، اما چنانچه نمرات پیش‌آزمون متغیر وابسته در دسترس می‌بود، امکان داشت از آنها به صورت مؤثری در فرآیند هم‌تاکردن، استفاده کرد. هر یک از اعضای یک زوج به صورت تصادفی برای یک نوع تیمار انتخاب می‌شوند. شیوه شیر یا خط می‌تواند برای تخصیص تصادفی به کار برده شود.

هم‌تاسازی به بهترین وجه می‌تواند در بررسی‌هایی که در آنها از گروه‌های نمونه کوچک استفاده می‌شود و طرح ۳ برای آنها مناسب نیست، به کار رود. طرح ۳ کاملاً متکی به انتخاب تصادفی برای دستیابی به گروه‌های هم‌ارز است. در مورد نمونه‌های کوچک، تأثیر اتفاق به تنهایی می‌تواند موقعیتی را به وجود آورد که گروه‌های تصادفی اصولاً بسیار متفاوت با یکدیگر باشند. طرح ۳ به هیچ وجه قابل مقایسه بودن واقعی گروه‌های کوچک را قبل از شروع آزمایش تضمین نمی‌کند. طرح هم‌تاکردن آزمودنی‌ها، به هر حال از میزان تفاوت آزمایشی که ناشی از تفاوت‌های اصلی بین گروهی است، می‌کاهد؛ بدین معنا که تفاوت‌های بین آزمودنی‌ها را از نظر متغیرهای مربوط به متغیر وابسته، که قبل از آزمایش وجود داشته‌اند، کنترل می‌کند. تخصیص تصادفی زوجها به گروه‌ها، به قدرت این طرح می‌افزاید.

#### ۴ - طرح آزمودنی‌های هم‌تا شده تصادفی، تنها پس‌آزمون با گروه گواه

پس‌آزمون	متغیر مستقل	گروه
$Y_1$	X	E
$Y_2$	—	C
		( $M_1$ )

طرح ۴ با همان مشکلاتی که قبلاً در مورد هم‌تاسازی به عنوان وسیله‌ای برای کنترل ذکر

کردیم، مواجه است. همتاسازی کلیه آزمودنیهای بالقوه، باید کامل باشد و تخصیص اعضای هر زوج به گروهها باید به صورت تصادفی صورت گیرد. چنانچه به دلیل مناسب انجام نشدن عمل همتاسازی نیاز به حذف یک یا چند آزمودنی باشد، این نمونه اریب می شود. وقتی از طرح ۴ استفاده می کنیم، باید هر آزمودنی را، حتی به صورت تقریبی، قبل از تخصیص تصادفی همتا کنیم. طرح ۴ در مورد بیش از دو گروه نیز کاربرد دارد و این امر با به وجود آوردن مجموعه های همتا و تخصیص تصادفی هر عضو یک مجموعه به یک گروه انجام می شود.

**طرح ۵ - طرح آزمودنیهای تصادفی، پیش آزمون - پس آزمون با گروه گواه**  
در طرح ۵، آزمودنیها با بهره گیری از روشهای تصادفی در گروههای آزمایشی و گواه توزیع می شوند و به آنها در مورد متغیر وابسته  $Y$  پیش آزمون و پس آزمون داده می شود. تیمار تنها در مورد آزمودنیهای آزمایشی برای زمان مشخص اعمال می شود و پس از آن، دو گروه از نظر متغیر وابسته مورد سنجش قرار می گیرند. تفاوت متوسط پیش آزمون و پس آزمون ( $Y_2 - Y_1$ ) برای هریک از گروهها محاسبه می شود. سپس به منظور اطمینان یافتن از اینکه تیمار آزمایشی موجب تغییر بیشتری از موقعیت کنترل شده است، نمره های تفاوت متوسط مقایسه می شوند. معنی داری تفاوت در تغییرات متوسط (که با تفریق تغییر متوسط گروه گواه از تغییر متوسط گروه آزمایش به دست می آید)، باید از طریق یک آزمون آماری مناسب، نظیر آزمون  $t$  یا آزمون  $F$  تعیین شود. روش دیگر آماری که از دقت بیشتری برخوردار است، عبارت است از انجام یک تحلیل کوواریانس که در آن نمرات پس آزمون به عنوان متغیر وابسته و نمرات پیش آزمون به عنوان متغیر کمکی تلقی می شود.

اگرچه گروه گواه، تیماری دریافت نمی کند، اما این امر بدین معنی نیست که گروه گواه هیچ گونه تجربه ای کسب نمی کند. در تحقیق در زمینه روشهای تدریس، گروه گواه با روشهای سنتی و معمولی آموزش می بینند. در پاره ای آزمایشهای یادگیری، رسم است که همزمان با آموزش ویدئو گروه آزمایشی، به گروه گواه فعالیتی غیر مربوط داده شود. در آزمایش بر روی اثرات دارویی خاص، ممکن است به گروه گواه یک دارونما (مانند قرص شکر) داده شود و این امر باید به نحوی صورت گیرد که آزمودنیها ندانند که با آنها به گونه ای متفاوت با افراد گروه آزمایشی رفتار شده است.

## طرح ۵- طرح آزمودنیهای تصادفی، پیش آزمون - پس آزمون با گروه گواه

پس آزمون	متغیر مستقل	پیش آزمون	گروه	
$Y_2$	X	$Y_1$	E	(R)
$Y_2$	—	$Y_1$	C	(R)

اندازه های قبل و بعد در طرح ۵، به محقق این امکان را می دهد که تغییرات را بررسی کند و این طرح معمولاً به عنوان طرح کلاسیک برای آزمایشهای تغییرات خوانده می شود. نقطه قوت اصلی این طرح، تصادفی بودن اولیه آن است که هم ارزی گروهها را قبل از آزمایش تضمین می کند. همچنین این واقعیت که آزمایشگر بر پیش آزمون کنترل دارد، می تواند یک واریسی اضافی بر یکسانی دو گروه را از نظر متغیر وابسته  $Y$  فراهم آورد. طرح ۵، با ماهیت تصادفی بودنش، اکثر متغیرهای نامربوط را که می توانند بر اعتبار درونی خدشه وارد کنند، کنترل می کند. برای مثال تأثیرات سابقه، بلوغ و انجام پیش آزمون برای هر دو گروه وجود دارد؛ در نتیجه، هرگونه تفاوتی را که در بین گروهها در  $Y$  وجود داشته باشد، نمی توان به این عوامل نسبت داد. انتخاب افتراقی آزمودنیها و رگرسیون آماری نیز از طریق شیوه انتخاب تصادفی کنترل می شود. طرح ۵ را می توان با بیش از دو گروه نیز اجرا کرد.

مسئله اصلی مورد نظر در کاربرد طرح ۵، اعتبار بیرونی است. مشکل، از استفاده از پیش آزمون ریشه می گیرد که خود از اصول این طرح است. چنانچه قبلاً ذکر کردیم، ممکن است تعاملی بین پیش آزمون و آزمودنیها وجود داشته باشد که این تعامل خود می تواند آزمودنیها را به طریقی تغییر دهد یا به آنها حساسیت بخشد. اگرچه هر دو گروه آزمایشی و گروه گواه پیش آزمون را دریافت می کنند و اثر حساسیت بخشیدن را تجربه می کنند، اما ممکن است که آزمودنیهای گروه آزمایشی، صرفاً به دلیل حساسیت زیادشان، به تیمار X به نوع خاصی پاسخ گویند. سؤال مهم مطرح این است که آیا اثر X بر روی آزمودنیهای آزمایشی، بدون قرار گرفتن در معرض پیش آزمون همین می بود؟ این مشکل بویژه در بررسیهای تغییر نگرش مشخص است. به محض آنکه در یک بررسی اولین مقیاس نگرش به عنوان پیش آزمون به کار برده شود، می تواند نوعی علاقه یا حساسیت نسبت به مطالبی که در مقیاس وجود دارد، در آزمودنی ایجاد

کند. در نتیجه زمانی که آزمودنی در موقعیت آزمایش (سخنرانی، فیلم و غیره) قرار می‌گیرد، بیشتر به ترکیب حساسیت برانگیخته شده‌اش نسبت به مطالب و تیمار آزمایشی پاسخ می‌دهد تا صرفاً تیمار X.

به مثال دیگری می‌پردازیم. فرض کنیم تعداد دانش آموزانی که ادعا می‌کنند منبع خبریشان روزنامه‌وال استریت<sup>(۱)</sup> است، معیاری برای موفقیت در روش جدید تدریس علوم اجتماعی در دبیرستان باشد. در طول دوره، تأکید بخصوصی بر روی این منبع ویژه گذارده نشده است، ولی این روزنامه همراه چند روزنامه، که از معروفیت کمتری برخوردارند، در دسترس دانش آموزان قرار گرفته است. چنانچه در طرح بررسی از طرح پیش‌آزمون و پس‌آزمون استفاده شده باشد، در پرسشنامه پیش‌آزمون ممکن است مواردی نظیر «آیا شما برای گرفتن خبر، روزنامه‌وال استریت را می‌خوانید؟»، آمده باشد. تنها همین سؤال می‌تواند حساسیت دانش‌آموزان گروه آزمایش را نسبت به این روزنامه برانگیزد؛ به نحوی که چنانچه در طول دوره این روزنامه در دسترسشان قرار بگیرد، از بین روزنامه‌های دیگر آن را انتخاب کنند. گروه گواه در معرض منابع خبری مختلف قرار نمی‌گیرند و در نتیجه سؤال پیش‌آزمون امکان حساسیت برانگیزی را بر روی آنها ندارد. چیزی که ممکن است در این بررسی اتفاق بیفتد، این است که گروه آزمایشی نسبت به گروه گواه استفاده بیشتری را از روزنامه‌وال استریت، در پس‌آزمون نشان دهند و علت این امر، به تنهایی، محتوای درس نباشد، بلکه ناشی از ترکیبات اثرات محتوای درس و پیش‌آزمون باشد. اما کلاس دیگری که با همان روش آموزش دیده باشند، ولی پیش‌آزمونی در موردشان اجرا نشده باشد، حساس نشده باشند و نسبت به گروه گواه توجه بیشتری را به این روزنامه نشان ندهند. چنین امری نشانه تعاملی بین پیش‌آزمون و تیمار آزمایشی است. از آنجاکه پیش‌آزمون ممکن است حساسیت و پاسخگویی آزمودنی را نسبت به تغییرات X کاهش یا افزایش دهد، در نتیجه یافته‌های به دست آمده از یک نمونه پیش‌آزمون شده ممکن است معرف اثرات متغیر آزمایشی در مورد جامعه پیش‌آزمون نشده‌ای که آزمودنیهای آزمایشی از آن انتخاب شده‌اند، نباشند. در نتیجه ما در امر تعمیم یافته‌ها با مشکل روبرو می‌شویم. احتمال دارد تنها بتوانیم یافته‌های آزمایشی را در مورد گروههای پیش‌آزمون شده و نه پیش‌آزمون نشده، تعمیم دهیم. این خود خدشه دار ساختن اعتبار بیرونی است.

به رغم این نقیصه، طرح ۵ به میزان زیاد مورد استفاده قرار می‌گیرد. دلیل کاربرد زیاد این طرح، جدی نبودن مشکل اثر تعامل در غالب تحقیقات در زمینه تعلیم و تربیت است. پیش آزمونها معمولاً نوعی آزمونهاى پیشرفت تحصیلی هستند و در نتیجه اثر حساسیت بخشی مهمی بر روی آزمودنیها، که با چنین آزمونهاى آشنا هستند، ندارند. به هر حال، چنانچه فرآیند آزمایش به گونه‌ای تازگی داشته باشد و یا اثرات انگیزشی داشته باشد، توصیه می‌شود که محقق طرحی را انتخاب کند که در آن پیش آزمون به کار نرود. از طرفی، هر زمان محقق در مورد اثرات تعاملی پیش آزمون شک داشته باشد، می‌توان گروه یا گروههای تازه‌ای به بررسی اضافه کرد. سولومون<sup>(۱)</sup> برای رفع نقطه ضعف طرح ۵، دو طرح با استفاده از گروه یا گروههایی را پیشنهاد می‌کند که در مورد آنها پیش آزمون انجام نمی‌شود.

#### طرح ۶ - طرح سه گروهی سولومون

در طرحهای اولیه سولومون، از سه گروه با استفاده از روش تخصیص تصادفی آزمودنیها به گروهها، استفاده می‌شود.

چنانچه ملاحظه می‌شود، دو سطر اول این طرح با طرح ۵ یکسان است. ولی، طرح سولومون از این مزیت برخوردار است که از گروه گواه ثانوی استفاده می‌کند و بدین وسیله بر مشکل ذاتی طرح ۵ فائق می‌آید. این مشکل، اثر تعاملی پیش آزمون و دستکاری آزمایشی است. در مورد گروه گواه دوم که با حرف C<sub>۲</sub> نشان داده شده است، پیش آزمون انجام نمی‌شود، ولی این گروه در معرض تیمار X قرار می‌گیرد. نمرات Y<sub>۲</sub> آنها برای ارزشیابی اثر تعامل به کار برده می‌شود.

#### طرح ۶ - طرح سه گروهی سولومون

پس آزمون	متغیر مستقل	پیش آزمون	گروه	
Y <sub>۲</sub>	X	Y <sub>۱</sub>	E	(R)
Y <sub>۲</sub>	—	Y <sub>۱</sub>	C <sub>۱</sub>	(R)
Y <sub>۲</sub>	X	—	C <sub>۲</sub>	(R)



ارزشیابی اثر تعامل از طریق مقایسه نمرات  $Y_7$  برای سه گروه میسر می شود. تنها نمرات پس آزمون در تحلیل وارد می شود. هرچند میانگین گروه آزمایشی به صورت معنی داری بالاتر از میانگین گروه گواه اول است، اما نمی توان مطمئن بود که این تفاوت ناشی از  $X$  است. این تفاوت می تواند به دلیل بالارفتن میزان حساسیت آزمودنیها پس از پیش آزمون و ناشی از تعامل حساسیت با  $X$  باشد. به هر حال، چنانچه میانگین  $Y_7$  برای گروه گواه دوم هم به صورت معنی داری بالاتر از میانگین گروه کنترل اول باشد، می توانیم فرض کنیم که تیمار آزمایشی عامل این تفاوت بوده است و نه اثر متقابل پیش آزمون و  $X$ ؛ زیرا در مورد گروه کنترل ۲ پیش آزمون انجام نشده بود. این گروه، هرچند در معرض تیمار  $X$  قرار گرفته است، ولی به عنوان گروه گواه عمل می کند و در نتیجه با علامت اختصاری  $C_7$  مشخص می شود.

#### ا / طرح ۷ - طرح چهار گروهی سولومون

طرح ۷، کنترل شدیدتری را با توسعه دادن طرح ۶ و اضافه کردن یک یا چند گروه گواه، اعمال می کند. در مورد گروه چهارم نه پیش آزمون انجام می شود و نه تیمار  $X$  را دریافت می کنند. در اینجا هم گروه سوم، هرچند تیمار  $X$  را دریافت می کند، به عنوان گروه گواه، عمل می کند.

طرح ۷، طرحی قوی است، زیرا از مزیت‌های چند طرح دیگر برخوردار است و نیز ویژگیهای خاص خود را دارد. دو سطر اول (طرح ۵)، عوامل خارجی، نظیر سابقه و بلوغ را کنترل می کند و سطر سوم (طرح ۶) کنترل اثر تعامل پیش آزمون  $X$  را امکان پذیر می سازد. چنانچه سطر چهارمی هم اضافه شود و طرح ۷ را بسازد، ما کلیه اثرات همزمان احتمالی را که می تواند در فاصله زمانی بین  $Y_1$  و  $Y_7$  اتفاق بیفتد، کنترل می کنیم. دو سطر آخر نشانه طرح ۳ است. در نتیجه عملاً ما ترکیبی از طرح پیش آزمون و پس آزمون، با کنترل آزمایشی داریم که در آن طرح توزیع تصادفی آزمودنیها به کار گرفته شده است.

علاوه بر آنکه نقاط قوت هریک از طرحها به تنهایی منظور شده است، از ویژگی تکرار نیز که از طریق آن آزمایش فراهم می آید، برخورداریم. این ترکیب از مزیت اطلاعات به دست آمده از فرایند پیش آزمون و پس آزمون بهره می گیرد و همزمان نشان می دهد که چگونه موقعیت آزمایشی گروه آزمودنیها را که پیش آزمون نشده اند، تحت تأثیر قرار می دهد.

## طرح ۷ - طرح چهارگروهی سولومون

پس آزمون	متغیر مستقل	پیش آزمون	گروه	
$Y_2$	X	$Y_1$	E	(R)
$Y_2$	—	$Y_1$	$C_1$	(R)
$Y_2$	X	—	$C_2$	(R)
$Y_2$	—	—	$C_3$	(R)

در طرح ۷ آزمایشگر می‌تواند برای تعیین اثر تیمار آزمایشی X، چندین مقایسه انجام دهد. چنانچه میانگین پس آزمون گروه آزمایشی به صورت معنی داری بزرگتر از میانگین اولین گروه گواه  $C_1$  باشد و اگر میانگین پس آزمون  $C_2$  به صورت معنی دار بزرگتر از  $C_1$  باشد، شواهدی در مورد مؤثر بودن تیمار آزمایشی در دست خواهیم داشت. اثرات شرایط آزمایشی را بر روی گروهی که پیش آزمون شده‌اند، می‌توان از طریق مقایسه پس‌آزمونهای گروه آزمایشی و گروه گواه  $C_1$  و یا تغییرات قبل و بعد E و  $C_1$  تعیین کرد. اثر آزمایش را بر روی آزمودنیهای یک گروه، که قبلاً پیش آزمون نشده‌اند، می‌توان با مقایسه  $C_2$  و  $C_3$  تعیین کرد. چنانچه تفاوت‌های متوسط بین نمره‌های پس آزمون  $E - C_1$  و  $C_2 - C_3$  تقریباً یکسان باشند، آزمایش باید اثر قابل مقایسه‌ای بر روی گروههای پیش‌آزمون شده و پیش‌آزمون نشده، داشته باشد.

طرح ۷ عملاً آزمایش را دوبار انجام می‌دهد، یک بار با و یک بار بدون پیش‌آزمونها. چنانچه نتایج این دو آزمایش آن طور که در بالا آمد، همخوانی داشته باشند، محقق می‌تواند نسبت به نتایج به دست آمده اعتماد بیشتری داشته باشد.

نقیصه اصلی این طرح، مشکل بودن اجرای آن در عمل است. در اجرای دو آزمایش به صورت همزمان، وقت و نیروی بیشتری صرف می‌شود؛ در حالی که مشکل یافتن آزمودنیهای یکسان که مورد نیازند نیز به جای خود باقی می‌ماند.

مشکل دیگر در رابطه با تحلیل آماری است. چهار مجموعه کامل از اندازه‌ها برای چهار گروه وجود ندارد. چنانچه در بالا آمد، می‌توانیم گروه E را با گروه گواه ۱ ( $C_1$ ) و گروه گواه ۲ ( $C_2$ ) را با گروه گواه ۳ ( $C_3$ ) مقایسه کنیم، ولی هیچ روش آماری وجود ندارد که ۶ اندازه موجود را همزمان آزمون کند. کمپ بل و استنلی طرح تحلیل واریانس دو عاملی را با نمرات پس آزمون به تنهایی توصیه می‌کنند. پیش‌آزمون همراه X، به عنوان یک متغیر مستقل دوم مورد نظر قرار می‌گیرد. طرح به صورت زیر است:

$X$	بدون $X$	
$Y_1$ ، آزمایشی	$Y_1$ ، گواه ۱	پیش آزمون شده
$Y_2$ ، گواه ۲	$Y_1$ ، گواه ۳	پیش آزمون نشده

از میانگینهای ستونی، محقق می‌تواند اثر اصلی  $X$ ، از میانگینهای ردیفی، اثر اصلی پیش‌آزمون و از میانگین خانه‌ها، تعامل آزمون و  $X$  را تعیین کند.

### // طرحهای عاملی

طرحهایی که تا اینجا ارائه شدند، طرحهای یک متغیری کلاسیک بودند که در آنها تغییرات در یک متغیر مستقل موجب تأثیر بر روی متغیر وابسته می‌شود. با این همه در مورد پدیده‌های پیچیده اجتماعی معمولاً چندین متغیر به صورت همزمان بر روی یکدیگر اثر تعاملی دارند و سعی در محدود ساختن یک بررسی به یک متغیر، تنها ساده‌انگاری تصنعی یک موقعیت پیچیده است. متغیر  $X$  به تنهایی احتمالاً همان اثری را که در مقابله با  $X$  دیگری نشان می‌دهد، ندارد. در نتیجه یافته‌های به دست آمده از یک طرح یک‌متغیری می‌تواند بدون معنا باشد. برای مثال، اثر یک روش خاص تدریس ممکن است به تعداد متغیر (نظیر سطح هوش دانش‌آموزان، ویژگیهای شخصیتی معلم، فضای عمومی کلاس و غیره)، بستگی داشته باشد. برای مثال، احتمال دارد آموزش با کمک کامپیوتر در مورد دانش‌آموزان کم‌هوشتر مؤثرتر از دانش‌آموزان باهوش باشد. یک طرح تک‌متغیری کلاسیک اثر تعاملی روش و سطح هوش را نشان نمی‌دهد. صحت اطلاعات به دست آمده از یک بررسی را می‌توان با استفاده از یک طرح که اثرات دو و یا چند متغیر مستقل را به صورت همزمان مشخص می‌کند، تضمین کرد. در حقیقت گفته می‌شود که پیشرفت غیر مترقیه تحقیقات در زمینه تعلیم و تربیت با تکامل طرحهای عاملی فیشر<sup>(۱)</sup> به دست آمده است.

طرح عاملی، طرحی است که در آن دو و یا چند متغیر به طور همزمان دستکاری می‌شوند تا اثر مستقل هر یک از آن متغیرها بر روی متغیر وابسته و همچنین اثرات ناشی از تعامل بین چند متغیر بررسی شود.

به طور کلی، طرحهای عاملی به دو نوع تقسیم می‌شوند. در نوع اول یکی از متغیرهای مستقل می‌تواند به طور آزمایشی دستکاری شود. در طرح نوع اخیر، آزمایشگر اصولاً می‌خواهد اثرات یک متغیر مستقل واحد را مورد بررسی قرار دهد، ولی باید دیگر متغیرهایی را که احتمال دارد بر متغیر وابسته اثر گذارند، در نظر گیرد. معمولاً این متغیرهای دیگر، متغیرهای خصیصه‌ای نظیر جنس، هوش، نژاد، موقعیت اجتماعی - اقتصادی، پیشرفت تحصیلی و نظایر اینها می‌باشند. اثرات این متغیرها با وارد کردن مستقیمشان در طرح عاملی، مورد بررسی (و همزمان تحت کنترل)، قرار می‌گیرند. آزمایشگر اثر متغیر مستقل اصلی را در هر یک از سطوح یک یا چند متغیر خصیصه‌ای مستقل، ارزشیابی می‌کند. سطوح مختلف متغیرهای خصیصه‌ای، معمولاً به صورت طبیعی گروههای منتخب آزمودنیها را شکل می‌دهد. برای مثال در یک بررسی برای تبیین اثر یک شیوه تدریس، از دانش آموزان تیزهوش و کندآموز، استفاده می‌شود. وارد نمودن متغیرهای خصیصه‌ای نه تنها دقت عمل یک بررسی، بلکه قابلیت تعمیم آن را نیز افزایش می‌دهد. از آنجا که آزمایشگر می‌تواند در مورد اینکه آیا تیمار بر روی کلیه سطوح، اثرات قابل ملاحظه دارد یا خیر، نتیجه‌گیری کند، قابلیت تعمیم یافته‌های آزمایشی افزایش می‌یابد.

در نوع دوم طرح عاملی، کلیه جنبه‌های مستقل به صورت آزمایشی دستکاری می‌شوند. در این نوع طرحها، آزمایشگر می‌خواهد چند متغیر مستقل فعال را بررسی کند و همچنین مایل است اثرات آنها را به تنهایی و نیز به صورت ترکیبی بسنجد. هر دو متغیر مستقل به صورت آزمایشی دستکاری می‌شوند. برای مثال، آزمایشگر ممکن است اثرات اندازه کلاس را همراه با آموزش از طریق کامپیوتر، بر روی یادگیری مفاهیم علمی، مقایسه کند. در این بررسی، هر دو متغیر دستکاری می‌شوند. دو تیمار از متغیر روش آموزش وجود دارد که عبارتند از: روش آموزش با کمک کامپیوتر، در مقابل روش سنتی. متغیر دوم، اندازه کلاس، نیز وجود دارد که دارای دو سطح کلاس بزرگ در مقابل کلاس کوچک است. چنین طرحی تحلیل اثرات اصلی برای هر دو متغیر آزمایشی و همچنین تحلیل تعامل بین تیمارها را امکان‌پذیر می‌سازد.

### طرح ۸ - طرح عاملی ساده

طرحهای عاملی در سطوح مختلف پیچیدگی، تکامل یافته‌اند. ساده ترین طرح عاملی، طرح

۲ در ۲ (۲×۲) است، در این طرح هر یک از دو متغیر مستقل دارای دو سطح هستند.

برای توضیح بیشتر، فرض می‌کنیم آزمایشگری می‌خواهد تأثیر دو نوع روش تدریس (روش الف و ب) را بر روی پیشرفت تحصیلی دانش آموزان کلاس نهم در درس علوم مقایسه کند.

طرح ۸ - طرح عاملی ساده

متغیر آزمایشی ( $X_1$ )		متغیر کنترل ( $X_2$ )
تیمار A	تیمار B	
خانه ۱	خانه ۳	سطح ۱
خانه ۲	خانه ۴	سطح ۲

او عقیده دارد که اثرات متفاوتی در اثر این دو روش، بر مبنای سطح هوش دانش آموزان، می‌تواند وجود داشته باشد. او بر مبنای نمرات هوشبهر، جامعه را به دو طبقه هوشبهر بالا و پایین طبقه بندی می‌کند و به صورت تصادفی، ۶۰ آزمودنی را از بین افراد گروه هوشبهر بالا انتخاب می‌کند، ۳۰ نفر آنها را به روش الف و ۳۰ نفر دیگر را به روش ب تخصیص می‌دهد. همین روش را در مورد گروه هوشبهر پایین اعمال می‌کند. معلمان را نیز به صورت تصادفی برای آموزش گروهها تعیین می‌کند.

در این نمونه آزمایش فرضی، دو نوع تیمار آزمایشی و دو سطح هوش داریم. طرح عاملی  $2 \times 2$  برای اندازه گیری اثر دو روش تدریس در یادگیری دانش آموزان در جدول ۱ - ۹ آمده است. به خاطر داشته باشید که طرح  $2 \times 2$  نیاز به ۴ گروه آزمودنی دارد. آزمودنیهای هر یک از دو سطح هوشبهر به صورت تصادفی به دو تیمار خاص تخصیص می‌یابند.

نمرات ۴ خانه، نشان دهنده میانگین نمرات متغیر وابسته یا آزمون پیشرفت در درس علوم مربوط به چهار گروه است. علاوه بر نمرات چهار خانه، که معرف ترکیبهای مختلف در سطوح است، نمرات میانگین حاشیه‌ای را نیز مشاهده می‌کنیم؛ دو میانگین برای ستونها و دو میانگین برای ردیفها. میانگینهای ستونی حاشیه‌ای مربوط به دو روش یا تیمارها و میانگین ردیفی حاشیه‌ای مربوط به دو سطح هوش است.

از اطلاعاتی که به دست می‌آید، ابتدا می‌توانیم اثرات اصلی<sup>(۱)</sup> دو متغیر مستقل را تعیین

کنیم. نمرات میانگین تیمار، بدون در نظر گرفتن سطح هوشبهر، نشان دهنده اثر اصلی تیمارهاست. چنانچه میانگین روش الف برای دو گروه (۶۷/۵) را با میانگین روش ب برای دو گروه (۶۸/۵) مقایسه کنیم، مشاهده می‌کنیم که تفاوت این دو میانگین فقط یک نمره است. بنابراین این احتمالاً وسوسه می‌شویم که چنین نتیجه‌گیری کنیم که روش ب به کار برده شده تأثیر کمی بر روی متغیر وابسته داشته است.

اکنون بهتر است نمره‌های میانگین برای سطوح مختلف را برای تعیین اثر اصلی  $X_1$  یا سطح هوش، بر روی پیشرفت تحصیلی بررسی کنیم. اثر اصلی برای سطوح مختلف، هیچ اثر افتراقی ناشی از تیمار را نشان نمی‌دهد. نمره میانگین برای دو گروه با هوشبهر بالا ۷۴ و برای دو گروه با هوشبهر پایین ۶۲ است. از آنجا که تفاوت ۱۲ واحد است، می‌توانیم این تأثیر را به هوشبهر نسبت دهیم. میانگین نمرات گروه با هوشبهر بالا به صورت قابل ملاحظه‌ای بالاتر است، بنابراین این بدون در نظر گرفتن تیمار می‌توان نتیجه گرفت که افراد با هوشبهر بالاتر از افراد با هوشبهر پایین، موفق تر بوده‌اند.

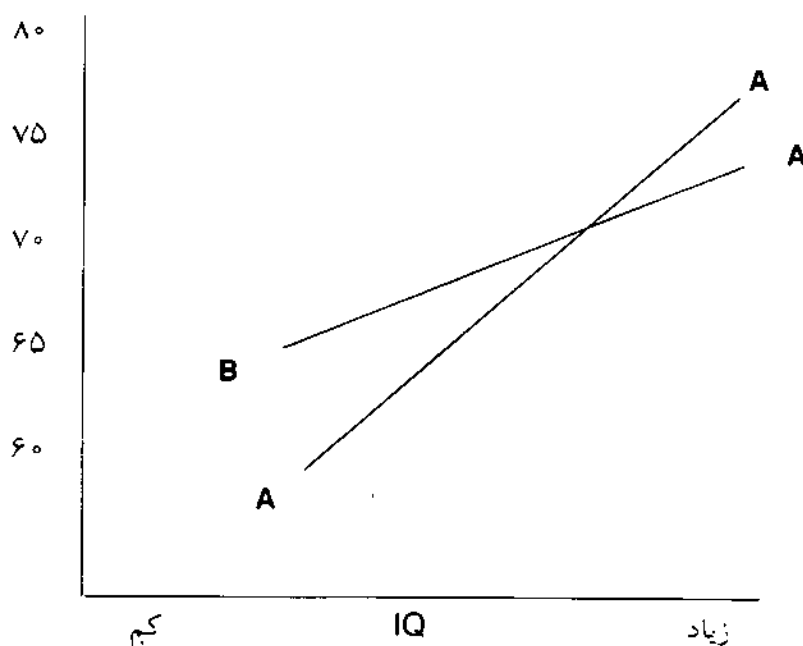
جدول ۹-۱ نمونه یک طرح عاملی

میانگین	روش آموزش ( $X_1$ )		هوشبهر
	روش ب	روش الف	( $X_2$ )
۷۴	۷۳	۷۵	بالا
۶۲	۶۴	۶۰	پایین
	۶۸/۵	۶۷/۵	میانگین

بعلاوه طرح عاملی امکان سنجش تعامل بین دو متغیر مستقل را برای محقق فراهم می‌سازد، یعنی اثرات مختلف یکی از متغیرها بر سطوح مختلف متغیر دیگر. چنانچه تعاملی وجود داشته باشد، اثری که تیمار بر روی یادگیری دارد، برای دو سطح هوشبهر متفاوت خواهد بود. چنانچه تعاملی وجود نداشته باشد، اثر تیمار برای هر دو سطح هوش یکسان خواهد بود. از بررسی نمرات میانگین جدول ۹-۱ در می‌یابیم که روش الف در مورد افراد با هوشبهر بالا، مؤثرتر از روش ب و روش ب در مورد افراد با هوشبهر پایین مؤثرتر از روش الف است. بنابراین ترکیب

خاصی از تیمار و سطح هوشیهر به گونه‌ای عمل می‌کند که اثرات بیشتری از دیگر ترکیبات دارد. اثر تعامل بین روش و سطوح هوش به صورت نگاره‌ای در شکل ۹-۲ ارائه شده است. چنانچه این تعامل از نظر آماری معنی دار باشد، نتیجه گیری می‌کنیم که مؤثر بودن روش بستگی به سطح هوشیهر دارد. یک روش در مورد یک سطح هوش مؤثرتر است و بالعکس همین امر در مورد سطح دیگر صادق است.

شکل ۹-۲ نمایش تعامل بین روش و سطح هوش



جدول ۹-۲ - نمونه یک طرح عاملی

میانگین	تیمار ( $X_1$ )		هوشیهر
	روش ب	روش الف	( $X_2$ )
۵۴	۵۸	۵۰	بالا
۴۴	۴۸	۴۰	پایین
	۵۳	۴۵	میانگین

حال به آزمون مجموعه دیگری از داده های حاصله از یک بررسی عاملی  $2 \times 2$  فرضی می پردازیم. جدول ۹-۲ نتایج بررسی طراحی شده جهت مطالعه اثر دو روش آموزش را بر روی پیشرفت تحصیلی نشان می دهد. در این مورد نیز، از آنجا که محقق انتظار دارد تأثیر روش به طور متفاوتی بستگی به سطح هوش آزمودنیها داشته باشد، از ابتدا دو سطح هوش را مشخص می کند. آزمودنیها در هر سطح به صورت تصادفی برای دو روش انتخاب می شوند. به دنبال اجرای مراحل آزمایشی، آزمون پیشرفت تحصیلی داده شده و نمرات هر آزمودنی ثبت می شود. چنانچه میانگین افراد دو گروه آموزش دیده با روش ب (۵۳) را با میانگین افراد دو گروه آموزش دیده با روش الف (۴۵) مقایسه کنیم، می بینیم که نمره اول تا اندازه ای بالاتر است. بنابراین، تیمار ب مؤثرتر از تیمار الف به نظر می رسد.

تفاوت بین میانگینها برای اثرهای اصلی ۱۰ است، یعنی ۵۴ در مقابل ۴۴. بدون در نظر گرفتن نوع تیمار، افراد گروه با هوشبهر بالاتر بهتر از افراد گروه با هوشبهر پایین تر عمل کرده اند. داده ها نشان می دهند که هیچ گونه تعاملی بین تیمار و سطوح هوشبهر وجود ندارد. روش ب در هر دو سطح مؤثرتر است. به عبارت دیگر، تیمار و سطح هوشبهر از یکدیگر مستقل هستند. عدم تعامل به صورت نگاره ای در شکل ۹-۳ نشان داده شده است. بدون استفاده از طرح عاملی، نشان دادن وجود و یا نبود چنین تعاملی امکان پذیر نیست.

طرح عاملی را می توان برای انجام آزمایشهای پیچیده تری، که در آنها تعدادی متغیر مستقل وجود دارد، بسط داد. مقادیر عددی ارقام، نمایانگر تعداد سطوح متغیرهای مستقل خاص هستند. برای مثال در طرح عاملی  $2 \times 3 \times 4$ ، سه متغیر به ترتیب با دو، سه و چهار سطح وجود دارد. چنین آزمایشی می تواند دو روش تدریس، سه سطح توانایی و چهار کلاس را در برگیرد. به طور نظری می توان گفت که یک طرح عاملی می تواند شامل هر تعداد متغیر مستقل، با هر تعداد سطح در هر یک از آنها باشد.

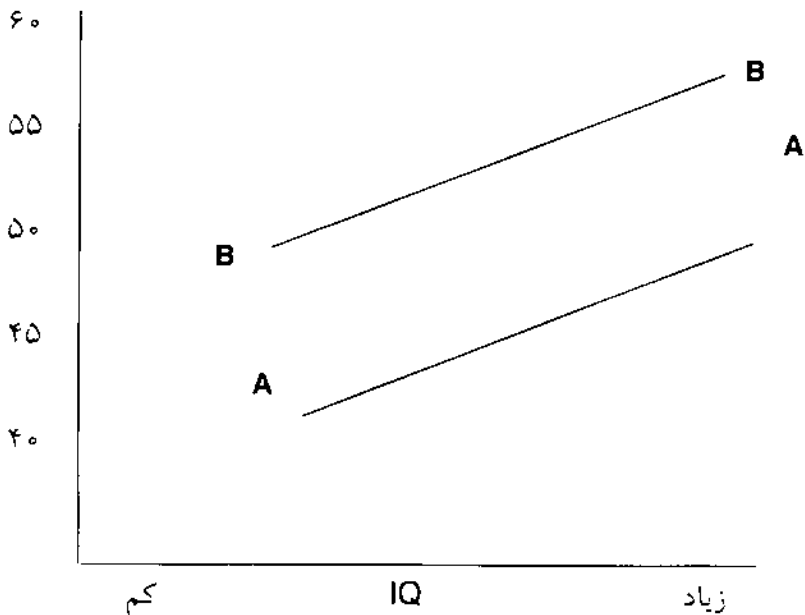
اما وقتی تعداد زیادی از عوامل به صورت همزمان دستکاری یا کنترل می شوند، بررسی و تحلیل آماری به صورت غیر ماهرانه در می آید و پاره ای از ترکیبات، تصنی می شود. تعداد گروههای مورد نیاز برای یک طرح عاملی برابر با حاصل ضرب ارقامی است که طرح عاملی را نشان می دهد، یعنی  $2 \times 3 \times 4 = 24$ . در یک طرح  $2 \times 3 \times 4$  برای ارائه تمام ترکیبات سطوح مختلف و متغیرهای مستقل نیاز به ۲۴ گروه داریم. تنها اندیشیدن در مورد پیچیدگی ترتیب دادن



تعداد زیاد آزمودنی تحت شرایط بسیار گوناگون، می تواند به خواننده کمک کند تا دلیل این امر را که محققان تعلیم و تربیت سعی دارند پاسخ سؤالاتشان را با ساده ترین طرح ممکن بیابند، درک کند و این کوشش حتی با وجود کامپیوترهای الکترونیک، که می تواند تحلیل آماری را انجام دهد، صورت می گیرد.

مزیت های طرح عاملی عبارتند از: ۱- می تواند آنچه را که نیاز به دو یا چند بررسی جداگانه دارد، در یک بررسی انجام دهد، ۲- امکان بررسی تعاملها را فراهم می آورد که معمولاً در تحقیقات آموزشی حائز اهمیت می باشند و ۳- برای آزمون فرضیه ها، آزمونهای قوی تری را ارائه می دهد.

شکل ۳-۹. نمایش عدم تعامل بین روش و سطح هوش



// طرحهای شبه آزمایشی

هدف محقق استفاده از طرحهایی است که کنترل کامل آزمایشی را از طریق استفاده از روش

توزیع تصادفی فراهم آورد. این گونه طرحها، طرحهای آزمایشی واقعی هستند که در بخشهای قبل ذکر شدند (طرحهای ۳ تا ۸). به هر حال، بسیاری موقعیتهای، بخصوص در زمینه تعلیم و تربیت، وجود دارند که در آنها امکان اجرای آزمایش واقعی وجود ندارد. به دلایل گوناگون، نه کنترل کامل برنامه ریزی شرایط آزمایشی و نه توانایی توزیع تصادفی، همیشه تحقق پذیر نیست. برای مثال، در تحقیقی که در شرایط کلاس درس اجرا می شود، احتمال دارد برای محقق امکان تخصیص تصادفی آزمودنیها به گروهها وجود نداشته باشد. در این مورد، محقق ناگزیر است به طرحهایی روی آورد که تا حد امکان در شرایط موجود، امکان کنترل را فراهم می آورند. این گونه طرحها به طرحهای شبه آزمایشی معروفند و در مواردی مورد استفاده قرار می گیرند که طرحهای آزمایشی واقعی قابل اجرا نباشند.<sup>(۱)</sup> از آنجا که طرح شبه آزمایشی امکان کنترل کافی را ایجاد نمی کند، از نظر محقق آگاهی داشتن از منابع اعتبار داخلی و خارجی و به کار گرفتن آنها در تفسیر، بسیار حیاتی است.

### طرح ۹ - گروه گواه غیر تصادفی با پیش آزمون و پس آزمون

گرچه تخصیص تصادفی آزمودنیها به گروهها کمال مطلوب است، ولی معمولاً در عمل امکان پذیر نیست. در موقعیت مدرسه، نمی توان برنامه های درسی را مختل کرد و یا به کلاسها دوباره سازمان داد تا بررسی آزمایش انجام شود. در این مورد، لازم است از گروههایی استفاده شود که از قبل در کلاسها سازمان یافته اند.

### طرح ۹ - گروه گواه غیر تصادفی با پیش آزمون و پس آزمون

گروه	پیش آزمون	متغیر مستقل	پس آزمون
E	$Y_1$	X	$Y_2$
C	$Y_1$	—	$Y_2$

احتمال دارد به محقق اجازه داده شود با چهار گروه دوره مقدماتی زبان انگلیسی در

1- Campbell, D. T., and J. C. Stanley (1966). *Experimental and Quasi - Experimental Designs for Research* (p. 34). Skokie, IL: Rand McNally.

دبیرستان، آزمایشی صورت دهد. از آنجا که جلسات درس در ساعات متفاوت تشکیل می شود، نمی توان آزمودنیها را به صورت تصادفی به تیمارها اختصاص داد، ولی محقق می تواند از روش تصادفی برای تعیین دو گروه آزمایشی و دو گروه گواه، استفاده کند. در مورد کلیه آزمودنیها، در ابتدای آزمایش، پیش آزمون و در خاتمه پس آزمون انجام می شود. از آنجا که در مورد گروههای آزمایشی و گواه، هم پیش آزمون و هم پس آزمون انجام می شود و طول زمان برای تمامی آزمودنیها یکسان است، انجام آزمون، وسیله اندازه گیری، بلوغ و میرایی به عنوان عوامل مختل کننده اعتبار داخلی تلقی نمی شوند.

چنانچه محقق مسئولیت کامل تیمار را بر عهده داشته باشد، تاریخچه، مشکلی نخواهد بود. اگر محقق تنها بر کار معلمان عادی که تیمارهای آزمایشی و گواه را ارائه می دهند، نظارت کند، تفاوتی بین معلمان می تواند به صورت نظام مند بر روی نتایج اثر بگذارد.

در هر دو طرح ۹ و ۵ - آزمودنیهای تصادفی، پیش آزمون - پس آزمون با گروه گواه - از تفاوتی بین نمرات پیش آزمون و پس آزمون به عنوان متغیر وابسته استفاده می شود. عدم انتخاب تصادفی آزمودنی در طرح ۹ می تواند متضمن سه نوع تهدید برای اعتبار داخلی باشد که در طرح ۵ به علت انتخاب تصادفی وجود ندارند.

۱- هر چند انتخاب، به تنهایی، مشکل اعتبار درونی تلقی نمی شود؛ زیرا آزمودنیها بر مبنای اهداف آزمایش و جهت دریافت تیمارهای آزمایشی و گواه به صورت تصادفی انتخاب نمی شوند، لکن تعامل بین انتخاب و بلوغ می تواند یک مسئله جدی اعتبار درونی باشد. فرض کنید کلاس درس انگلیسی گروه اول دانشجویان سال اول، همزمان با کلاس ریاضی تقویتی تشکیل می شود، در حالیکه کلاس گروه سوم دانشجویان سال اول با کلاس درس جبر پیشرفته همزمانی دارد. ما انتظار خواهیم داشت که گروه اول بهترین عملکرد تحصیلی را از خود نشان دهد، زیرا بسیاری از دانشجویان ضعیف تر آن، کلاس ریاضی تقویتی را انتخاب می کنند. همچنین انتظار خواهیم داشت که گروه سوم پیشرفت تحصیلی ضعیف تری داشته باشد، زیرا دانشجویان قوی تر آن، کلاس جبر پیشرفته را انتخاب می کنند. زمانبندی کلاسهای ریاضی بر روی ترکیب گروههای مختلف دانشجویان سال اول تأثیر می گذارد، و این امر به دلیل تعامل بین انتخاب و بلوغ، بر پیشرفت تحصیلی مؤثر است.

چنانچه بخواهیم افراد داوطلب شرکت در آزمایش را با افراد غیر داوطلب مقایسه کنیم،

تعامل انتخاب - بلوغ به صورت مشکلی ویژه، بروز می‌کند. برای مثال، کلاس تقویتی قرائت برای داوطلبان در مدرسه تشکیل شده است. میانگین نمرات پیش آزمون قرائت، تفاوتی را بین افراد داوطلب شرکت در این کلاسها و افراد غیر داوطلب نشان نمی‌دهد. چنانچه نمرات بعد از تیمار، یادگیری بیشتری را در مورد گروه تیمار نسبت به گروه گواه نشان دهد، نمی‌توانیم با اطمینان این یادگیری بیشتر را به تیمار نسبت دهیم. احتمال زیادی می‌رود که افراد مایل به شرکت در کلاس تقویتی مدرسه، به قرائت اهمیت بیشتری می‌دادند و احتمالاً این پیشرفت را بدون ارائه تیمار نیز نشان می‌دادند.

۲- رگرسیون آماری، مشکل اصلی اعتبار درونی در مورد طرح ۹ است؛ بدین معنا که نمرات بالا و پایین را به برگشت به طرف میانگین مشترک در اندازه گیریهای بعدی متعایل می‌سازد. چنانچه گروههای مورد استفاده در بررسی، از جامعه‌هایی با میانگینهای متفاوت انتخاب شده باشند، این تأثیر رگرسیونی در طرح وارد می‌شود. حتی اگر گروهها در پیش آزمون هم ارز باشند، تأثیر رگرسیون موجود می‌تواند منجر به یک جابجایی یا تغییر از زمان پیش آزمون تا پس آزمون شود که به غلط به عنوان تأثیر آزمایش تلقی می‌شود.

فرض کنیم در یک بررسی، میانگین نمرات پیش آزمون گروه آزمایشی ۷۵ باشد و این میانگین، پایین تر از میانگین جامعه‌ای باشد که نمونه از آن گرفته شده است، در حالی که میانگین ۷۵ گروه گواه بالاتر از میانگین جامعه خود است. از آنجا که در آزمون دوباره، دو گروه به طرف میانگین جامعه خود گرایش پیدا می‌کنند، گروه آزمایشی در پس آزمون به میانگین بالاتری می‌رسد. این امر چه X ارائه شود و چه نشود، صورت می‌گیرد. از طرف دیگر میانگین گروه گواه به طرف پایین میل می‌کند. در نتیجه ظاهراً چنین به نظر می‌رسد که در طول دوره آموزش، گروه آزمایشی پیشرفت بیشتری نسبت به گروه گواه داشته است و این امر سهواً به تأثیر X نسبت داده می‌شود.

۳- در طرح ۹ مشکلات جدی اعتبار درونی از نمره‌های تغییر یا «افزوده» پدید می‌آید، زیرا آزمودنیها به طور تصادفی به گروههای تیمار یا گواه تخصیص نیافته‌اند. هرچند پرداختن به مشکلات مربوط به نمرات تغییر به صورت تفصیلی در این کتاب غیرممکن است، ولی به چند مورد از مسائل می‌پردازیم. معمولاً یک همبستگی منفی میان نمرات پیش آزمون و میزان پیشرفت که از تفاوت بین پیش آزمون و پس آزمون به دست می‌آید، ملاحظه می‌شود. آیا این

بدان معناست که دانش آموزانی که نمرات اولیه آنها پایین تر بوده است، یادگیری بیشتری (بر مبنای اندازه گیری تغییر نمرات) از دانش آموزان با نمرات اولیه بالاتر، خواهند داشت؟ احتمالاً خیر. همبستگی منفی به احتمال زیاد ناشی از ویژگیهای روان سنجی خاص نمره تغییر است. به دلیل مشترکی، غالب آزمونهای آموزشی دارای سقف هستند، یعنی دامنه پیشرفت در پاسخگویی به سؤالات آزمون محدود است. چنانچه دانش آموزی از یک پیش آزمون ۱۰۰ سؤالی به ۹۲ سؤال جواب صحیح دهد، در پس آزمون تنها می تواند به میزان ۸ نمره پیشرفت نشان دهد. از سوی دیگر دانش آموزی که در یک پیش آزمون نمره ۴۲ می گیرد، در پس آزمون می تواند به اندازه ۵۸ نمره پیشرفت نشان دهد. به دلیل وجود این سقف، دانش آموزان متعلق به گروهی که در پیش آزمون، نمرات بالا داشته اند، از نظر نمره تغییر در پس آزمون، محدودیت دارند. مشکل مربوط به نمره های تغییر تنها از نظر اعتبار درونی، تهدیدی به شمار می رود و این مشکل زمانی پیش می آید که آزمودنیها برای تیمار به صورت تصادفی انتخاب نشده باشند. با انتخاب تصادفی، هر نوع تحریفی که به لحاظ استفاده از نمره های تغییر بوجود آمده است، در بلندمدت تأثیر یکسانی بر هر دو گروه آزمایشی و گواه خواهد داشت. در طرح ۹، مشکلات ناشی از نمرات تغییر وقتی بارز می شود که میانگین نمرات پیش آزمون گروههای آزمایشی و گواه به صورت قابل توجهی تفاوت داشته باشند؛ زیرا اثر سقف و رگرسیون احتمالاً بر نمرات پس آزمون یک گروه بیش از گروه دیگر خواهد بود. برای مثال، چنانچه طرح ۹ برای مقایسه اثرات دو روش آموزش املا به کار برده شود و در آن از آزمون املای یکسانی، حاوی ۱۰۰ سؤال در پیش آزمون و پس آزمون، استفاده شود و میانگین پیش آزمون یک گروه ۸۰ و گروه دیگر ۵۰ باشد، اثر سقف، میزان یادگیری احتمالی گروه اول را محدودتر از گروه دوم نشان خواهد داد.

در طرح ۹، خطرات مترتب بر اعتبار بیرونی، همان خطراتی هستند که در طرح ۵ با آن مواجه شدیم. به هر حال، یک مزیت طرح ۹ نسبت به طرح ۵، کنترل آسانتر تاثیرات واکنشی آزمایش است. چنانچه کلاسها به صورت دست نخورده مورد آزمون قرار گیرند، آزمودنیها کمتر نسبت به آزمایشی که در حال انجام است، وقوف خواهند داشت، تا این که از میان افراد انتخاب شوند و در جلسات آزمون شرکت کنند. این امر در قابلیت تعمیم یافته ها، دخیل است. همچنین احتمالاً ملاحظه خواهد شد که چنانچه آزمایشگر برای اجرای طرح خود از کلاس دست نخورده (مانند طرح ۹) استفاده کند، از نظر اخذ تأیید اداری هم با مشکل کمتری برخورد خواهد کرد.

هرچه گروه‌های آزمایشی و گواه در ابتدای آزمایش شباهت بیشتری داشته باشند و هرچه وجود این شباهت با همانندی میانگینهای گروهها در پیش آزمون تأیید شود، نتایج بررسی در طرح گروه گواه غیر تصادفی با پیش آزمون و پس آزمون، از اعتبار بیشتری برخوردار خواهد بود. چنانچه نمرات پیش آزمون مشابه باشند و بتوان نشان داد که احتمالاً دلیل تفاوت‌های مشاهده شده در پس آزمون، ناشی از تعامل انتخاب - بلوغ و یا رگرسیون نیست، نتایج این طرح شبه آزمایشی تا حدود زیادی قابل اعتماد خواهد بود.

حتی اگر تفاوت بین میانگین گروهها قبل از تیمار قابل توجه باشد، باز به کارگرفتن تک گروه گواه غیر هم ارز، بهتر از به کار نگرفتن گروه گواه به طور کلی است. این طرح نسبت به طرحهای ماقبل آزمایشی که فاقد گروه گواه هستند، برتری دارد. طرح ۹ را می‌توان به گونه‌ای بسط داد که در آن بیش از دو گروه به کار گرفته شوند.

### طرح ۱۰ - طرح متوازن

طرح ۱۰، طرح دیگری است که می‌توان در آن از گروه‌های کلاسی دست نخورده استفاده کرد و در فاصله زمانی انجام آزمایش، جای گروهها را با یکدیگر تغییر داد. برای مثال در نیمه اول آزمایش، می‌توان برای گروههای آزمایشی و گواه به ترتیب از روشهای الف و ب استفاده کرد و سپس در نیمه دوم آزمایش به ترتیب روشهای ب و الف را اعمال کرد. وجه تمایز طرح ۱۰ در این است که در طول آزمایش تمامی آزمودنیها کلیه تیمارها را دریافت می‌کنند. در حقیقت، این طرح شامل مجموعه‌ای از تکرار آزمایشهاست. در هر تکرار، گروههای آزمایشی به گونه‌ای جابجا می‌شوند که در انتهای آزمایش هر گروه در معرض تمام تیمارها قرار گرفته باشد. ترتیب قرار گرفتن در معرض موقعیتهای آزمایشی برای هر گروه متفاوت است. طرح متوازن معمولاً در مواردی به کار برده می‌شود که می‌خواهیم چند تیمار را آزمون کنیم، ولی می‌توان آن را تنها با دو تیمار نیز اجرا کرد.

در طرح ۱۰، هر ردیف نشان دهنده یک تکرار است. در هر تکرار، گروهها جابجا می‌شوند، به گونه‌ای که گروه الف ابتدا تیمار  $X_1$ ، بعد  $X_2$  و  $X_3$  و بالاخره  $X_4$  را تجربه می‌کند. هرخانه طرح شامل نمرات میانگین متغیر وابسته به گروه، تیمار و تکرار است. نمره میانگین هر ستون، نشان دهنده چگونگی عملکرد چهار گروه، در مورد متغیر وابسته مربوط به تیماری است که در بالای ستون مشخص شده است.

طرح ۱۰ - نمونه‌ای از طرح متوازن

تیمار آزمایشی				
$X_f$	$X_p$	$X_r$	$X_1$	تکرار
D	C	B	گروه A	۱
B	D	A	گروه B	۲
C	A	D	گروه C	۳
A	B	C	گروه D	۴

میانگین ستونی      میانگین ستونی      میانگین ستونی      میانگین ستونی

یک آموزگار می‌تواند جهت مقایسه میزان تأثیر دو روش تدریس در یادگیری علوم از بررسی متوازن استفاده کند، او می‌تواند دو کلاس و همچنین دو متن علوم را که از نظر مشکل بودن، طولانی بودن و ... تقریباً یکسان باشند، انتخاب کند. هم ارزی متون از نظر پیچیدگی و مشکل بودن مفاهیم آنها، ضروری است. در طول اولین اجرای طرح، متن ۱، برای کلاس ۱، با روش الف و برای کلاس ۲، با روش ب تدریس می‌شود. آزمون پیشرفت تحصیلی برای متن ۱ در مورد دو گروه انجام می‌شود. سپس برای کلاس ۱، متن ۲ را با روش ب و برای کلاس ۲، همان متن را با روش الف تدریس می‌کنند و بعد از هر دو گروه در مورد متن ۲ آزمون به عمل می‌آورند. ترتیب طرح در جدول ۳-۹ آمده است.

بعد از انجام بررسی، میانگینهای ستونها، به منظور مشخص کردن میانگین پیشرفت هر دو گروه (کلاسها)، با استفاده از روش تدریس قید شده در عنوان ستون، محاسبه می‌شود. مقایسه این میانگینهای ستونی، با بهره‌گیری از تحلیل عاملی، میزان تأثیر روشها را در پیشرفت یادگیری درس علوم مشخص می‌کند.

طرح ۱۰ فاقد پاره‌ای نقاط ضعف طرح ۹ است. بدین ترتیب چنانچه استفاده از کلاسهای دست نخورده ضرورت داشته باشد، تفاوت‌های احتمالی بین گروهها را از دور خارج می‌کند. از آنجا که تمام تیمارها برای تمام گروهها به کار برده می‌شود، نتیجه به دست آمده برای هر  $X$  را نمی‌توان به تفاوت‌های از پیش موجود بین آزمودنیها نسبت داد. چنانچه افراد یک گروه به طور متوسط از گروه دیگر باهوش‌تر باشند، تمام تیمارهای  $X$ ، از این هوش بالاتر برخوردار شده‌اند.

## جدول ۳-۹ نمونه یک طرح متوازن

تیمارهای آزمایشی		
تکرار	روش الف	روش ب
متن ۱	کلاس ۱	کلاس ۲
متن ۲	کلاس ۲	کلاس ۱
میانگین ستونی		میانگین ستونی

نقیصه اساسی طرح ۱۰، احتمال وجود یک اثر انتقالی از یک X به دیگری است. در نتیجه، از این طرح باید تنها در مواردی استفاده شود که تیمارهای آزمایشی به گونه‌ای باشند که قرار گرفتن در معرض یکی، هیچ گونه اثری بر تیمار بعدی نداشته باشد. برآوردن این ضرورت در بسیاری از تحقیقات آموزشی، امری دشوار است. علاوه بر این، تهیه مطالب یادگیری هم ارز، که در هر تکرار آزمایش باید مورد استفاده قرار گیرند، یک ضرورت به شمار می‌رود. شاید همواره امکان یافتن مطالب هم ارز وجود نداشته باشد. دیگر نقیصه طرح متوازن احتمال ایجاد خستگی در دانش آموزان به دلیل آزمونهای مکرر است، که لازمه این طرح می‌باشد.

## // طرحهای توالی زمانی

## / طرح ۱۱ - طرح توالی زمانی یک گروهی

طرح ۱۱ اندازه‌گیری دوره‌ای بر روی یک گروه و ارائه یک تیمار آزمایشی در طول این توالی زمانی اندازه‌گیریها است. همان گونه که در طرح مشاهده می‌شود، بر روی یک متغیر وابسته، چندین بار اندازه‌گیری صورت می‌گیرد. سپس تیمار X ارائه و اندازه‌گیریهای بعدی Y انجام می‌شود. از طریق مقایسه اندازه‌گیریها، قبل و بعد از ارائه X، می‌توان اثر X را بر نحوه عملکرد گروه در Y ارزیابی کرد. در شرایط مدرسه‌ای، یک طرح توالی زمانی را می‌توان برای بررسی اثرات تغییرات عمده در سیاستهای اداری مدرسه بر مسائل انضباطی، به کار برد. یا ممکن است یک بررسی شامل اندازه‌گیریهای مکرر نگرش دانش آموزان و اندازه‌گیری اثر یک فیلم مستند باشد، که برای تغییر نگرش دانش آموزان تهیه شده است.



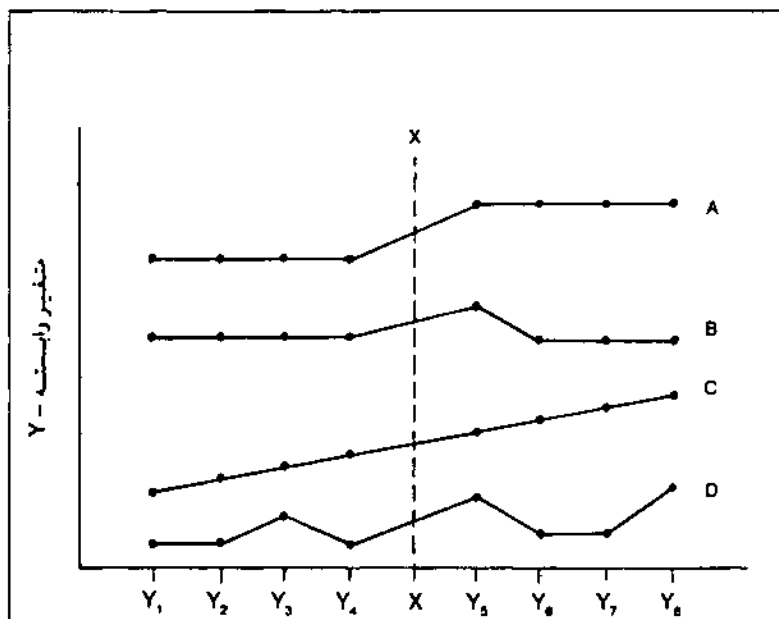
۱۱- طرح توالی زمانی یک گروهی

$Y_8$	$Y_7$	$Y_6$	$Y_5$	X	$Y_4$	$Y_3$	$Y_2$	$Y_1$
-------	-------	-------	-------	---	-------	-------	-------	-------

شکل ۴-۹، پاره‌ای از الگوهای احتمالی از بررسیهای توالی زمانی را نشان می‌دهند که در آنها یک تیمار آزمایشی ارائه شده است. این شکل مجموعه‌ای از اندازه‌گیریهای  $Y_1$  تا  $Y_8$  را با ارائه تیمار آزمایشی در نقطه X نشان می‌دهد. ما می‌توانیم اثرات X را با بررسی ثبات اندازه‌گیریهای مکرر، بسنجیم.

از بررسی تفاوت بین  $Y_4$  و  $Y_5$  در الگوی A، می‌توان فرض کرد که X بر روی متغیر وابسته اثر دارد. الگوی B نیز احتمال وجود اثر آزمایشی X را نشان می‌دهد. به هر حال، نمی‌توان فرض کرد که X در الگوهای C و D تغییری ایجاد می‌کند. الگوی C می‌تواند ناشی از عوامل بلوغ و یا اثری مشابه باشد. ماهیت عجیب و غریب طرح D، عملکرد عوامل نامربوط را خطاظر نشان می‌سازد.

شکل ۴-۹ نمونه‌ای از الگوهای احتمالی یک طرح توالی زمانی



طرح ۱۱ از این نظر که اندازه های قبل و بعد را مورد استفاده قرار می دهد و فاقد گروه گواه است، به طرح ۱ شباهت دارد، ولی از پاره ای جهات بر طرح ۱ مزیت دارد و همین امر باعث می شود که برای تحقیقات در زمینه تعلیم و تربیت مفیدتر باشد. آزمونهای متعدد بر خطرات متداول مترتب بر اعتبار درونی، نوعی واریسی اعمال می کند. اثرات عوامل بلوغ، آزمون کردن و رگرسیون را می توان با این دلیل موجه رد کرد که تغییری که بین  $Y_4$  و  $Y_5$  مشاهده می شود، در نمرات آزمونهای قبلی دیده نشده است. توصیه می شود که وسایل اندازه گیری به هیچ وجه در طول بررسی تغییر نکنند. از این طریق، اثر وسایل اندازه گیری در توصیف احتمالی تفاوت بین  $Y_4$  و  $Y_5$  حذف می شود.

نقطه ضعف عمده طرح ۱۱، فقدان کنترل گذشت زمان است. بدین ترتیب که باید این امکان را که به جای  $X$ ، تغییر مشاهده شده را رویداد همزمانی به وجود آورده باشد، حذف کرد. شاید عواملی چون فصل، تغییرات هوا و یا عوامل مربوط به مدرسه، مانند امتحانات، باعث ایجاد چنین تغییری شده باشند. در مطالعه ای که برای سنجش اثر ارائه یک سخنرانی توأم با نمایش فیلم، بر روی نگرش دانش آموزان نسبت به یکپارچگی مدرسه طراحی شده است، شورش سیاهپوستان در شهری دوردست، که در سطح کشور مورد بحث و گفتگو قرار گرفته است، تا چه اندازه بر اندازه گیریهای نگرش تأثیر خواهد گذاشت؟ محقق در حالی که سعی می کند یافته هایش را تفسیر کند، باید مشخص کند که گذشت زمان (رویدادهای کنترل نشده همزمان با انجام بررسی) تا چه میزان یک عامل تبیینی پذیرفتنی می باشد.

همچنین باید اعتبار بیرونی طرح زمانی را مد نظر داشت. از آنجا که در این طرح آزمونهای مکرر صورت می گیرد، نوعی اثر تعاملی از آزمون کردن می تواند وجود داشته باشد که موجب می شود یافته ها تنها به همان جامعه ای محدود شود که در معرض آزمونهای متعدد قرار گرفته اند. به هر حال تا آنجا که این اندازه گیریها به صورت معمولی و به گونه ای مورد استفاده قرار گیرند که معرف موقعیت مدرسه ای باشند، محدودیت جدی محسوب نمی شوند. بعلاوه نوعی تعامل «انتخاب -  $X$ » ممکن است بروز کند، بویژه در صورتی که گروه خاصی انتخاب شود که معرف جامعه نباشد.

تفسیر آماری در مورد داده های زمانی می تواند مشکل ویژه ای به شمار آید. از آنجا که نمرات فردی و میانگین در طول دوره آزمایش به میزان زیادی تغییر می کند، چنین تمایلی وجود دارد که

این تغییر را ناشی از تیمار X بدانیم، در حالی که ممکن است در واقع نتیجه متغیرهای دیگر باشد. آزمونهای معمولی معنی داری ممکن است برای طرح زمانی مناسب نباشند. در اینجا خوانندگان را ارجاع می‌دهیم به کتاب کمپ بل و استنلی، بخش مربوط به بحث در مورد آزمونهای آماری که در طرح زمانی مورد استفاده قرار می‌گیرد.

### / طرح ۱۲ - طرح توالی زمانی با گروه گواه

طرح ۱۲ تکامل یافته طرح ۱۱ است که در آن گروه گواه اضافه می‌شود. گروه گواه، در اینجا نیز معرف همان گروه کلاسی دست نخورده است که همزمان با گروه آزمایشی مورد آزمون قرار می‌گیرد، ولی در معرض تیمار X قرار نمی‌گیرد. این طرح بر نقیصه طرح ۱۱، که همان فقدان گذشت زمان به عنوان منبع واریانس بیرونی است، فائق می‌آید. گروه گواه مقایسه لازم را امکان پذیر می‌سازد. چنانچه گروه آزمایشی از موقعیت  $Y_p$  به  $Y_o$  یادگیری نشان دهد، ولی گروه گواه چنین یادگیری‌ای را نشان ندهد، می‌توان اثر را ناشی از X دانست، و نه از رویداد همزمان دیگری که بر هر دو گروه اثر می‌گذارد.

انواع دیگری از طرحهای توالی زمانی با افزودن گروههای گواه بیشتر، مشاهده بیشتر و یا تیمارهای آزمایش بیشتر به دست می‌آید.

### طرح ۱۲ - طرح توالی زمانی با گروه گواه

گروه									
$Y_8$	$Y_7$	$Y_6$	$Y_5$	X	$Y_p$	$Y_p$	$Y_p$	$Y_1$	E
$Y_8$	$Y_7$	$Y_6$	$Y_5$	—	$Y_p$	$Y_p$	$Y_p$	$Y_1$	C

### // مشکلات اعتبار طرحهای آزمایشی

پاره‌ای از منابع عدم اعتبار در طرحهای آزمایشی یک متغیری در جدول ۹-۴ خلاصه شده است. از این خلاصه کوچک، نمی‌توان به عنوان راهنمای مطلق جهت انتخاب استفاده کرد، بلکه طرح را باید همراه با آنچه به صورت کامل در متن ارائه شده است، در نظر داشت تا خواننده بتواند نقاط قوت و ضعف خاصی را که هر طرح مشخص می‌کند، درک نماید.

## // طرحهای آزمایشی تک آزمودنی

چنین به نظر می‌رسد که عنوان طرح آزمایشی تک آزمودنی نوعی تناقص گویی است. چگونگی امکان دارد که آزمایش را با نمونه‌ای یک نفری انجام داد؟ روشن است که تخصیص تصادفی و استفاده از گروه گواه وجود نخواهد داشت. به رغم مسائل یاد شده، استفاده از تحقیقاتی که در آن تنها از یک آزمودنی استفاده می‌شود، در ۲۰ سال اخیر متداول شده است. دلایل طرفداران استفاده از این روش این است که کنترل آزمایش از طرقی غیر از راههای سنتی نیز امکان‌پذیر است. پس از توصیف دو شیوه اصلی تحقیقات تک‌آزمونی و استدلالی که در ورای آن قرار دارد، به بررسی نقاط قدرت و ضعف این نوع تحقیق در مقایسه با دیگر طرحهای مرسوم می‌پردازیم.

بررسی بر روی فرد، همواره در تحقیقات آموزشی و روان‌شناختی جای خود را داشته است. در این مورد پژوهشهای فروید و مشاهدات پیازه بر روی کودکان به صورت منفرد، نمونه‌های بارزی از این نوع تحقیق هستند.

اگرچه، هم در مورد پژوهیها (رجوع شود به فصل ۱۱) و هم در آزمایشهای تک آزمودنی، محقق به بررسی فرد می‌پردازد، لکن تمایز این دو شیوه در این است که در آزمایش تک‌آزمودنی محقق تماماً یک یا چند متغیر مستقل را دستکاری می‌کند، در حالی که در مورد پژوهی، مشاهده‌گر صرفاً به بررسی واکنشهای فرد در مقابل حوادثی می‌پردازد که به صورت طبیعی اتفاق می‌افتند. طرحهای تک آزمودنی بویژه کاربردهای کلینیکی - بالینی دارد؛ زیرا در این گونه موارد تأکید بر روی ارزشهای درمانی، اقدامی است که به مراجعه‌کننده توصیه می‌شود. برای مثال، آموزگار کودکان کند ذهن، نیاز به اطلاعاتی در مورد میزان اثربخشی یک شیوه بخصوص در مورد یک کودک منفرد دارد. چنین احساس می‌شود که بررسیهایی که تفاوت‌های میانگین و متوسط را برای گروهها گزارش می‌کنند، در مداوای یک فرد معنایی ندارند.

طرحهای تک آزمودنی در واقع تکامل یافته طرحهای زمانی یک گروهی شبه آزمایشی (طرح ۱۱) هستند. متداول‌ترین آنها عبارتند از: ABAB و طرحهای خط پایه چندتایی.

طرح ABAB شامل یک دوره بدون تیمار یا خط پایه A می‌شود که در طول آن رفتار مورد نظر چندین بار اندازه‌گیری می‌شود. از این رفتارها، به عنوان مثال می‌توان از تعداد دفعاتی که دانش‌آموز استثنایی صندلیش را ترک می‌کند و یا تعداد دفعاتی که یک کودک در خود مانده،

جدول ۴-۹ عواملی که اعتبار طرحهای آزمایشی را خدشه دار می سازند

منابع مخدوش کننده اعتبار	طرحها*											
	مقابل آزمایشی	آزمایشی دائمی										
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
<u>اعتبار درونی</u>												
گذشت زمان	-	+	+	+	+	+	+	+	+	+	+	+
فرآیندهای بلوغ	-	+	+	+	+	+	+	+	+	+	+	+
شیوه های پیش آزمون کردن	-	+	+	+	+	+	+	+	+	+	+	+
وسایل اندازه گیری	-	+	+	+	+	+	+	+	+	+	+	+
رگرسیون آماری	?	+	+	+	+	+	+	+	+	+	+	+
انتخاب افتراقی آزمودنیها	+	-	+	+	+	+	+	+	+	+	+	+
افت آزمایشی	+	-	+	+	+	+	+	+	+	+	+	+
تعامل بین انتخاب و بلوغ و موارد نظیر	-	-	+	+	+	+	+	+	+	+	+	+
<u>اعتبار بیرونی</u>												
تعامل انتخاب و X	-	-	?	?	?	?	?	?	?	?	?	?
تعامل پیش آزمون و X	-	-	?	-	+	-	+	-	+	+	+	+
شیوه های آزمایشی واکنشی	?	?	?	?	?	?	?	?	?	?	?	?
تداخل چند تیمار	-	-	-	-	-	-	-	-	-	-	-	-

۵ - طرحها عبارتند از :

۱- بکن گروهی پیش آزمون و پس آزمون

۲- مقابله گروهی ایستا

۳- گروه گواه، آزمودنیهای تصادفی، تنها پس آزمون

۴- آزمودنیهای همتا شده تصادفی، تنها پس آزمون با گروه گواه

۵- آزمودنیهای تصادفی، پیش آزمون - پس آزمون با گروه گواه

۶- سه گروهی سولوگون

۷- چهار گروهی سولوگون

۸- عاملی ساده

۹- گروه گواه غیر تصادفی با پیش آزمون و پس آزمون

۱۰- متوازن

۱۱- نوالی زمانی یک گروهی

۱۲- نوالی زمانی با گروه گواه

علامت «علاوه» نشان دهنده عامل کنترل شده؛ علامت «منها» نشان دهنده فقدان کنترل و علامت «سؤال» نشان دهنده منبع دخالت احتمالی، و دجایهای خالی، نشان دهنده عامل نامربط است.

سرش را به دیوار می‌زند، نام برد. این مرحله ارزشیابی قبل از تیمار، در واقع به عنوان دوره گواه تلقی می‌شود و اثرات تیمار، با اطلاعات به دست آمده از این مرحله، مقایسه می‌شود. پس از آنکه تصویر ثابتی از رفتار قبل از دریافت تیمار به دست آمد، مرحله B با ارائه تیمار شروع می‌شود. در مرحله تیمار می‌توان برای مثال به کودک برای اینکه زمان خاصی در صندلیش باقی بماند، ژتونهایی داد (این ژتونها قابل تبدیل به مزیت مورد نظر او خواهد بود) و یا کودک دیگر پس از نشان دادن هر عکس‌العملی حاکی از صدمه زدن به خود، یک شوک ملایم دریافت کند. رفتار مرتباً در طول مرحله تیمار مشاهده و ثبت می‌شود و این عمل معمولاً تا زمانی که به نظر رسد اثر تیمار، ظاهر و میزان رفتار، ثابت شده است، تکرار می‌شود. بعلاوه کنترل آزمایش از طریق اجرای مرحله دوم A صورت می‌گیرد. معمولاً این مرحله حذف تیمار است، لکن در برخی موارد، مرحله دوم A عملاً نقض تیمار است؛ به این معنا که رفتاری که با پاسخ مورد نظر همخوانی ندارد، تقویت می‌شود. قطع ژتون برای کودک عقب مانده، حذف تیمار است؛ در حالی که دادن ژتون در صورت ترک صندلی، نقض تیمار است. در هر دو مورد، با حذف تیمار انتظار می‌رود که رفتار به سطح رفتار اصلی (خط پایه اول) برگردد. چنانچه آزمایش را در مرحله دوم A خاتمه دهیم، این اشکال پیش می‌آید که بیمار به همان وضعیت قبل از شروع آزمایش بر می‌گردد. به دلایل اخلاقی و همچنین به منظور افزایش قوت طرح با تکرار روش، مرحله B تیمار دوباره اجرا می‌شود. در طرح ABAB ایجاد تغییرات مختلفی امکان پذیر است. برای مثال با استفاده از طرح ABAB می‌توان بیش از یک تیمار را آزمود. در این طرح، A، مرحله بدون تیمار و B و C تیمارهای مختلف هستند.

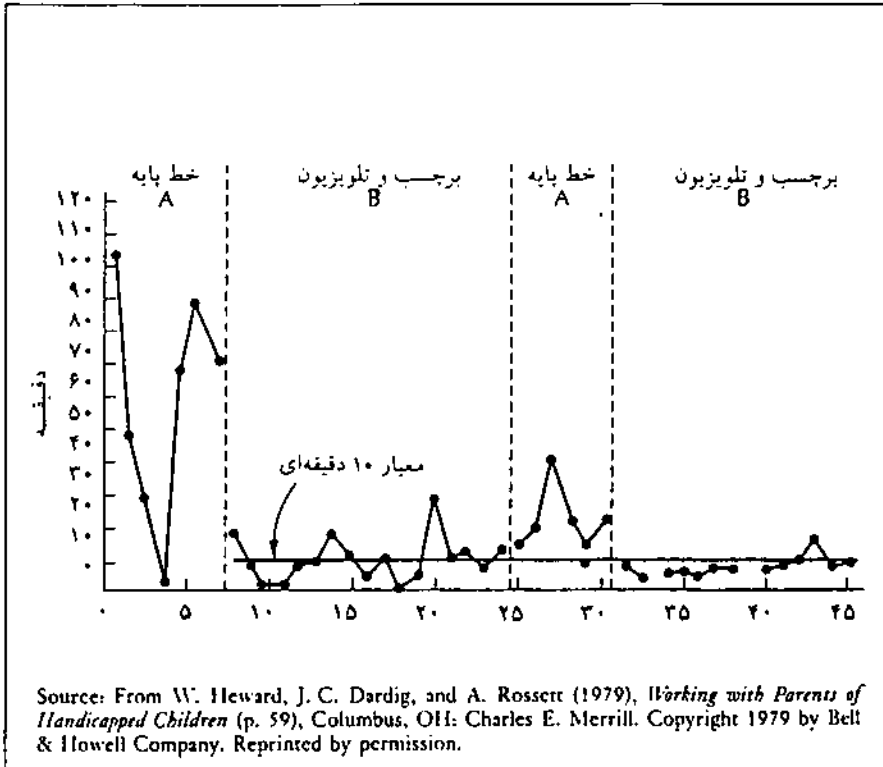
شکل ۹-۵ یک طرح ABAB را نشان می‌دهد. تمام افراد خانواده یک پسر ۸ ساله عقب‌افتاده به نام کرت<sup>(۱)</sup>، صبحها مشکل داشتند، زیرا دو ساعت طول می‌کشید تا او لباس بپوشد. یادآوری مداوم، خواهش کردنها و سرزنش کردنهای گاه‌به‌گاه، هیچکدام اثری نداشتند. یک بار در طول

دوره یک هفته‌ای اندازه‌گیری خط پایه A، کرت در مدت فقط ۴ دقیقه لباس پوشید و در نتیجه والدینش فهمیدند که او قادر است خودش در زمان قابل قبولی لباس بپوشد.

در طول مرحله تیمار (B)، هر روز صبح لباس کرت در حمام قرار داده می‌شد. او را بیدار می‌کردند و زمان سنجی برای مدت ده دقیقه در آشپزخانه تنظیم می‌شد. چنانچه کرت قبل از به صدا درآمدن زمان سنج لباس می‌پوشید، یک برچسب قرمز بر چهارچوب در اطاقش نصب می‌شد که بدان معنی بود که او می‌تواند آن شب تلویزیون تماشا کند. چنانچه قبل از پایان ده دقیقه لباس نمی‌پوشید، باید تا زمان لباس پوشیدن در حمام باقی می‌ماند و برای آن روز جریمه می‌شد و نمی‌توانست تلویزیون تماشا کند. کرت در صورتی که قبل از زمان موعود لباس می‌پوشید، مورد تحسین واقع می‌شد و در غیر این صورت، او را نادیده می‌گرفتند.

در طول هفته شروع، میانگین زمان لباس پوشیدن کرت ۵۹ دقیقه بود. در طول ۱۷ روز مرحله اول B، ۹ بار قبل از زمان موعود لباس پوشید و میانگین او ۱۰ دقیقه بود. در سه مورد، زمان لباس پوشیدنش صفر دقیقه ثبت شد، زیرا قبل از بیدار شدن والدینش بیدار شده و لباس پوشیده بود. در طول ۶ روز بعد، لباس کرت آماده می‌شد، ولی زمان سنج حذف شده بود و بدون در نظر گرفتن زمانی که او صرف لباس پوشیدن می‌کرد، اجازه داشت تلویزیون تماشا کند. در طول این مرحله دوم A، میانگین زمان لباس پوشیدن او ۲۳ دقیقه شد. زمانی که تیمار برای مدت ۱۲ روز از ۱۳ روز باقی‌مانده طول طرح، دوباره ارائه شد، میانگین زمان لباس پوشیدن کرت به ۸ دقیقه رسید. از آنجا که تیمار همواره با تغییری در زمان لباس پوشیدن همراه بود، رابطه قابل ملاحظه‌ای میان زمان لباس پوشیدن و تیمار به وجود آمده بود. عنصر کلیدی در طرح ABAB عبارت است از بازگشت به سطوح قبلی رفتار، زمانی که وضعیت خط پایه دوباره برقرار می‌شود. فرض اینکه مداخله، دلیل به وجود آمدن تغییر است، تحت شرایط خاصی به صورت قبلی ملاحظه‌ای ضعیف می‌شود. چنانچه تغییرات قابل ملاحظه‌ای در مرحله خط پایه مشاهده شود، می‌توان چنین نتیجه‌گیری کرد که رفتار پس از تیمار با رفتار قبل از تیمار تفاوت معنی‌داری ندارد. در صورتی که تغییر از تیمار (B) به خط پایه A قابل اغماض باشد، بخش اعظم قوت طرح ABAB از بین رفته است. خط پایه دوم معمولاً همان شدت رفتار خط پایه اول را نشان نمی‌دهد، و در برخی موارد بازگشتی به میزانهای قبلی اتفاق نمی‌افتد. در برخی موارد، دیگر عوامل چون، سابقه و بلوغ می‌توانند عوامل اثر مشاهده شده باشند.

شکل ۵-۹ طرح ABAB استفاده از تقویت افتراقی جهت تقلیل زمان لباس پوشیدن یک پسر ۸ ساله در صبحها



کازدین<sup>(۱)</sup> (۱۹۸۰) چنین می‌گوید که بازگشت به شرایط خط پایه، تحت شرایطی، اخلاقی نیست. آیا واقعاً می‌خواهیم کودک در خود مانده به سطح قبل از تیمار برگردد و سرش را به دیوار بکوبد، تنها برای آنکه نشان دهیم که دلیل تقلیل رفتار ناهنجار او، تیمار بوده است؟ به دلیل این مشکلات بالقوه طرح ABAB، طرح خط پایه چند تایی مطرح شد. از نظر کلینیکی، این طرح از این مزیت برخوردار است که در آن تداوم پیشرفت وجود دارد؛ بدون آنکه به رفتاری که مطلوبیت کمتری دارد، برگردیم.

در یک طرح خط پایه چند تایی، اندازه‌گیری رفتارهای مختلف به صورت همزمان در شرایط

1- Kazdin (1980), *Research Design in Clinical Psychology*, New York, Harper and Row.



خط پایه صورت می‌گیرد. برای مثال، محقق می‌تواند تعداد دفعاتی را که دانش‌آموز بدون اجازه در کلاس صحبت کرده، تعداد تکالیفی که انجام داده و تعداد دفعاتی که دانش‌آموزی، دانش‌آموز دیگر را کنک زده است، ثبت کند. یا در مورد دو یا چند کودک کلاس آمادگی می‌تواند یک رفتار مشخص را در مرحله شروع، مثل میزان برقرار کردن تماس با معلم از طریق نگاه، ثبت کند. در هر دو مورد تیمار (B) در تمام شرایط و برای تمام دانش‌آموزان یکسان است. کنترل آزمایشی در طرح خط پایه چند تایی، به جای بازگشت به مرحله شروع از طریق شروع تیمار در مقاطع زمانی مختلف برای هر رفتار، یا هر فرد مورد نظری انجام می‌گیرد. در نتیجه بعد از آنکه خط پایه تعیین شد، تیمار برای رفتار ۱ ارائه و از آنجا خط پایه برای رفتار ۲ و ۳ شروع می‌شود. زمانی که رفتار ۳ تعیین شد، تیمار برای رفتار ۱ و خط پایه برای رفتار ۳ ادامه می‌یابد. در نهایت، تیمار برای رفتار ۳ مقرر می‌شود. چنین انتظار می‌رود که هر رفتار در جهت مطلوب در مرحله‌ای که تیمار شروع می‌شود، تغییر کند و این تغییر قبل و بعد از آن صورت نگیرد.

در نتیجه طرح خط پایه چند تایی از AB به عنوان واحد مبنا استفاده می‌کند. چنانچه حادثه‌ای سوای تیمار عامل واقعی به وجود آمدن تغییرات باشد، قاعدتاً باید تمام کودکان و یا تمام رفتارها را در یک زمان تحت تأثیر قرار دهد. فرض در مورد این طرح بدین صورت وجود دارد که تیمار رفتارهای مختلف را به صورت ویژه تحت تأثیر قرار می‌دهد. دلیلی ندارد تقویت یک رفتار (مثلاً تمام کردن مسائل حساب)، پاسخی دیگر (میزان خواندن) را افزایش دهد. چنانچه بخواهیم بررسی انجام شده با استفاده از طرح خط پایه چند تایی اثرات قابل تفسیر ارائه دهد، باید رفتارها یا موقعیتها کاملاً از یکدیگر مستقل باشند (بین آنها همبستگی وجود نداشته باشد). در عمل، دستیابی به رفتارهای مستقل دشوار است. تغییر یک رفتار (مثلاً صحبت کردن در کلاس)، بر رفتارهای موردنظر (تمام کردن تکالیف در زمان مقرر) اثر می‌گذارد.<sup>(۱)</sup>

### // مقایسه طرحهای تک آزمودنی و گروهی

هدف آزمایشگر از انجام هر دو شیوه آزمایشی تک آزمودنی و گروهی آن است که حتی الامکان، رابطه بین دستکاری متغیر مستقل (تیمار) و اثرات آن بر متغیر وابسته (رفتار) را به

1- Kazdin, A., and S. Kopel (1975). On resolving ambiguities of the multiple baseline design: Problems and recommendations. *Behavior Therapy*, 6, 601 - 608

صورت روشن و بدون ابهام تعیین کند. در طرحهای گروهی، تخصیص تصادفی آزمودنیها به گروههای آزمایشی و گواه، بسیاری از تبیینهای هماورد را در مورد تفاوتیابی که بعد از تیمار مشاهده می‌شوند، حذف می‌کند.

با استفاده از آزمونهای آماری مناسب می‌توان اثرات تیمار (بین گروهها) را نسبت به اثرات تغییرپذیری بین آزمودنی (درون گروه) مورد سنجش قرار داد. نتایج این آزمون نشان می‌دهد که آیا عامل احتمال به تنهایی می‌تواند توجیه معتبری برای نتایج حاصل از بررسی باشد. طرح تک آزمودنی با استفاده از روشهای دیگر اعتبار کسب می‌کند. آزمایشگر میزان زمانی را که مراحل خط پایه و تیمار اثربخش هستند، کنترل می‌کند و زمان مرحله خط پایه را می‌توان تا زمانی که رفتار تثبیت شود، گسترش داد. به منظور دستیابی به تفسیری روشن و بدون ابهام، خط پایه باید حدوداً یکنواخت باشد و یا متمایل به سمت مخالف جهتی باشد که پس از ارائه تیمار انتظار می‌رود. اشکالی که به کنترل آزمایشگر در مورد طول مرحله تیمار وارد است، عبارت است از تمایل به ادامه تیمار تا زمانی که «اتفاقی بیفتد». چنانچه رفتار، بعد از مدت کوتاهی پس از شروع تیمار بروز نکند، احتمال دارد که عامل دیگر، که می‌تواند متغیر غیر آزمایشی هم باشد، دلیل تغییر مشاهده شده باشد.

طرحهای آزمایشی تک آزمودنی یکی از نقاط ضعف طرحهای گروهی را که همان تغییرپذیری بین آزمودنی است، حذف می‌کند. هر فردی به عنوان گواه خودش تلقی می‌شود و در نتیجه، قابل مقایسه بودن مسئله‌ای تلقی نمی‌شود. یکی از ابزار عمده کنترل، تکرار است که به ندرت در طرحهای گروهی امکان پذیر است. طرح ABAB تنها در یک مورد از تکرار در مورد همان آزمودنیها استفاده می‌کند، در حالی که طرح خط پایه چند تایی بیش از یک تیمار را تکرار می‌کند. تکرار طرح خط پایه چند تایی این احتمال این را که اثرات منتسب به تیمار در واقع معلول حوادث خارجی و یا متغیرهای مربوط به آزمودنی باشند، کاهش می‌دهد.<sup>(۱)</sup>

تحقیقات تک آزمودنی که بخوبی طرح ریزی شده باشند، معیارهای اعتبار داخلی دارند، لکن، مسئله اعتبار خارجی - تعمیم پذیری یافته‌های آزمایشی - با طرحهایی که تنها از یک یا چند آزمودنی استفاده می‌کنند، حاصل نمی‌شود. می‌توان نشان داد که چنانچه در مورد یک

1- Birnbauer, J., C. Peterson, and J. Solnick (1974). Design and Interpretation of studies of single subjects *American Journal of Mental Deficiency*, 79, 191 - 203

نوجوان دچار اختلال رفتاری، اجازه گوش دادن به موسیقی راک را مشروط به انجام تکالیف مدرسه‌اش کنیم، میزان تکلیف انجام شده توسط این نوجوان افزایش می‌یابد، ولی چگونه می‌توان به این سؤال پاسخ داد که آیا این تیمار در مورد دیگر نوجوانان و یا تمامی نوجوانان دچار اختلال رفتاری نیز مؤثر است یا خیر. هر چند هر بررسی تک آزمودنی از نظر اعتبار بیرونی ضعیف است، ولی انجام تعدادی بررسی مشابه که دقیقاً آزمودنیها، شرایط بررسی و تیمارها را توصیف کنند، می‌تواند زمینه را برای کاربرد وسیع اثرات خاص یک تیمار فراهم سازد. برای توضیحات بیشتر در زمینه طرحهای تک‌آزمودنی به هارسن و بارلو<sup>(۱)</sup> یا کازدین<sup>(۲)</sup> مراجعه کنید.

### /// خلاصه

آزمایش، سخت‌ترین و مطلوبترین شکل بررسی علمی است. شرایط کنترل شده که مشخصه آزمایش است، امکان شناخت روابط کارکردی اثبات شده بین پدیده‌های موردنظر دست‌اندرکاران تعلیم و تربیت را فراهم می‌سازد. آزمایشگرهایی که شرایط وقوع رویداد را کنترل می‌کنند، مسلماً بر مشاهده‌گرانی که تنها یک رویداد را مشاهده و بررسی می‌کنند و کنترلی بر آن ندارند، مزیت‌های روشنی دارند: ۱- آنها می‌توانند شرایط را به صورت نظام‌مند دستکاری و یا در آنها تغییر ایجاد کنند و تغییرات به وجود آمده در نتایج را در نظر گیرند. ۲- آنها می‌توانند ترتیبی دهند تا رویداد در زمانی اتفاق افتد که برای مشاهده و اندازه‌گیری دقیق آمادگی دارند و ۳- آنها می‌توانند به منظور اثبات صحت و سقم یافته‌ها، مشاهداتشان را تحت همان شرایط تکرار و کنترلی مستقلی بر روی نتایج اعمال کنند.

طرحهای واقعی آزمایشی بهترین کنترل را از نظر اعتبار داخلی فراهم می‌آورند. با این‌گونه طرحها، آزمودنیها به صورت تصادفی برای تیمارها تخصیص داده می‌شوند. در طرحهای شبه آزمایشی محقق می‌تواند بر روی تیمار کنترل داشته باشد، ولی نمی‌تواند آزمودنیها را به صورت تصادفی به تیمارها تخصیص دهد. طرحهای شبه آزمایشی، که اثر تیمار را بر روی یک آزمودنی بررسی می‌کنند، در تحقیقات رفتاری مفید هستند.

1- Hersen, M., and D. H. Barlow (1976). *Single - case experimental designs: Strategies for studying behavior*. New York: Pergamon.

2- Kazdin, A. E. (1982). *Single case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.

// مفاهیم اصلی

آزمایش تک آزمودنی	آزمایشهای آزمایشگاهی
آزمایشهای میدانی	آزمودنیهای به صورت تصادفی جفت شده
اثر آزمایشگر	اثر هائورن
استفاده از آزمودنیها برای کنترل خود	اعتبار بوم شناختی
اعتبار جامعه آماری	اعتبار خارجی
اعتبار خارجی عملیات	انتخاب انفرادی
انتخاب همگن	بلوغ به عنوان مشکل اعتبار داخلی
پس آزمون گروه گواه	پیش آزمون به عنوان مشکلی برای اعتبار داخلی
تحقیقات آزمایشی	تحقیق شبه آزمایشی
تحلیل کواریانس	تخصیص تصادفی (تصادفی کردن)
تعامل بین ویژگیهای آزمودنی و تیمار	تعامل در طرح عاملی
حساس سازی پیش آزمون	دستکاری متغیرهای مستقل
رگرسیون آماری به عنوان مشکل اعتبار داخلی	سابقه به عنوان مشکل اعتبار داخلی
طرح آزمایشی	طرح آزمودنیهای تصادفی، پیش آزمون - پس آزمون با گروه گواه
طرح آزمودنیهای همناشده تصادفی، تنها پس آزمون با گروه گواه	طرح توالی زمانی با گروه گواه
طرح توالی زمانی یک گروهی	طرح چهار گروهی سولومون
طرح گروه گواه، آزمودنیهای تصادفی، تنها پس آزمون	طرح متوازن
طرح ABAB	طرحهای آزمایشی واقعی
طرحهای عاملی	طرحهای ماقبل آزمایشی
قانون متغیر با اهمیت منفرد	کنترل تفاوت‌های وضعیتی
کنترل متغیرها	گروه آزمایشی
گروه گواه	گروه گواه غیر تصادفی
متغیرهای مداخله کننده	متغیرهای مخدوش کننده
مقایسه گروهی ایستا	نمونه گیری تصادفی
وسایل اندازه گیری به عنوان مشکل اعتبار داخلی	همتا کردن تصادفی
هم ارزی آماری اعتبار داخلی	

### /// تمرینها

۱- پژوهشگری از گروهی از دانش آموزان که برای درس علوم اجتماعی در دبیرستان نام نویسی کرده بودند، به صورت تصادفی ۶۰ دانش آموز را انتخاب کرد. در مرحله بعد، دانش آموزان به صورت تصادفی به دو گروه ۳۰ نفری A و B تخصیص یافتند. گروه A برنامه درسی سنتی علوم اجتماعی را دریافت کردند و گروه B از طریق برنامه جدید طراحی شده برای بررسی تاریخچه گروههای خاص نژادی، آموزش دیدند. در پایان نیم سال، دو گروه با مقیاسی که برای اندازه گیری نگرشها نسبت به گروههای نژادی طراحی شده بود، مقایسه شدند. در این طرح بگویید هر یک از موارد زیر کدامند؟

الف - متغیر مستقل

ب - متغیر وابسته

ج - گروه گواه

د - گروه آزمایشی

ه - روش یا روشهای استفاده شده برای کنترل تفاوتهای بین گروهها

و - طرح تحقیقی مورد استفاده

۲- مسئله تحقیقاتی زیر را در نظر بگیرید: آیا آموزش شاگردان زبان فرانسه به روش سمعی - گفتاری، به جای روش دستور زبان گشتاری باعث می شود که دانش آموزان در پایان سال عملکرد متفاوتی در یک آزمون استاندارد شده گرامر، خواندن و واژگان، نشان دهند؟  
الف - آزمایش مطلوب برای پاسخگویی به این سؤال را طراحی کنید. فرض کنید که هیچگونه محدودیت اداری و غیره وجود ندارد.

ب - آزمایشی را که به احتمال زیاد در وضعیت یک دبیرستان نوعی لازم است، طراحی کنید.

ج - مزایای طرح آزمایشی مطلوب الف را در مقایسه با طرح ب مشخص کنید.

۳- تفاوت بین نمونه گیری تصادفی و تخصیص تصادفی چیست؟ رابطه آنها با اعتبار درونی

و بیرونی چیست؟

۴- طرحهای تحقیقی زیر را از نظر روشهای مورد استفاده و کنترلی که فراهم می سازد،

ارزیابی کنید. در صورت لزوم، پیشنهادهايتان را برای بهبود روش ارائه دهید.

الف - محقق می خواهد تعیین کند که آیا گروه بندی همگن، یادگیری را در اولین واحد درس

زیست‌شناسی، بهبود می‌بخشد یا خیر. محقق یکی از دو دبیرستان موجود در یک شهر کوچک را به عنوان مدرسه آزمایشی و مدرسه دیگر را به عنوان مدرسه گواه انتخاب کرد. هر دو مدرسه در زیست‌شناسی، تقریباً تعداد مساوی دانش آموز داشتند. در مدرسه آزمایشی، دانش آموزان به طور همگن بر حسب هوشبهر و نمراتشان در آزمونهای پیشرفت تحصیلی در درس علوم گروه بندی شده بودند. در مدرسه گواه، دانش آموزان به صورت تصادفی در بخشهای چهارگانه جای داده شده بودند. در پایان سال تحصیلی به کلیه دانش آموزان، آزمون استاندارد شده زیست‌شناسی داده شد. نتایج آماری نشان دادند که گروه آزمایشی در این آزمون بر گروه دیگر برتری داشتند. محقق چنین نتیجه‌گیری کرد که گروه بندی همگن می‌تواند باعث یادگیری بهتر در درس زیست‌شناسی شود.

ب - یک معلم تاریخ به مسئله عدم آشنایی دانش آموزان با حکومت ایالت و کشورشان و نیز رویدادهای جاری، توجه داشت. او تصمیم گرفت که با مطالب و روشهای جدید آزمایش کند و ببیند که آیا می‌تواند پیشرفتی حاصل کند. در کلاسهای A و B مطالب و روشهای جدید را به کار برد و در کلاسهای C و D از روشهای سنتی استفاده کرد. به هر دو کلاسهای A و B پیش آزمون و پس آزمون داده شد، در حالی که به کلاسهای C و D تنها پس آزمون داده شد. مقایسه نمرات نشان داد که کلاسهای A و B برتری داشتند. عملکرد برتر آنها به ارائه مطالب و روشهای جدید نسبت داده شد.

۵- آزمایش ایده آل برای آزمون فرضیه های زیر را طرح کنید: کودکانی که فیلمهای با محتوای متعادل از نظر واکنش نژادی را می‌بینند، نسبت به کودکانی که فیلمهای ترسیم‌کننده تضادهای نژادی را می‌بینند، نگرش مثبت تری نسبت به اقلیتهای نژادی نشان خواهند داد.

۶- دوباره به مسئله تحقیق سؤال ۲ برگردید و فرض کنید که همچنین می‌خواهید بدانید که آیا در روش آموزش فرانسه در مورد دختران و پسران اثرات افتراقی دارد یا خیر. به طور خلاصه طرح آزمایشی را که امکان پاسخگویی به این سؤال را همزمان با سؤال قبلی فراهم سازد، شرح دهید.

۷- فرض کنید محقق از دو روش آموزش ( $A_1$  و  $A_2$ ) در مورد دو گروه دانش آموزان ( $B_1$  و  $B_2$ )، که سطح انگیزه پیشرفشان متفاوت است، استفاده کرده است. مقایسه گروهها از طریق یک آزمون پیشرفت در انتهای بررسی صورت گرفته است. میانگینها در زیر ارائه شده است. تفسیر شما از این نتایج چیست؟

$A_1$	$A_2$	
۱۵	۳۵	$B_1$
۱۵	۱۵	$B_2$

۸- شرط اعتبار نتایج در طرح تک آزمودنی چیست؟

/// پاسخها

- ۱- الف - نوع برنامه درسی علوم اجتماعی
- ب - نمرات در یک مقیاس سنجش نگرش نژادی
- ج - گروه A، برنامه درسی حاضر
- د - گروه B، برنامه درسی حاوی تاریخ نژادی
- ه - انتخاب تصادفی نمونه ها از جامعه آماری و تخصیص تصادفی نمونه به گروههای آزمایشی و گروه گواه

و - طرح ۳، طرح گروه گواه، آزمودنیهای تصادفی، تنها پس آزمون

۲- الف - از طرح ۳ استفاده کنید، یعنی دانش آموزان درس زبان فرانسه را به صورت تصادفی به دو گروه دستور زبان گشتاری (گروه گواه) و یا سمعی - گفتاری (گروه آزمایشی) تخصیص دهید. برای هر دو گروه شرایط زمان تدریس، معلمان و تسهیلات کلاسی یکسان را فراهم کنید و تنها، روش تدریس برای دو گروه متفاوت باشد. در پایان سال، آزمون اراثة و میزان پیشرفت دانش آموزان مقایسه شود.

ب - به صورت تصادفی دانش آموزان درس زبان فرانسه را در همان شرایط موجود کلاسی برای دو روش تدریس تخصیص دهید. هر معلم تعداد مساوی از دو نوع کلاس را تدریس خواهد کرد.

ج - در طرح ایده آل، خطرات مترتب بر اعتبار داخلی و خارجی از طریق انتخاب تصادفی فرد فرد دانش آموزان بهتر کنترل می شود. طرح سؤال ب می تواند با آزمودنیهایی که قبل از اراثة تیمار هم ارز نیستند، مشکل داشته باشد، زیرا تفاوتهای مشاهده شده در نمرات آزمون می تواند به عوامل دیگری بجز تفاوت تیمار مربوط باشد.

۳- در نمونه گیری تصادفی برای بیرون کشیدن یک نمونه از یک جامعه آماری از روشی استفاده می شود که همه افراد نمونه، بخت یکسان برای انتخاب شدن داشته باشند. از آنجا که این شیوه به این مسئله می پردازد که تا چه حد نتایج به دست آمده از یک نمونه را می توان به

جامعه‌ای که نمونه از آن انتخاب شده است، تعمیم داد. این روش یک راهبرد برای افزایش اعتبار بیرونی محسوب می‌شود.

۴- الف - از آنجا که محقق نمی‌تواند دانش آموزان را به صورت تصادفی به دبیرستانها اختصاص دهد، اعتبار درونی با چند خطر مواجه است. احتمال دارد دانش آموزان مدرسه آزمایشی باهوشتر بوده باشند و یا از پشتوانه علمی بیشتری نسبت به دانش آموزان مدرسه گواه برخوردار باشند. تفاوت‌های موجود در کیفیت تدریس بیولوژی در مدارس کنترل نشده است. از آنجا که محقق تنها مدارس یک شهر کوچک را انتخاب کرده است، نتایج بررسی را نمی‌توان به دبیرستانهای مناطق دیگر تعمیم داد.

محقق می‌توانست پیشرفت اولیه در درس علوم و هوشبهر را برای مدارس مقایسه کند تا ببیند آیا گروهها قبل از تیمار، همگن هستند یا خیر. استفاده از چند دبیرستان و انتخاب تصادفی کلاسهای آن دبیرستانها برای شرایط آزمایشی، می‌تواند عوامل خاص یک مدرسه فرضی را کنترل کند.

ب - کلاسهایی که به صورت تصادفی تخصیص نیافته باشند، ممکن است هم ارز نباشند. از پیش آزمون شاید می‌توانستیم برای تعیین هم ارزی استفاده کنیم، لکن پیش‌آزمون تنها به گروههای آزمایشی داده شده است. دادن پیش‌آزمون به گروههای آزمایشی به تنهایی می‌تواند آن گروهها را حساس سازد و در نتیجه بر روی تفاوت‌های مشاهده شده اثر گذارد. حتی در صورتی که نتوانیم فرد فرد دانش آموزان را به صورت تصادفی تخصیص دهیم، کلاسها باید به صورت تصادفی برای تیمارها تخصیص یابند.

۵- در طرح ایده‌آل، دانش آموزان به صورت تصادفی به گروهها تخصیص می‌یابند. از نتایج به دست آمده از یک اندازه‌گیری پس از تیمار در مورد نگرشها نسبت به اقلیتهای نژادی، می‌توان برای مقایسه گروههای آزمایشی و گواه استفاده کرد.

۶- مسئله به یک طرح عاملی با دو گروه پسر و دو گروه دختر نیاز دارد که به صورت تصادفی به شرایط آزمایشی و گواه تخصیص داده شوند.

۷- به نظر می‌رسد که تعاملی بین انگیزه پیشرفت و نوع تدریس وجود داشته باشد. روش  $A_1$  در مورد دانش آموزانی که انگیزه پیشرفت در سطح  $B_1$  داشتند، مؤثرتر است، در حالی که روش  $A_2$  برای دانش آموزان سطح  $B_2$  مناسبتر بوده است. معنی دار بودن این تعامل را می‌توان با استفاده از آزمون  $F$  تعیین کرد. انگیزش یا روش تدریس هیچ اثر کلی ندارند، زیرا میانگینها برای  $A_1$  و  $A_2$  و  $B_1$  و  $B_2$  یکسان هستند.

۸- در صورت ایجاد تغییر در تیمار، باید تغییر مشخصی نیز در رفتار بروز نماید.



## فصل دهم

### تحقیقات علی - مقایسه‌ای

#### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- تحقیقات علی - مقایسه‌ای را توصیف و آن را با تحقیقات آزمایشی مقایسه کند.
- ۲- شرایط لازم را برای استنتاج یک رابطه علی بیان کند.
- ۳- تبیین‌های جایگزین را برای تحقیقات علی - مقایسه‌ای توصیف و مواردی را که این تبیین‌ها قابل قبول هستند، یا نیستند، شناسایی کند.
- ۴- روشهای کنترل محدود (ناقص) را توصیف و مواردی را که این کنترل مفید است، شناسایی کند.
- ۵- مسائلی را که پاسخ به آنها از طریق انتخاب این شیوه انجام می‌شود، شناسایی کند.

در بررسی مسائل آموزشی نظیر اینکه چرا بعضی از کودکان متون نوشتاری را بهتر از دیگر کودکان می‌خوانند و چرا بعضی از نوجوانان بزهکار می‌شوند، در حالی که دیگران چنین سرنوشتی را پیدا نمی‌کنند، پی می‌بریم که تنها برخی از مسائل، از طریق تحقیقات آزمایشی قابل بررسی هستند. در صورتی که بخواهیم تأثیر متغیرهایی چون محیط‌خانه، انگیزش، هوش، عادات خواندن والدین و... را بررسی کنیم، نمی‌توانیم دانش آموزان را به صورت تصادفی به طبقات مختلف این متغیرها، تخصیص دهیم. این گونه متغیرهای مستقل، متغیرهای خصیصه‌ای خواننده می‌شوند. متغیر خصیصه‌ای آن ویژگی است که آزمودنی قبل از شروع بررسی دارد.

بالعکس، متغیر مستقل که محقق می‌تواند مستقیماً دستکاری کند، متغیر فعال است. محقق می‌تواند از قبل تعیین کند که کدامیک از دانش‌آموزان به آزمایشگاه کوچک دسترسی داشته و کدامیک نداشته باشند، کدام برای یادگیری درس جبر از برنامه A و کدام از برنامه B استفاده کنند. در صورتی که بررسی بر روی متغیرهای مستقل فعال مورد نظر باشد، می‌توان تحقیقات آزمایشی یا نیمه آزمایشی را به کار گرفت.

در صورتی که با متغیرهای خصیصه‌ای مستقل سر و کار داشته باشیم، طرحهای علی-مقایسه‌ای را به کار می‌گیریم (این گونه طرحها گاهی مواقع تحقیقات بعد از واقعه<sup>(۱)</sup> خواننده می‌شوند).

عنوان Ex Post Facto کلمه لاتینی به معنای «پس از واقعیت»<sup>(۲)</sup> است و برای نمایاندن این به کار می‌رود که تحقیق مورد نظر، بعد از آن که طی یک دوره طبیعی از رویدادها، تغییرات در متغیر مستقل مشخص شده‌اند، انجام می‌گیرد. در این گونه تحقیقات، محقق کنترل مستقیم بر متغیر مستقل ندارد؛ زیرا یا این متغیرها ماهیتاً قابل دستکاری نیستند و یا اینکه قبل از شروع تحقیق اتفاق افتاده‌اند.

محققان نه با دستکاری مستقیم متغیر، بلکه از طریق انتخاب افرادی که آن متغیر خاص در آنها وجود دارد یا ندارد، قوی است یا ضعیف و غیره، به تغییرات مورد نظر خود دست می‌یابند. آنها از کودکانی که دچار ضایعه مغزی شده‌اند و نیز کودکانی که دچار این ضایعه نشده‌اند، انجام یک عمل ادراکی واحد را می‌خواهند، یا عملکرد کودکان باهوش و کم هوش را در یک سطح مشخص اضطراب مقایسه می‌کنند.

### /// مقایسه روش تحقیق علی - مقایسه‌ای باروش آزمایشی

در هر دو نوع تحقیق، نکتهٔ مورد نظر، کشف و یا تبیین روابط بین متغیرها از طریق داده‌های حاصله است. تحقیقات بعد از واقعه نیز مانند تحقیقات آزمایشی می‌توانند فرضیه‌هایی را در مورد رابطهٔ بین یک متغیر مستقل  $X$  و متغیر وابسته  $Y$ ، آزمون کنند. در منطق پایه، روشهای آزمایشی و علی - مقایسه‌ای شبیه یکدیگرند. هدف هر دو روش مقایسهٔ دو گروه است که از نظر تمام ویژگیها به استثنای یکی مشابه هستند. این مقایسه به منظور اندازه‌گیری تأثیرات آن ویژگی انجام می‌شود. بنابراین، بسیاری از اطلاعاتی را که یک آزمایش فراهم می‌آورد، می‌توان از طریق تحلیل تحقیق علی - مقایسه‌ای به دست آورد.

با این همه، از طریق انجام یک آزمایش، در مقایسه با بررسیهای علی - مقایسه‌ای، می‌توان به شواهد قانع‌کننده‌تری درباره روابط علت و معلولی و کارکردی در بین متغیرها دست یافت. علت این امر، امکان کنترل تأثیر متغیرهای نامرتبط، به وسیلهٔ شرایط آزمایشی است. بعلاوه متغیر فرضی مستقل پیشابند برای تعیین اثر آن بر روی متغیر وابسته، مستقیماً دستکاری می‌شود. چنانچه مشاهده شود که در وضعیت کنترل شده،  $Y$  به صورت تابعی از تغییرات  $X$  تغییر می‌کند، محقق به شواهدی در جهت اعتبار فرضیهٔ رابطهٔ علت و معلولی بین  $X$  و  $Y$  رسیده است. از طرف دیگر، در یک تحقیق از نوع علی - مقایسه‌ای، محقق نمی‌تواند متغیر مستقل را به وسیلهٔ دستکاری و یا انتخاب تصادفی کنترل کند. در این گونه تحقیقات تغییرات متغیر مستقل قبلاً به وقوع پیوسته است، محقق با مسئلهٔ کوشش برای تعیین پیشابندهای پیامدهای مشاهده شده روبرو می‌شود. به علت عدم کنترل در یک طرح از نوع علی - مقایسه‌ای، نتیجه‌گیری در مورد وجود یک رابطهٔ اصیل بین  $X$  و  $Y$  مخاطره‌آمیز است.

در اینجا با بررسی هر دو روش در مورد یک مسئلهٔ تحقیقاتی واحد، تفاوت بین دو روش آزمایشی و علی - مقایسه‌ای را نشان می‌دهیم. فرض کنیم بررسی اثرات اضطراب دانش‌آموزان در یک آزمون پیشرفت تحصیلی بر روی نتایج امتحان آنها، مورد نظر باشد. رویکرد علی - مقایسه‌ای در مورد این مسئله این است که سطح اضطراب را که در موقعیت امتحان وجود دارد، اندازه‌گیری و سپس عملکرد دانش‌آموزانی را که از نظر میزان اضطراب در «سطح بالا» و «سطح پایین» بودند، مقایسه کند. نقطه ضعف این رویکرد ناشی از این واقعیت است که نمی‌توان نتیجه‌گیری کرد که تفاوتی که در عملکرد آزمون پیشرفت تحصیلی مشاهده می‌شود، لزوماً ناشی از

اضطراب دانش آموز است. هر دو مجموعه نمره‌ها می‌توانند از عامل سومی مانند هوش عمومی تأثیر پذیر باشند. هوش عمومی می‌تواند عامل اصلی در سطح اضطراب و همچنین نتایج آزمون پیشرفت تحصیلی باشد.

در مورد همین مسئله، رویکرد آزمایشی بدین صورت عمل می‌کند که آزمایش را در دو موقعیت، که از هر نظر به استثنای جنبه برانگیزنده یکسان هستند، اجرا می‌کند. در یکی از این دو، موقعیت اضطراب برانگیز و در دیگری خنثی است. آزمایشگر می‌تواند با گفتن مطالبی نظیر: نمره نهایی آنها بستگی به عملکرد این آزمون دارد، آزمون بشدت مشکل است و یا از نتیجه این آزمون برای شناختن افراد ناتوان استفاده خواهد شد، در آزمودنیها اضطراب ایجاد کند. محقق می‌تواند آزمودنیها را به طور تصادفی به دو گروه تخصیص دهد. در این مورد چنانچه گروه مضطرب بهتر از گروه خنثی عمل کنند، می‌توان نتیجه گرفت که اضطراب، اثری تسهیل کننده بر روی عملکرد آزمون دارد. چنین نتیجه گیری به لحاظ کنترل حاصل از تخصیص تصادفی گروهها به تیمارها و دستکاری مستقیم متغیر مستقل توسط آزمایشگر می‌تواند عقلانی باشد. اضطراب، یکی از نادر متغیرهایی است که می‌تواند هم یک متغیر مستقل فعال و هم مستقل خصیصه‌ای باشد. بدین معنا که می‌توان آن را به صورت فعالانه، چنانچه در قبل توصیف شد (رویکرد آزمایشی)، دستکاری کرد و یا آزمودنیها را بر مبنای نمره آنها در یک آزمون اندازه گیری میزان اضطراب طبقه بندی کرد (رویکرد علی - مقایسه‌ای).

به عبارتی، می‌توان به مطالعه علی - مقایسه‌ای به عنوان رویکردی عکس رویکرد آزمایشی نگریست. در تحقیقات علی - مقایسه‌ای، به جای آنکه گروههای هم ارز را در معرض تیمارهای متفاوت قرار دهند، با بررسی گروههای متفاوت شروع و پیشایندهای این تفاوتها را تعیین می‌کنند. تحقیق علی - مقایسه‌ای با توصیف موقعیت موجود شروع می‌شود و فرض می‌کند این موقعیت در اثر پاره‌ای عوامل فعال قبلی بوجود آمده باشد و تلاش می‌کند با جستجو در رویدادهای گذشته، عوامل پیشاینده فرضی را، که عمل آن در زمان قبل شروع شده است، تعیین کند.

چنین روشی کنترلی را که ویژه روش آزمایشی است و لازمه استنباط یک رابطه علت و معلولی است، فراهم نمی‌آورد. چنانچه محقق در یک مطالعه از نوع علی - مقایسه‌ای، رابطه‌ای بین متغیرها پیدا کند، تنها شواهدی را مبتنی بر تغییرات همراه به دست آورده است. از آنجا که

محقق، X و یا متغیر احتمالی دیگری را که ممکن است تعیین کننده Y باشد، کنترل نکرده است، مبنایی برای استنباط یک رابطه علت و معلولی بین X و Y ندارد. محقق برای آنکه بتواند نوعی رابطه پیشاینده - پیامد بین متغیرها برقرار سازد، باید شواهدی گردآوری کند تا نشان دهد که Y مقدم بر X نیست، یا اثر Y به وسیله عامل دیگری که به عامل پیشاینده فرضی مرتبط است، ایجاد نشده است. برای مثال، چنانچه یک آزمون ادراک را به دو گروه از کودکان با ضربه مغزی و بدون ضربه مغزی بدهیم، تفاوت عملکرد می‌تواند منعکس کننده اثرات ضربه مغزی و یا عوامل دیگری مانند تفاوت‌های میزان اضطراب ناشی از برخی انواع بیماری باشد.

حال به گزارش قدیمی تری از راجرسن و راجرسن<sup>(۱)</sup> می‌پردازیم. در این گزارش آمده است که یک گروه کودک که با شیر مادر تغذیه شده بودند، در دوران دبستان موفقیت تحصیلی بیشتری از کودکانی که با شیرخشک تغذیه شده بودند، نشان دادند. از چنین یافته‌هایی نمی‌توان نتیجه‌گیری کرد که تغذیه با شیر مادر می‌تواند عامل بالا رفتن میزان موفقیت در مدرسه باشد. به احتمال زیاد، رابطه مشاهده شده توسط محققان یاد شده ناشی از تغییراتی در یک یا چند متغیر است که هم بر نوع تغذیه کودکان و هم بر سطح عملکرد بعدی آنان در مدرسه، تاثیر گذاشته است.

مؤلفان گزارش می‌کنند که بررسی در کلینیکی انجام شده است که در آن مادران به تغذیه کودک با شیر مادر تشویق می‌شدند و عدم موفقیت در این امر، نتیجه وضعیت جسمانی ضعیف نوزاد یا مادر یا هر دو بوده است. بنابر این، منطقی‌ترین توجیه می‌تواند به این صورت باشد که هم نوع تغذیه کودک و هم عملکرد بعدی او در مدرسه، می‌تواند ناشی از تندرستی باشد. تندرستی بیشتر، منجر به تغذیه با شیر مادر و عملکرد بهتر در مدرسه شده است. در نتیجه، به رغم رابطه مشاهده شده بین دو عامل، مطمئناً نمی‌توانیم نتیجه‌گیری کنیم که سطح عملکرد در مدرسه به صورت مستقیم ناشی از نوع تغذیه است.

با وجود آنکه، تحقیق علی - مقایسه‌ای، جانشین کاملی برای تحقیق آزمایشی نیست، اما روشی را در اختیار ما می‌نهد که می‌توان از آن در اوضاع و شرایطی استفاده کرد که اکثر تحقیقات آموزشی باید تحت آن شرایط انجام شوند. این نوع تحقیق روش مفیدی است که اطلاعات ارزشمندی را در زمینه تصمیم‌گیری آموزشی فراهم می‌آورد.

### /// شرایط لازم برای استنباط روابط علی

در صورتی که بخواهیم نتیجه گیری کنیم که متغیر  $X$  علت متغیر  $Y$  است، وجود سه نوع شواهد ضروری است:

۱- رابطه آماری بین  $X$  و  $Y$  وجود داشته باشد.

۲-  $X$  از نظر زمانی جلوتر از  $Y$  باشد.

۳- عوامل دیگر تعیین کننده  $Y$  نباشند.

طرحهای آزمایشی به دلیل دارا بودن کنترلها (عوامل حفاظتی) پاسخگویی تمام این ضرورتها هستند و بنابر این، نتیجه گیری علی در آنها امکان پذیر است؛ در حالی که در طرحهای علی - مقایسه‌ای که فاقد کنترل آزمایشی هستند، تفسیر رابطه علی به مراتب مخاطره آمیزتر است.

چنانچه محقق در یک بررسی علی - مقایسه‌ای، بین دو متغیر رابطه برقرار کرده باشد، باید برای یافتن شواهدی دال بر دو مورد دیگر، کار را ادامه دهد. محقق باید ترتیب زمانی را در نظر بگیرد، بدین معنا که باید ببیند آیا احتمال دارد  $Y$  قبل از  $X$  اتفاق افتاده باشد، و از اینرو، معلول  $X$  نباشد. تصمیم گیری در مورد رابطه زمانی بین  $X$  و  $Y$  به دو صورت امکان پذیر است. این امر یا از طریق منطقی صورت می‌گیرد و یا به عنوان نتیجه اندازه گیریهایی که نشان می‌دهند گروهها قبل از قرار گرفتن در معرض  $X$ ، از نظر  $Y$  یکسان بوده‌اند.

همچنین، توجه محقق به این امر که آیا عوامل دیگری بجز  $X$ ، عامل  $Y$  بوده‌اند یا خیر، از اهمیت فوق العاده‌ای برخوردار است. محقق برای بررسی این احتمال، متغیرهای مربوط دیگر را در تحلیل خود دخالت می‌دهد و نحوه تأثیر این متغیرهای اضافی را بر رابطه بین  $X$  و  $Y$  مشاهده می‌کند.

امکان دارد محقق به این نتیجه برسد که حتی با وجود وارد کردن متغیرهای دیگر، ارتباط بین  $X$  و  $Y$  به جای خود باقی می‌ماند. در این صورت برای تأیید نتیجه گیری علی، شواهدی وجود دارد. از طرف دیگر، محقق ممکن است دریابد که رابطه بین  $X$  و  $Y$ ، با حضور متغیرهای دیگر تغییر کرده است و یا حتی از بین می‌رود. در این صورت چنین نتیجه گیری می‌کند که  $X$  تعیین کننده  $Y$  نیست یا رابطه بین  $X$  و  $Y$  واقعی نیست.

### // احتمالات نتایج کاذب در یک تحقیق علی - مقایسه‌ای

تفاوت بین یک متغیر مستقل فعال و یک متغیر مستقل خصیصه‌ای، بسیار حایز اهمیت

است. در صورتی که محققان بتوانند تیمار X را کنترل و سپس متغیر وابسته Y را مشاهده کنند، دلایل کافی برای اینکه X بر Y اثر گذارده است، در دست دارند. در صورتی که نتوانند X را کنترل کنند، ممکن است به نتیجه‌گیری نامناسبی برسند؛ زیرا رابطه مشاهده شده ممکن است یک رابطه کاذب باشد، یعنی رابطه مشاهده شده ناشی از دیگر عوامل و نه اثر X بر Y باشد. از بین منشأهای احتمالی رابطه‌های کاذب می‌توان از عامل مشترک، علیت معکوس و وجود متغیرهای مستقل دیگر نام برد.

### / عامل مشترک

در یک بررسی از نوع علی - مقایسه‌ای، باید این احتمال را که متغیر مستقل و وابسته بررسی، صرفاً نتایج مجزای متغیر سومی هستند، در نظر داشت. برای مثال، چنانچه میانگین حقوق معلمان به عنوان متغیر مستقل و میزان فروش سالیانه نوشابه‌های الکلی از زمان لغو قانون منع فروش در ایالات متحده را متغیر وابسته تلقی کنیم، همبستگی مثبت بالایی بین دو متغیر به دست می‌آوریم. آیا این نتیجه بدین معناست که هرگاه حقوق معلمان افزایش یافته است، آنها پولشان را صرف خرید الکل کرده‌اند؟ تبیین معقول تر این یافته‌ها به این صورت است که هم میزان حقوق معلمان و هم مقدار فروش نوشابه‌های الکلی نتیجه تورم و رفاه فزاینده بعد از سال ۱۹۳۳ است.

چنین ثابت شده است که درآمد متوسط فارغ التحصیلان دبیرستانهای خصوصی بالاتر از فارغ التحصیلان دبیرستانهای عمومی و یا دبیرستانهای وابسته به کلیسا است. آیا این بدان معناست که دبیرستانهای خصوصی دانش آموزان را برای موفقیت اقتصادی، بهتر تربیت می‌کنند؟ یا تفاوت موجود ناشی از این واقعیت است که خانواده‌هایی که برای فرستادن فرزندانشان به مدارس خصوصی، به اندازه کافی پول دارند، توانایی فراهم کردن آموزش حرفه‌ای برای فرزندانشان، وارد کردن آنها به تجارت و یا خرید کرسی نمایندگی سنای ایالات متحده را نیز دارند؟

در شهر X مشاهده کردیم که طی بیست سال گذشته، افزایش مصرف برق با افزایش موارد بیماری روانی همراه بوده است. آیا این بدان معناست که افزایش استفاده از نیروی برق به افزایش بیماریهای روانی، منجر می‌شود؟ کنترل ارقام سرشماری نشان می‌دهد که جمعیت شهر X طی

این سالیان افزایش یافته است و افزایش مصرف برق و موارد حاکی از بیماری روانی، هر دو تابعی از رشد جمعیت هستند.

محققی که روش علی - مقایسه‌ای را به کار می‌برد، باید همواره احتمال وجود علت یا علل مشترک یک رابطه مشاهده شده را در نظر گیرد. در مثالهای ما علل مشترک تقریباً واضحی تشخیص داده می‌شود، در حالی که در تحقیقات بعد از واقعه، همواره شک آزردهنده‌ای وجود دارد که شاید برای توجیه روابط، عوامل مشترکی وجود داشته باشند که به ذهن کسی خطور نکرده‌اند. آمار نشان داده است که میزان صدمات وارده به رانندگانی که از کمربند ایمنی استفاده می‌کنند، کمتر از رانندگانی است که از آن استفاده نمی‌کنند. آیا این امر بدان دلیل است که استفاده از کمربند ایمنی میزان صدمات را کاهش می‌دهد یا آنکه رانندگان محتاط، (الف) از کمربند ایمنی استفاده می‌کنند و (ب) تصادفهای منجر به آسیبهای بدنی کمتری برایشان پیش می‌آید؟

### علت معکوس /

محقق در تفسیر رابطه مشاهده شده یک بررسی علی - مقایسه‌ای، باید این احتمال را نیز در نظر بگیرد که عکس فرضیه پیشنهادی هم می‌تواند موجد آن رابطه باشد، یعنی به جای این که بگوییم  $X$  علت  $Y$  است،  $Y$  علت  $X$  باشد. برای مثال، این یک واقعیت است که نسبت پروان کلیسای اسقفی که نامشان در کتاب «اشخاص سرشناس در آمریکا»<sup>(۱)</sup> فهرست شده است، به مراتب بیشتر از نسبت این گروه در کل جامعه است. آیا این بدان معناست که پیروی از کلیسای اسقفی منجر به نوعی موفقیت می‌شود که در نهایت نام این افراد در کتاب اشخاص سرشناس درج می‌شود؟ به همین اندازه و یا حتی بیشتر معقول است که چنین فرضیه‌ای را طرح کنیم که «افراد موفق، تمایل دارند به کلیسای اسقفی بپیوندند».

چنانچه یافته‌های ما نشان دهد که میانگین نمرات دانش آموزانی که توشابه‌های الکلی می‌نوشند، نسبت به دیگر دانش آموزان، پایین تر است، نمی‌توانیم بلافاصله نتیجه‌گیری کنیم که مصرف الکل از موفقیت تحصیلی می‌کاهد. شاید نمرات بد، باعث روی آوردن دانش آموزان به الکل باشد (یا البته، تعدادی عامل مشترک باشند که هم عامل نوشیدن الکل و هم عامل کسب نمرات ضعیف تر باشند).



تحقیقات انجام شده بر روی روشهای تربیتی کودک نشان داده‌اند که کودکانی که بیشتر تنبیه می‌شوند، رفتار پرخاشگرانه بیشتری بروز می‌دهند. آیا به این ترتیب کسی می‌تواند نتیجه‌گیری کند که تنبیه والدین باعث پرخاشگر شدن کودکان می‌شود یا آنکه کودکان پرخاشگر احتمالاً بیشتر تنبیه می‌شوند؟

پرداختن به فرضیه علیت معکوس آسانتر از فرضیه عامل مشترک است. با فرضیه عامل مشترک این احتمال وجود دارد که تعدادی عوامل مشترک وجود داشته باشند که بتوانند رابطه کاذب ایجاد کنند. با فرضیه علیت معکوس تنها یک احتمال وجود دارد: به جای آنکه  $X$  عامل  $Y$  باشد،  $Y$  عامل  $X$  است.

در هر صورت چنانچه  $X$  همواره از نظر زمانی جلوتر از  $Y$  باشد، ماهیت داده‌های ما به خودی خود، احتمال علیت معکوس را از بین می‌برد. برای مثال، نتایج تعداد زیادی بررسی نشان داده است که متوسط درآمد سالیانه فارغ التحصیلان دانشگاه بالاتر از درآمد سالیانه کسانی است که به دانشگاه نرفته‌اند. می‌توانیم فرضیه علیت معکوس را حذف کنیم، زیرا فارغ التحصیل دانشگاه بودن یا نبودن، مقدم بر درآمد سالیانه بعدی است، لکن نمی‌توانیم عامل مشترک احتمالی مختلف را حذف کنیم.

یک روش جهت تعیین ترتیب زمانی متغیرها عبارت است از اندازه‌گیری همان آزمودنیها در فاصله‌های زمانی مختلف. فرض کنیم محقق می‌خواهد رابطه بین پذیرش فلسفه سازمانی و ترفیع شغلی را در آن سازمان بررسی کند. چنانچه محقق صرفاً نمونه‌ای از کارمندان را انتخاب و با آنها مصاحبه کند و این نتیجه را به دست آورد که کسانی که در موقعیتهای بالاتر سازمانی قرار دارند، نگرشها و دیدگاههایشان با سیستم ارزشی سازمان هماهنگ است، نخواهد دانست که پذیرش ارزشها و هدفهای سازمانی موجب ترفیع شغلی شده است، یا کسب موقعیت بالاتر منجر به پذیرش بیشتر سیستم ارزشی سازمان گردیده است. محقق برای حذف رابطه علت و معلولی معکوس، به عنوان یک تبیین، می‌تواند با گروهی از کارمندان جدید مصاحبه کند و به وسیله یک پرسشنامه با مقیاس درجه بندی و یا چیزهایی نظیر آن، اندازه‌ای از «پذیرش فلسفی سازمانی» را به دست آورد. سپس، بعد از یک دوره زمانی، شاید ۱۸ ماه، محقق می‌تواند از طریق پرونده‌های سازمان بفهمد که کدام یک از آنها ترفیع گرفته‌اند. چنانچه نتایج چنین نشان دهد که نسبت بالایی از کسانی که نگرشها و عقاید هماهنگی با فلسفه سازمان را ابراز کرده‌اند، در شغل

خود ترقی کرده‌اند، محقق می‌تواند نتیجه‌گیری کند که هماهنگی با فلسفه سازمان عامل ترقی بوده است (در هر حال هنوز احتمال وجود برخی عوامل مشترک وجود دارد که دلیل بروز تفاوت هم در فلسفه سازمانی و هم در ترفیع شغلی می‌باشند).

### دیگر متغیرهای مستقل احتمالی

امکان دارد که در یک تحقیق علی-مقایسه‌ای، متغیر مستقل دیگری بجز متغیر مورد بررسی وجود داشته باشد که می‌توانسته است اثر مشاهده شده بر روی متغیر  $Y$  را به وجود آورده باشد؛ بدین معنی که علاوه بر  $X_1$ ، متغیرهای  $X_2$  و  $X_3$  می‌توانند عوامل ایجادکننده تفاوتها در متغیر وابسته باشند.

می‌دانیم که تعداد موارد خودکشی ثبت شده در سوئد، جزء بالاترین آمارهای خودکشی در جهان است. آیا این بدان معناست که محیط سوئد مردم را به خودکشی وامی‌دارد؟ آیا این بدان معناست که سوئدی‌ها برای خودکشی مستعدتر هستند؟ شاید یکی یا هر دو این فرضیه‌ها صحیح باشند، لکن به همین میزان احتمال دارد که متغیر مستقل واقعی میزان صداقت پزشکی قانونی سوئد باشد و پزشکان قانونی سوئدی، نسبت به پزشکان قانونی دیگر کشورها، از صداقت بیشتری برخوردار باشند. در کشورهایی که به خانواده‌هایی که در آنها خودکشی اتفاق می‌افتد، مهر بدنامی زده می‌شود، پزشک قانونی با استفاده از تدابیر ممکن، علت مرگ را به جای خودکشی، تصادف اعلام می‌کند. بنابر این، تفاوت در آمار خودکشیهای گزارش شده می‌تواند معلول طرز برخورد پزشکی قانونی باشد و بس.

در کنفرانسی با شرکت فرمانداران، فرماندار  $X$  با غرور، پایین بودن آمار جنایت را در ایالت محل خدمتش اعلام می‌کند. فرماندار دیگر اظهار می‌کند که پلیس ایالت محل خدمت فرماندار  $X$  شدیداً دچار کمبود نیروی انسانی است و میزان پایین بودن جنایات تنها به این دلیل است که اصولاً تعداد کمی از جنایات گزارش می‌شوند. کارخانه داری از مدیر امور اداری کارخانه‌اش سؤال می‌کند که چرا فارغ التحصیلان دانشگاه Old Siwash را به رغم این واقعیت که به دلیل با کفایت بودنشان نردبان ترقی را به سرعت می‌پیمایند، استخدام نمی‌کند. مدیر امور اداری با درایت پاسخ می‌دهد که دلیل این پدیده، کفایت این افراد نیست، بلکه چون خود کارخانه‌دار فارغ التحصیل دانشگاه نامبرده است، ناخودآگاه نسبت به فارغ التحصیلان آن نظر مثبت دارد و این نظر خود را در ترفیعات اعمال می‌کند.

اولین وظیفه مسلم یک محقق این است که کوشش کند فهرستی از تمامی متغیرهای مستقل احتمالی دیگر تهیه کند. سپس با ثابت نگاه داشتن بقیه، می‌تواند هر یک از متغیرها را مورد آزمون قرار دهد تا مشخص شود که آیا با  $Y$  رابطه دارند یا خیر. چنانچه بتوانیم متغیرهای مستقل جایگزین را با نشان دادن اینکه رابطه‌ای با  $Y$  ندارند، حذف کنیم، می‌توانیم شواهدی برای تأیید فرضیه اصلی رابطه بین  $X$  و  $Y$  به دست آوریم.

داده‌های گزارش زیر رابطه‌ای بین استفاده از اتومبیل و میزان موفقیت تحصیلی نشان می‌دهد.

### آیا اتومبیل و کسب دانش با هم همخوانی دارند؟

دبیرستان مدیسون<sup>(۱)</sup>، از شهر رکس برگ، آیداهو،<sup>(۲)</sup> گواهی بر منفی بودن این پاسخ است. بررسی انجام شده بر روی معدل نمرات چهار ساله کلاس آخر دبیرستان مدیسون و دارندگان اتومبیل این دبیرستان، اطلاعات زیر را ارائه می‌دهد:

هیچیک از دانش آموزانی که میانگین آنها  $A$  تمام بود، از اتومبیل استفاده نمی‌کردند.

تنها ۱۵٪ از دانش آموزان با میانگین  $B$ ، با اتومبیل به دبیرستان می‌آمدند.

۴۱٪ دانش آموزان با میانگین  $C$  اتومبیل به مدرسه می‌آوردند.

۷۱٪ دانش آموزان با میانگین  $D$ ، با اتومبیل به دبیرستان می‌آمدند.

۸۳٪ دانش آموزان با میانگین  $E$ ، با اتومبیل به دبیرستان می‌آمدند.

### برنامه عملی

بر مبنای نتایج فوق، هیئت امنای رکس برگ راه حلی را در ۱۱ اوت ۱۹۵۸ برگزید که مشخص می‌ساخت:

۱- دانش آموزان کلاس یازدهم و دوازدهم، که با اتومبیل به مدرسه می‌آمدند، باید طی پرستنامه‌ای به هیئت امنای مدرسه توضیح دهند که چرا و به چه دلیلی نیاز دارند که اتومبیلشان را به مدرسه آورند. پس از بررسی، هیئت امنا به گروهی اجازه می‌دهد که اتومبیل به مدرسه آورند.

۲- دانش آموزانی که با کسب اجازه، اتومبیل به مدرسه می آورند، باید اتومبیلشان را در محل تعیین شده پارک کنند.

۳- مجوز دانش آموزانی که از قوانین فوق سرپیچی کنند، باطل می شود و در صورت تکرار، توسط هیئت امنای مدرسه، اخراج می شوند.

### همکاری جامعه

کل جامعه جهت پشتیبانی از تصمیم هیئت امنای بسیج شدند. شورای حمایت از قانون و مجمع جوانان رکس برگ پشتیبانی کلیه سازمانها و کلوپها و نهادهای جامعه چون کلوپ لاتاری، روزنامه‌ها و پلیس را در این امر اعلام کردند. بسیاری از شهروندان سوگندنامه‌ای را برای اجرای طرح امضا کردند و کارت عضویتشان را از شورا و مجمع دریافت کردند. مدیر دبیرستان مدیسون، و. ج. نلسون<sup>(۱)</sup> قویاً تصویبنامه هیئت امنای را در زمینه جلوگیری از استفاده نامحدود از اتومبیل توسط دانش آموزان دبیرستانی، تأیید و چنین اظهار کرد: «ما بر این باوریم که محدودیت استفاده از اتومبیل باعث می شود که دانش آموزان با دقت و توجه بیشتری به درس پردازند. هنوز به این موفقیت دست نیافته‌ایم، ولی مسلماً در راه آن قدم بر می داریم.»

### اظهارات یک مدیر مدرسه در تگزاس

یک مدیر مدرسه در تگزاس، در اظهار نظر نسبت به برنامه رکس برگ چنین می گوید: «همیشه اعتقاد داشتم که همبستگی بالایی بین میزان رد شدن، کسب دانش و مالکیت و رانندگی نامحدود اتومبیل توسط دانش آموزان دبیرستانی وجود دارد. ما در جلب علاقه و شوق پسرانی که از نظر بدنی آمادگی ورزش را دارند، برای کار کردن در رشته های ورزشی، همواره مشکل داشته‌ایم. بسیاری از پسران مجبورند برای خرید اتومبیل و یا تأمین مخارج تعمیرات آن بعد از مدرسه به صورت نیمه وقت کار کنند و در نتیجه به دنبال ورزش نمی روند و چنانچه برای تعمیرات آن پول کافی داشته باشند، ترجیح می دهند بعد از مدرسه این طرف و آن طرف شهر رانندگی کنند. تعداد بسیار کمی از والدین هستند که کنترل و تسلط کافی بر فرزندانشان دارند تا

آنها را از استفاده غیر معقول از اتومبیل باز دارند، ولی تعداد آنها بسیار قلیل است.»<sup>(۱)</sup>  
 آیا این نتیجه‌گیری که استفاده از اتومبیل عامل پایین آمدن میزان موفقیت تحصیلی است، قابل قبول است؟ حال به فرضیه‌های احتمالی دیگر بنگریم.

۱- عامل مشترک - آیا متغیرهای دیگری وجود دارند که هم استفاده از اتومبیل و هم کسب دانش را تحت تأثیر قرار دهند؟ می‌دانیم که طبقه اجتماعی با دانش اندوزی ارتباط دارد. چنانچه طبقه اجتماعی بر استفاده از اتومبیل نیز اثر داشته باشد، دیگر رابطه آشکار بین معدل درسی و استفاده از اتومبیل، رابطه‌ای از نوع علت و معلولی نخواهد بود، بلکه دو جنبه از تفاوت‌های طبقات اجتماعی محسوب می‌شود. تفاوت‌های موجود در شیوه زندگی و ارزشهای دانش آموزان نیز می‌توانند در این رابطه آشکار، تأثیر داشته باشند. چنانچه برای دانش آموزانی رانندگی اتومبیل از ارزش بالایی برخوردار باشد، در حالی که علاقه کمی به دانش اندوزی داشته باشند، دیگر باز داشتن آنها از استفاده از اتومبیل لزوماً باعث افزایش علاقه آنها به دانش نمی‌شود. می‌توانیم تعدادی از فرضیه‌های معتبر عامل مشترک را ارائه دهیم.

۲- علیت معکوس - آیا این احتمال هست که گرفتن نمره بد، عامل روی آوردن به استفاده از اتومبیل باشد؟ منطقاً می‌توانیم چنین فرضیه‌ای تبیین کنیم که عملکرد ضعیف در مدرسه باعث می‌شود تا دانش آموز برای پذیرفته شدن در اجتماع به دنبال راههای دیگری باشد و شاید استفاده از اتومبیل یکی از این راهها باشد.

۳- دیگر متغیرهای مستقل احتمالی - آیا احتمال این هست که معلمان، دانش آموزانی را که از اتومبیل استفاده می‌کنند، به عنوان دانش آموزان بی‌علاقه تلقی کنند و در نتیجه نمره‌ای پایین‌تر از آنچه سزاوارند، به آنها بدهند؟

بسیاری فرضیه‌های متغیر جایگزین دیگر وجود دارند که نشان می‌دهند باید از تفسیر داده‌ها به صورت یک رابطه علت و معلولی بپرهیزیم. آشر و شوستر<sup>(۲)</sup> همین مسئله را با بهره‌گیری از طرح تحقیق دیگری که کنترل بیشتری را نسبت به تحقیق قبلی اعمال می‌کرد، انجام دادند. به جای آنکه تنها معدل کنونی استفاده‌کنندگان از اتومبیل را در نظر گیرند، نمرات کلاس دوازدهم

1- National Association of Secondary School Principals (1959). Do automobiles and scholarship mix? *NASSP Spotlight on Junior and Senior and High Schools*, 36, 3.

2- Asher, W., and M. M. Schuster (1967). Students' grades and access to cars. *Journal of Educational Research*, 60, 10.

آزمودنیها را پس از پایان نیمسال اول و همچنین نمرات همان آزمودنیها را هنگامی که دانش آموز کلاس دهم بودند، ثبت کردند. چنانچه استفاده از اتومبیل بر موفقیت تحصیلی اثر گذارد، انتظار می رود که نمرات دانش آموزان نسبت به کلاس دهم، که گواهینامه رانندگی نداشتند، افت کرده باشد. آشر و شوستر معدل نمرات کلاس دهم و دوازدهم را به عنوان متغیر وابسته تحقیقاتشان تلقی نمودند.

آنها همچنین هوشبهر را به عنوان متغیری که در هر دو مورد متغیر کمکی است، در نظر گرفتند (به فصل ۹ مراجعه شود). آنها هیچگونه رابطه معنی داری بین استفاده از اتومبیل و تغییر در معدل نیافتند و چنین نتیجه گیری کردند که «به نظر نمی رسد دسترسی به اتومبیل عامل ظهور الگویی رایج از کاهش و یا افزایش نمرات باشد».

### /// کنترل محدود در تحقیق علی - مقایسه ای

راهبردهایی برای افزایش اعتبار تحقیق علی - مقایسه ای وجود دارد، هرچند هیچیک از این راهبردها نمی توانند نقیصه ذاتی این روش تحقیق، یعنی عدم کنترل بر متغیر مستقل را به اندازه کافی جبران کنند. بین این راهبردها می توان به نمره های تغییر، همتاسازی، تحلیل کواریانس، همبستگی جزئی، گروه های همگن و وارد کردن متغیرهای خارجی در طرح اشاره کرد.

### // نمره های تغییر

در مثال استفاده دانش آموزان از اتومبیل، دریافتیم که نتیجه حاصل وقتی که تفاوت بین نمره های کلاس دوازدهم و کلاس دهم در نظر گرفته شد، با نتیجه ای که فقط حاصل نمره های کلاس دوازدهم بود، با هم فرق می کرد. مسلماً نتیجه اول نسبت به نتیجه دوم از اعتبار بیشتری برخوردار است، زیرا نتیجه دوم به طور کلی این احتمال را نادیده می گیرد که اصولاً پیش از آنکه این دو گروه به اتومبیل دسترسی پیدا کنند، استفاده کنندگان از اتومبیل دانش آموزان ضعیف تری بوده اند.

در نمره های تغییر، محقق نمرات قبلی  $Y$  را نیز در نظر می گیرد و صرفاً به نمرات حاضر  $Y$  نمی پردازد. اگر محقق میزان تغییرات گروه تیمار را با گروه کنترل مقایسه کند، احتمالاً کمتر از آن هنگام که صرفاً به نمرات فعلی می پردازد، دچار اشتباه می شود. لکن، استفاده از نمره های تغییر

تنها تا حدودی مسئله را حل می‌کند و نتایج چنین بررسی‌هایی باید محتاطانه به کار رود. برای مثال، مدیر مدرسه A، شیوه جدید تدریس قرائت را برای کلاس پنجم مدرسه‌اش به کار می‌برد و می‌خواهد پس از استفاده از این شیوه برای مدت یک سال، اثر آن را با اثر روش قبلی مقایسه کند. در همان منطقه، در مدرسه B از روش قدیمی استفاده می‌شود. در پایان سال تحصیلی، در هر دو مدرسه یک آزمون استاندارد شده قرائت داده می‌شود. مدیر مدرسه A میانگین نمرات قرائت کلاس پنجمی‌های دو مدرسه را مقایسه می‌کند. او می‌بیند که میانگین نمره کلاس پنجم او ۶ است، در حالی که میانگین نمره کلاس پنجم مدرسه B، ۴ می‌باشد. آیا این دلیل هیجان‌انگیز برای مؤثر بودن روش جدید است؟ مدیر مدرسه A می‌داند که تفاوت بین میانگینها می‌تواند ناشی از تفاوتی باشد که دانش آموزان دو مدرسه قبل از شروع کلاس پنجم داشته‌اند. او نمرات آزمون قرائت را در پایان کلاس چهارم به دست می‌آورد. می‌بیند که دانش آموزان او میانگینی معادل  $4/8$  و دانش آموزان مدرسه B میانگینی معادل  $3/2$  داشته‌اند. در نتیجه او باید این واقعیت را قبول کند که دو گروه دانش آموزان در شروع کلاس پنجم یکسان نبوده‌اند. مدیر برای دانش‌آموزانی که هر دو نمره کلاس چهارم و پنجم آنها در دسترس است، نمره کلاس چهارم را از نمره کلاس پنجم کم می‌کند. میانگین تفاوت نمرات دو سال برای مدرسه A  $1/2+$  و برای دانش آموزان مدرسه B،  $0/8+$  می‌شود. تفاوت  $0/4$  بین میانگین نمره‌های تغییر دو گروه، از تفاوت بین آن دو، یعنی نتیجه‌ای که حاصل در نظر گرفتن تنها نمره‌های کلاس پنجم است و هیچ تعدیلی در مورد عملکرد قبلی آنها صورت نگرفته است، کمتر چشمگیر، اما بیشتر قانع‌کننده است.

آیا اکنون مدیر مدرسه A می‌تواند نتیجه‌گیری کند که روش جدید مؤثرتر از روش قدیم است؟ یا توجه به ماهیت نمره‌های تغییر، پاسخ منفی است. تفاوت در نمره‌های تغییر می‌تواند ناشی از تداوم الگوهای قبلی باشد. در این مورد برای یافتن الگوی تغییرات، مدیر مدرسه A، میانگین نهایی به دست آمده در پایان سال چهارم را تقسیم بر چهار می‌کند (تعداد سالهای تحصیلی در مدرسه). میزان متوسط پیشرفت خواندن برای دانش آموزان مدرسه A، که در پایان سال چهارم میانگین  $4/8$  داشته‌اند،  $1/2$  برای هر سال و برای دانش آموزان مدرسه B،  $0/8$  بوده است. در هر دو مورد، میزان پیشرفت بر مبنای الگوی نمره‌های تغییر، همان چیزی بوده که انتظار می‌رفته است و در نتیجه شواهدی دال بر قبول فرضیه مؤثرتر بودن روش جدید، وجود

ندارد. یک نقیصه ذاتی در نمره های تغییر این است که پیشرفت قابل ملاحظه ظاهری در مقایسه یک گروه با گروه دیگر، می تواند صرفاً ناشی از تداوم الگوی تغییرات گذشته باشد و به هیچوجه ربطی به تیمار نداشته باشد.

مثال دیگری را در نظر بگیریم. چنین فرض می شود که یادگیری لاتین موجب پربارتر شدن گنجینه واژگان انگلیسی دانش آموزان دبیرستان می شود. محققان نمرات واژگان کلاس دهم و دوازدهم دانش آموزانی را که زبان لاتین را یاد گرفتند و آنهایی را که این زبان را یاد نگرفتند، ثبت کردند. چنانچه به شکل ۱۰-۱ نگاه کنیم، می بینیم که میزان پیشرفت نمره واژگان گروهی که لاتین خوانده بودند، از کلاس نهم تا دوازدهم (از ۸۰ تا ۹۵) بیشتر از گروهی بوده است که لاتین نخوانده اند (از ۵۴ تا ۶۳). با این همه، در شکل ۱۰-۱ مشاهده می شود که آن دسته از دانش آموزانی که درس لاتین را انتخاب کرده بودند، صرفاً الگوی قبلی رشد فزاینده واژگان را دنبال می کنند و آنهایی که درس لاتین را انتخاب نکرده بودند، همان الگوی قبلی خود را دنبال می نمایند. میزان پیشرفت هر دو گروه در طول سه سال کلاس ۹ تا ۱۲ همسانی بود که بین کلاسهای ۶ تا ۹ بوده است.

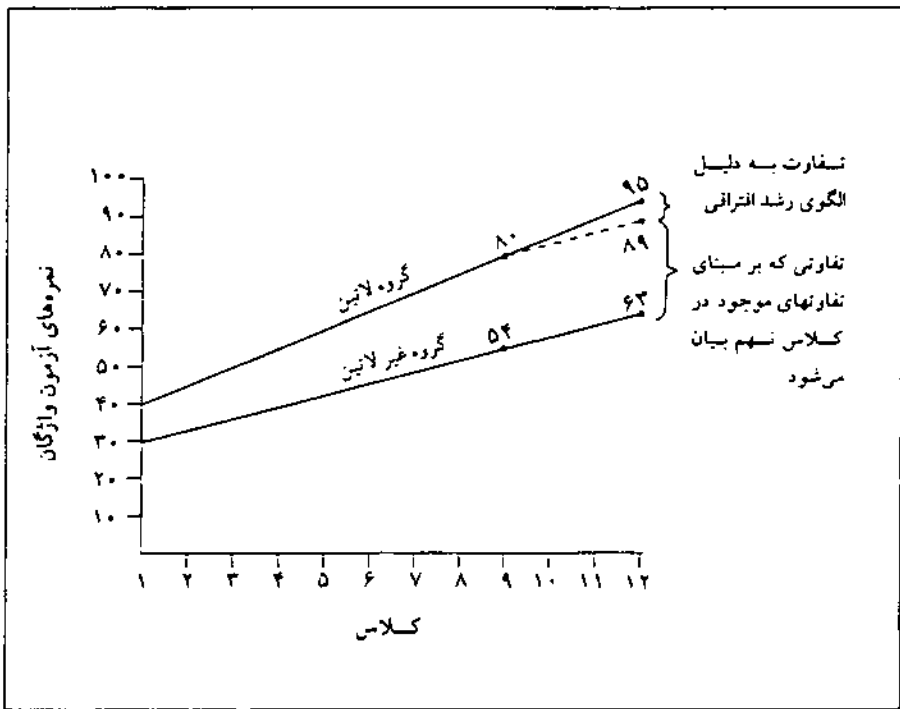
یک محقق بسیار بی تجربه، که تنها نمرات کلاس دوازدهم دو گروه را در نظر می گیرد و تفاوت نقطه آغاز آنها را نادیده می انگارد، ممکن است این چنین نتیجه گیری کند که ۳۲ نمره تفاوت (۶۳-۹۵) ناشی از تجربه یادگیری زبان لاتین است. یک محقق باتجربه تر، با استفاده از نمره های تغییر دو گروه از کلاس ۹ تا کلاس ۱۲، ۶ نمره تفاوت (۹-۱۵) را معلول یادگیری لاتین می داند. به هر حال، این محقق نیز به نتیجه ای اشتباه رسیده است. کدامیک بیشتر احتمال دارد که افراد ناآگاه را منحرف سازد؟ نتیجه گیری محقق اول کمتر محتمل است که مورد قبول واقع شود؛ زیرا این که دو گروه می توانند از نظر تبحر اولیه در واژگان انگلیسی کاملاً با هم متفاوت باشند، امری مسجل است و خواننده گزارش تحقیقاتی به احتمال زیاد نسبت به این نتیجه گیری، که بر مبنای تنها نمرات کلاس دوازدهم استوار است، شک می کند. به نظر می رسد که استفاده از نمره های تغییر، مشکل تفاوت های از قبل موجود را حل می کند و نتیجه گیری محقق دوم به نظر معتبرتر می آید. به هر حال، استفاده از نمره های تغییر، مسئله را تا حدودی حل می کند.

نمره های تغییر، در مورد الف (گروههایی که از نقاط متفاوتی آغاز می کنند) سازگاری دارند. نمره های تغییر، برای ب (تفاوت های از پیش موجود در میزان رشد) سازگاری ندارند. هرگاه



تفاوت‌های از قبل موجود در میزان رشد وجود داشته باشد، تفاوت‌های نمرات بعد از تیمار ناشی از هم الف و هم ب است. استفاده از نمره‌های تغییر اثر کاذب الف را حذف می‌کند، لکن اثر کاذب ب را حذف نمی‌کند. در نتیجه نمره‌های تغییر، تنها یک راه حل ناقص است. سازگاری ناقص می‌تواند ما را گمراه سازد.

شکل ۱ - ۱۰ تأثیر الگوهای قبلی رشد واژگان بر نمره‌های تغییر



### // همتاسازی

یک روش متداول برای تأمین کنترل محدود بررسیهای علی - مقایسه‌ای این است که آزمودنیها را در گروههای آزمایشی و کنترل، حتی المقدور بر مبنای متغیرهای بیرونی هرچه بیشتری همتا کنیم. عمل همتا کردن معمولاً بر مبنای آزمودنی به آزمودنی برای تشکیل زوجهای همتا صورت می‌گیرد. برای مثال، چنانچه کسی بخواهد رابطه بین تجربیات پیشاهنگی و

بزهکاری را مطالعه کند، می‌تواند دو گروه از پسران را با در نظر گرفتن معیارهای خاصی، به عنوان مجرم و غیر مجرم طبقه بندی کند. در چنین بررسی‌ای، عاقلانه است که زوجیهایی از گروههای همتا شده بر مبنای موقعیت اجتماعی - اقتصادی، ساخت خانوادگی و دیگر متغیرهایی که ممکن است هم با تجربیات پیشاهنگی و هم با بزهکاری رابطه داشته باشند، تشکیل شود.

به منظور تعیین اینکه آیا پیشاهنگی مشخصه غیر بزهکاران است یا نه و آیا در سوابق بزهکاران چنین تجربه‌ای وجود نداشته است، می‌توان از تحلیل داده‌های به دست آمده از گروههای همتا شده استفاده کرد. در یک تحقیق علی - مقایسه‌ای، فرد ممکن است با همان مشکلات شیوه همتاسازی که در روش آزمایشی دیده شد، برخورد کند (فصل ۹). در وهله اول، لازمه استفاده از همتا کردن در یک تحقیق علی - مقایسه‌ای این است که محقق بداند عوامل مربوط، یعنی عواملی که امکان دارد با متغیر وابسته همبستگی داشته باشند، کدامند. به علاوه، همتاسازی به میزان زیادی از تعداد آزمودنیهایی که برای تحلیل نهایی مورد استفاده قرار می‌گیرند، می‌کاهد. از دست دادن موردها، که ذاتی شیوه همتاسازی در یک تحقیق علی - مقایسه‌ای است، حتی جدی‌تر از تحقیق آزمایشی است، زیرا در آن همتاسازی، مقدم بر معرفی متغیر مستقل است. حداقل در یک روش آزمایشی (اگرچه توصیه نمی‌شود) این احتمال وجود دارد که موارد جدید جایگزین آنهایی بشود که همتا نمی‌شوند. این امر در تحقیقات علی - مقایسه‌ای امکان پذیر نیست.

یکی از تحقیقات معروف چاپین<sup>(۱)</sup> چگونگی از دست دادن آزمودنیها را در نتیجه عمل همتا کردن، نشان می‌دهد. چاپین می‌خواست درباره اثرات فارغ التحصیلی از دبیرستان، بر روی موفقیت فرد و سازگاری او با اجتماع، پس از یک دوره ده ساله، مطالعه کند. داده‌های اولیه روشن کرد که آنهایی که دبیرستان را به پایان می‌رسانند، موفق‌تر بودند، لکن بررسی بایگانیهای دبیرستانی روشن ساخت که آنهایی که از دبیرستان فارغ التحصیل شده بودند، از نمرات بالاتری در دبستان، سن کمتر، موقعیت شغلی بالاتر والدین، همسایگان بهتر و ... برخوردار بودند. این عوامل می‌توانستند عوامل «علی» هم در امر اتمام دبیرستان و هم در موفقیت‌های بعدی بوده باشند. چاپین به منظور کنترل این عوامل متداول، گروههای فرعی دانش آموزان را بر مبنای تمام

1- Chapin, F. S. (1947). *Experimental Designs in Sociological Research*. New York: Harper & Row.

این عوامل زمینه‌ای همتا کرد، به گونه‌ای که آنها تنها از نظر اتمام دبیرستان متفاوت بودند. عمل همتا کردن، تعداد موارد را از ۱۱۹۴ نفر، به ۴۶ نفر (۲۳ فارغ التحصیل و ۲۳ غیر فارغ التحصیل دبیرستان) کاهش داد، که این رقم کمتر از ۴٪ افرادی است که با آنها مصاحبه شده بود.

مسئله‌ای که حتی از کاهش تعداد آزمودنیها هم جدی‌تر است، نقشی است که رگرسیون آماری در طرحهای علی - مقایسه‌ای با استفاده از گروههای همتا بازی می‌کند. برای روشن شدن مسئله، بهتر است به مثالی که قبلاً در مورد بررسی نحوه یادگیری خواندن آوردیم، برگردیم. به منظور استفاده از گروههای همتا به جای نمره‌های تغییر غیر همتا شده، مدیر مدرسه A، دانش آموزی را از مدرسه‌اش انتخاب می‌کند که در درس قرائت کلاس چهارم نمره ۱/۳ گرفته است و او را با دانش آموزی از دبیرستان B، که در درس قرائت کلاس چهارم همان نمره را گرفته است، همتا می‌سازد. سپس دانش آموزی با نمره ۴/۸ را از مدرسه‌اش با دانش آموزی از مدرسه B با همان نمره و . . . همتا می‌سازد. یا این همه، از آنجا که میانگین جامعه B پایین‌تر از میانگین جامعه A بوده است، در بین جامعه B تعداد زیادی دانش آموز با نمره پایین خواهند بود که همتایی در جامعه A ندارند و بالعکس، در جامعه A تعداد زیادی دانش آموز با نمره بالا حضور دارند که همتایی در جامعه B ندارند. نمرات کلیه دانش آموزانی که همتایی برایشان یافت نشده است، از تحلیل اطلاعات حذف می‌شود. برای آن دسته از دانش آموزانی که در موردشان همتاسازی انجام شده است، میانگین نمرات قرائت کلاس چهارم برای گروه A و B یکسان خواهد بود. در نتیجه، ما عملاً در تشکیل گروهی از مدرسه B که از نظر پیشرفت در درس قرائت نظیر گروه منتخب از مدرسه A هستند، موفق بوده‌ایم.

تا بدینجا اشکالی به نظر نمی‌رسد. حال آیا می‌توانیم تفاوت‌های نمرات قرائت کلاس پنجم را به تفاوت تأثیر گذاری شیوه‌های قدیم و جدید تدریس، نسبت دهیم؟ متأسفانه، پاسخ منفی است. گروههای همتا شده ما، اساساً از دانش‌آموزان ضعیف‌تر جامعه آماری A و دانش‌آموزان قویتر جامعه آماری B، در درس قرائت، تشکیل شده‌اند. نمرات دانش‌آموزان همتا شده جامعه A به طرف میانگین کل جامعه A، یعنی به سمت بالا و نمرات دانش‌آموزان جامعه B به سمت پایین، یعنی به طرف میانگین کل جامعه B میل می‌کند. در نتیجه زمانی که نمرات قرائت کلاس پنجم گروههای همتا شده را مقایسه می‌کنیم، انتظار داریم حتی اگر شیوه جدید مؤثرتر از شیوه قدیم هم نباشد، میانگین A بالاتر از میانگین B باشد.

همتاسازی از این نظر که گروههای آزمایشی و گواهی را شکل می‌دهد که از نظر متغیر یا متغیرهای قبل از تیمار یکسان بوده‌اند، به نظر مفید می‌رسد. با این همه، در مواردی که گروههای همتا از جوامع مختلف انتخاب شده‌اند و این جوامع یکسان نیستند، انتظار می‌رود برگشت به سوی میانگینهای واقعی جامعه‌ها، نتایج کاذب ارائه دهد. معمولاً همتاسازی از داخل یک جامعه واحد، راهبردی مفید است. همتاسازی آزمودنیها از یک جامعه، با آزمودنیهایی از جامعه دیگری، راهبردی نامطلوب تلقی می‌شود. همانند استفاده از نمره‌های تغییر، همتاسازی نیز تنها تا حدودی تفاوت‌های از پیش موجود بین گروهها را حل می‌کند. این سازگاری ناقص می‌تواند همانند مورد نمره‌های تغییر، ما را منحرف سازد.

### // تحلیل کوواریانس

مزیت ANCOVA در این است که به جای داده‌های زوجهای همتا شده، می‌توان از داده‌های مربوط به تمام آزمودنیها استفاده کرد. چنانچه در فصل ۹ دیدیم، ANCOVA شیوه‌ای عالی برای بالا بردن سطح دقت در آزمایش است که آزمودنیهای آن به طور تصادفی انتخاب شده باشند. همچنین در بعضی موارد از ANCOVA برای تعدیل جزئی تفاوت‌های اولیه در یک طرح علی-مقایسه‌ای استفاده می‌شود. با این وجود، از آنجا که این تعدیل در تحقیق علی-مقایسه‌ای ناقص است، ANCOVA مشکل تفاوت‌های اولیه را بین گروهها حل نمی‌کند، بلکه از شدت آن می‌کاهد. در تفسیر نتایج تحقیقات علی-مقایسه‌ای، فرض اینکه ANCOVA به میزان رضایت بخشی تفاوت‌های اولیه را تعدیل کرده است، شایسته نیست.

نمونه‌ای شناخته شده از مخاطرات همتاسازی و تحلیل واریانس، در بررسی علی-مقایسه‌ای دانشگاه اوهایو و وستینگهاوس در زمینه اثرات برنامه هد استارت، مشاهده می‌شود.<sup>(۱)</sup> این بررسی پیشرفت تحصیلی دانش آموزانی را که در برنامه هد استارت شرکت داشتند، با دانش آموزانی که در این برنامه شرکت نداشتند، مقایسه کرد. کودکانی که در برنامه

1- Cicirelli, V., et al. (1969). *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*. (A report Presented to the Office of Economic Opportunity pursuant to contract B88-4536.) Westinghouse Learning Corporation and Ohio University. (Distributed by the Clearinghouse for Federal Scientific and Technical Information, U.S. Department of Commerce, National Bureau of Standards, Institute for Applied Technology. PB 184 328.)

شرکت داشتند، با کودکان نواحی مجاور که در برنامه شرکت نداشتند، با در نظر گرفتن متغیرهای جنس، گروههای نژادی و قومی و حضور در کودکستان، همتا شدند. به منظور تعدیل تفاوت‌های موجود در درآمد سرانه، سطح تحصیلی و سطح اشتغال پدر خانواده، از تحلیل کوواریانس استفاده شد. نتایج نشان داد که حتی با وجود آنکه نمرات از نظر تفاوت‌های اولیه تعدیل شدند، موفقیت گروهی که در برنامه شرکت نداشتند، بیشتر از گروهی بود که در برنامه شرکت داشتند و چنین نتیجه‌گیری شد که برنامه هد استارت اثری مخرب دارد.

کمپبل و ارل باچر<sup>(۱)</sup> اشاره می‌کنند از آنجا که هر دو شیوه همتا سازی و تحلیل کوواریانس از تعدیل کامل تفاوت‌های اولیه بین گروهها عاجز هستند، انتظار داریم نمرات تعدیل شده بعد از تیمار گروه محروم، پایین تر از نمرات تعدیل شده بعد از تیمار گروه غیر محروم باشد. از آنجا که مقدار تعدیل ناقص ناشناخته است، نمی‌توانیم با اطمینان نتیجه‌گیری کنیم که تجربه هد استارت اثری مخرب یا مفید دارد یا اصلاً اثری ندارد.

### // گروههای همگن

همان گونه که در مبحث کنترل آزمایشها گفته شد، امکان دارد که اثرات یک متغیر را با انتخاب نمونه‌ها، به گونه‌ای که تا حد امکان از نظر یک متغیر همگن باشند، کنترل کرد. شیوه مشابهی نیز در تحقیقات علی - مقایسه‌ای ممکن است دنبال شود. محقق به جای آنکه یک گروه ناهمگن را انتخاب و گروههای فرعی همتا شده را در گروه اصلی با هم مقایسه کند، می‌تواند در نمونه خود تنها آزمودنیهایی را که از لحاظ متغیر مورد نظر، همگن هستند، بگنجانند و در نتیجه آن متغیر را کنترل کند. چنانچه هوش متغیر نامرتبط باشد، محقق اثر آن را می‌تواند با استفاده از آزمودنیهایی که فقط از نظر هوش در یک سطح هستند، کنترل کند. با این شیوه به هدف جداسازی متغیر مستقل مورد نظر از سایر متغیرها، که عموماً همبسته هستند، دست می‌یابیم، به گونه‌ای که می‌توان هر اثری را که دریافت می‌شود، به طور موجه به متغیر مستقل نسبت داد. فرض کنید فردی بخواهد در مورد اینکه داشتن یک اتاق آرام در منزل، میز تحریر، کتابها و ... برای درس خواندن، چه اثراتی بر روی نمرات دانش آموزان دبیرستان دارد، تحقیق کند.

1- Campbell D.T., and A. Erlebacher (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In V. Hellmuth (ed.), *Compensatory Education: A National Debate*, vol. 3: Disadvantaged Child, New York: Brunner/Mazel.

چنانچه محقق صرفاً مقطعی از دانش آموزان دبیرستانی را انتخاب کند و از هر کدام از آنها سؤال کند که آیا اطاق آرامی برای مطالعه دارند یا نه و سپس نمرات آنهایی را که چنین اطاقی دارند، با آنهایی که ندارند، مقایسه کند، ممکن است اشتباهاً نتیجه گیری کند که داشتن یک اطاق آرام برای مطالعه، عامل کسب نمرات خوب می شود. توصیف دیگر به این صورت است که دیگر عوامل مرتبط با طبقه اجتماعی وجود دارند که هم چگونگی محیط مطالعه فرد و هم نمراتی را که در دبیرستان می گیرد، تحت تأثیر قرار می دهند. برای مثال، خانواده های با درآمد کم در مقایسه با خانواده های متعلق به طبقه اجتماعی متوسط و بالا، احتمالاً بیش از حد پر جمعیت هستند و معمولاً تأکید کمتری بر روی پیشرفت تحصیلی و کسب معلومات دارند. چنانچه محقق بخواهد اثرات طبقه اجتماعی را کنترل کند، می تواند بررسی اش را به آزمودنیهای متعلق به یک طبقه اجتماعی - اقتصادی محدود کند. چنانچه بین چگونگی مطالعه و نمرات (در یک طبقه اجتماعی) رابطه ای پیدا کند، می تواند اطمینان بیشتری نسبت به این نتیجه گیری داشته باشد که تفاوت های مشاهده شده در نمرات به دلیل چگونگی محیط مطالعه بوده است و نه تفاوت های اجتماعی - اقتصادی. البته چنین روشی، تعمیم یافته ها را فقط به همان طبقه اجتماعی که در بررسی مورد استفاده قرار گرفته است، محدود می کند. فرد نمی تواند بداند که آیا این رابطه در دیگر طبقات اجتماعی هم وجود دارد یا خیر.

استفاده از نمونه های همگن، تنها یک راه حل جزئی برای مشکلات تحقیق علی - مقایسه ای است. با انتخاب نمونه هایی از گروههایی که احتمال مشترک بودن آنها در یک متغیر علی وجود دارد، می توانیم برخی از متغیرهای علی مشترک را کنترل کنیم. نمی توانیم مطمئن باشیم که آزمودنیها از نظر کلیه متغیرهای علی مشترک احتمالی و غیراحتمالی، یکسان هستند. در مثال یاد شده، کاملاً احتمال دارد که در خانواده های متعلق به طبقه متوسط، هم داشتن و هم نداشتن اطاقی آرام برای مطالعه و پیشرفت تحصیلی معلول تعداد افراد خانواده و یا متغیرهای دیگر باشد. مسئله رگرسیون نیز مشکل دیگری در مورد نمونه های همگن است. برای مثال، در مورد تحقیقی که قبلاً ذکر کردیم، اگر مدیر مدرسه A در دو مدرسه تنها کودکانی را انتخاب می کرد که دقیقاً نمراتشان در امتحان قرائت کلاس چهارم، ۴ بود، رگرسیون سبب می شد که گروه A به سمت بالا و گروه B به سمت پایین تمایل پیدا کند؛ همانگونه که در مورد نمونه های هم مشاهده دیدیم.

## // وارد کردن متغیرهای نامرتبط در طرح

این امکان وجود دارد که متغیرهای مستقل بیرونی مرتبط را در یک طرح علی - مقایسه‌ای وارد و از روش تحلیل واریانس عاملی استفاده کرد.

برای مثال، فرض کنیم هوش یک متغیر بیرونی مرتبط است و امکان کنترل آن از طریق هم‌سازی و یا وسیله‌ای دیگر وجود ندارد. در این مورد، می‌توان هوش را به عنوان یک متغیر مستقل دیگر به طرح اضافه کرد و آزمودنیهای بررسی را بر مبنای سطح هوش طبقه بندی کرد. سپس اندازه‌های متغیر وابسته از طریق تحلیل واریانس تجزیه و تحلیل خواهند شد و اثرات اصلی و تعاملی هوش تعیین می‌شوند. چنین روشی هرگونه تفاوت معنی داری را در گروهها، از نظر متغیر وابسته، روشن می‌سازد، ولی نمی‌توان رابطه علت و معلولی را بین هوش و متغیر وابسته فرض کرد. دیگر متغیرهای بیرونی نیز می‌توانند به گونه‌ای عمل کنند که هم عامل اثر اصلی بوده و هم اثر تعاملی داشته باشند.

برای مثال، تحقیقات بسیاری وجود رابطه بین ترتیب تولد و میزان هوش و موفقیت شغلی را گزارش کرده‌اند؛ بدین ترتیب که: اولین فرزند بهتر از دومی، فرزند دوم بهتر از سومی و ... عمل می‌کند. پس از بازنگری دقیق و همه جانبه پیشینه تحقیقاتی، شولر<sup>(۱)</sup> چنین نتیجه‌گیری کرد: به نظر می‌رسد که بیشترین تفاوت‌هایی که در مورد ترتیب تولد گزارش شده است - یعنی قرار گرفتن فرزند اول خانواده در میان گروههایی که از هوش استثنایی یا پیشرفت شغلی فوق‌العاده برخوردارند - به طور موجز بر حسب تفاوت‌هایی که در روند تعداد افراد خانواده طبقات مختلف اجتماعی دیده می‌شود، قابل تبیین است. از آنجا که میان طبقه اجتماعی و تعداد افراد خانواده، رابطه‌ای معکوس وجود دارد، متوسط موقعیت اجتماعی فرزند پنجم پایین‌تر از متوسط موقعیت اجتماعی فرزند چهارم خواهد بود و الی آخر. چون طبقه اجتماعی با موفقیت رابطه دارد، رابطه آشکار بین ترتیب تولد و موفقیت می‌تواند صرفاً ناشی از تفاوت‌های بین موقعیت اجتماعی متوسط خانواده‌ها باشد و نه ترتیب تولد به خودی خود.

بعدها بلمونت و مارولا<sup>(۲)</sup>، بررسی‌ای بر روی ترتیب تولد و نمرات آزمون ماتریس‌های مدرج ریون<sup>(۳)</sup> (یک آزمون غیر کلاسی هوش) با نمونه‌ای شامل حدود ۴۰۰۰۰۰ مرد جوان در

1- Schooler, C. (1972). Birth order effects: Not here, not now! *Psychological Bulletin*, 72, 161-175.

2- Belmont, L., and F. A. Marolla (1973). Birth order, family size, and intelligence. *Science*, 182, 1096 - 1101.

3- Raven Progressive Matrices Test

هلتند، انجام دادند. چنین نتیجه گرفتند که در میان همه خانواده‌ها با هر تعداد فرزند، میزان هوش بر حسب ترتیب تولد، کاهش می‌یابد. از آنجا که آنها نمره‌ها را فقط در داخل خانواده‌های پنج فرزندی، در داخل خانواده‌های چهار فرزندی و... مقایسه کردند، تفاوت طبقات اجتماعی نمی‌توانست عامل نتایج حاصله باشد.

حال آیا نتایج بررسی بلمونت و مارولا، ما را قادر می‌سازد که بگوییم ترتیب تولد به خودی خود با هوش رابطه دارد؟ عاقلانه است که این رابطه را یک نتیجه آزمایشی تلقی کنیم. دلیل برخورد محتاطانه ما این است که احتمال دارد متغیرهایی بجز طبقه اجتماعی، عامل این رابطه آشکار باشند. مسلماً سن متوسط والدین کودک چهارم خانواده بالاتر از سن متوسط والدین کودک سوم است و این خود می‌تواند نقشی در این پدیده داشته باشد. شاید عاملی موجب این امر شده باشد که ما حتی تصویری از آن نداریم.

وارد کردن متغیرهای دیگر در طرح علی - مقایسه‌ای تنها یک راه حل جزئی است، اما هرگز نمی‌توانیم مطمئن باشیم که متغیرهای درست را انتخاب کرده‌ایم و یا تمامی متغیرهایی را که باید مورد مطالعه قرار گیرند، در طرح خود گنجانده‌ایم.

در تمام طول این فصل به این مسئله اشاره شد که نقیصه اصلی طرحهای علی - مقایسه‌ای، فقدان کنترل در آن است. از آنجا که به کارگیری شیوه تخصیص تصادفی آزمودنیها به گروهها، یا دستکاری مستقیم متغیر مستقل در یک موقعیت کنترل شده، امکان پذیر نیست، همواره این احتمال وجود دارد که متغیرهای کنترل نشده‌ای وجود داشته باشند که عامل بروز تغییرات در متغیر وابسته باشند. به این دلیل، نمی‌توان فرض کرد که گروهها در ابتدای کار مشابه بوده‌اند. از آنجا که محقق کنترلی بر این امر ندارد که چه کسی در معرض آن تجربه خاص قرار گرفته و چه کسی قرار نگرفته، کاملاً محتمل است که عامل دیگری در گراگرد افراد و یا محیط آنها تعیین کند که چه کسی در معرض آن تجربه قرار گیرد. بنابر این، همان «چیز دیگر» ممکن است متغیر مستقل حیاتی باشد.

در نتیجه کنترل ناکافی در یک بررسی علی - مقایسه‌ای، تفسیر یافته‌ها می‌توانند به صورت خاص مخاطره‌آمیز باشند. احتمال نادرستی بسیار زیاد است. زمانی که بر مبنای یک بررسی علی - مقایسه‌ای، رابطه‌ای بین دو متغیر برقرار می‌شود، در واقع فقط اولین قدم تجزیه و تحلیل برداشته شده است. محقق باید فرضیه‌های جایگزین محتمل را در نظر گیرد و آزمون کند و حتی



پس از آن نیز این احتمال را بدهد که رابطه آشکار بین متغیرها می‌تواند ناشی از علت درک ناشدنی دیگری باشد. فرایندهایی چون به کارگیری نمره‌های تغییر، همنا سازی، تحلیل کوواریانس، همبستگی جزئی، گروه‌های همسان و وارد کردن متغیرهای بیرونی، می‌توانند کمک کنند تا از خطاهای عظیم در بررسی‌های علی - مقایسه‌ای اجتناب کنیم، ولی همه آنها تفاوت‌های پیش از تیمار بین گروهها را فقط تا حدی تعدیل می‌کنند. فهرست کردن فرضیه‌های جایگزین (عامل مشترک، علیت معکوس و متغیرهای مستقل دیگر) می‌تواند در امر ارزشیابی واقعی تر نتایج علی - مقایسه‌ای، به ما کمک کند.

### /// طرح ریزی یک بررسی علی - مقایسه‌ای

روال اصلی تحقیقات علی - مقایسه‌ای عبارتند از: (۱) با آزمودنی‌هایی که از نظر متغیر مستقل متفاوت هستند، شروع و فرضیه‌هایی را در این زمینه که آنها در متغیر وابسته چگونه با هم تفاوت خواهند داشت، آزمون کنیم و (۲) با آزمودنی‌هایی که از نظر متغیر وابسته متفاوت هستند، آغاز و فرضیه‌های مربوط به متغیرهای مستقل احتمالی را آزمون کنیم. مثالی از روال اول، تحقیقی است که در مورد مقایسه عملکرد مسئله‌گشایی دانشجویان خلاق و غیر خلاق دانشگاه می‌شود. فرضیه چنین تبیین می‌شود: دانشجویان خلاق، سرعت و دقت بیشتری را در مسئله‌گشایی نسبت به دانشجویان غیر خلاق نشان می‌دهند. این فرضیه به طور روشن نیاز به یک طرح علی - مقایسه‌ای را مشخص می‌سازد؛ زیرا محقق نه می‌تواند خلاقیت را تغییر دهد و نه دانشجویان را به صورت تصادفی در گروهها تقسیم کند. او باید با دو گروه که در حال حاضر از نظر متغیر مستقل (خلاقیت) متفاوتند، شروع و سپس آنها را از نظر متغیر وابسته (عملکرد مسئله‌گشایی) مقایسه کند.

محقق باید «دانشجوی خلاق» و «دانشجوی غیر خلاق» را با واژه‌های کاربردی دقیق تعریف کند. دانشجوی خلاق می‌تواند به دانشجوی دوره لیسانس در دانشکده علوم تربیتی که نمره‌اش هم در آزمون «کاربردها و پیامدهای جایگزین گیلفورد»<sup>(۱)</sup> و هم در آزمون آناگرام<sup>(۲)</sup> در چارک بالایی قرار دارد، اطلاق شود. دانشجویی که نمره‌اش زیر چارک اول باشد، دانشجوی غیر خلاق محسوب می‌شود.

محقق باید سعی کند متغیرهای دیگر، بجز خلاقیت، را که می‌توانند بر روی متغیر وابسته (عملکرد مسئله‌گشایی) اثر گذارند، شناسایی کند و قدمهایی برای هم‌ارز سازی دو گروه آزمایشی و گواه در مورد این متغیرها، از طریق شیوه‌های هم‌تاسازی یا به وسیله تکنیکهای آماری، بردارد. برای مثال، در این بررسی، دیگر متغیرهای مستقل که باید کنترل شوند، عبارتند از: هوش، جنس و یا احتمالاً رشته تحصیلی و سال تحصیلی در دانشگاه. یک دانشجوی تیزهوش سال دوم در گروه خلاق باید با هم‌تایش در گروه غیرخلاق زوج شود.

پس از تشکیل گروههای هم‌تا، به هر دو گروه یک تکلیف مسئله‌گشایی، که اندازه‌ای از متغیر وابسته است، داده می‌شود. تحلیل عمیق تر داده‌ها به وسیله آزمون t، هرگونه تفاوت معنی‌دار در عملکرد مسئله‌گشایی بین هر دو گروه، و احتمالاً رابطه‌ای را بین خلاقیت و عملکرد مسئله‌گشایی نشان می‌دهد. هرچند ممکن است محقق بتواند از این بررسی نتیجه‌گیری کند که رابطه وجود دارد، ولی نمی‌تواند فرض کند که بین خلاقیت و عملکرد مسئله‌گشایی، پیوند علت و معلولی وجود دارد. احتمال دارد متغیرهای غیرکنترل شده دیگری وجود داشته باشند که به تنهایی و یا با هم، بر مسئله‌گشایی اثر گذارده باشند. یک محقق هوشیار به ضرورت بررسی تبیین‌های جایگزین پذیرفتنی دیگر درباره یافته‌های یک تحقیق علی-مقایسه‌ای، آگاهی دارد. برخی مواقع متغیر مستقل مانند جنس، موقعیت اجتماعی-اقتصادی و یا قومی، ویژگیهایی می‌باشد که نمی‌توان آنها را تغییر داد. برای مثال، رابینسون - آوانا، کهل و جنسن<sup>(۱)</sup>، عزت نفس<sup>(۲)</sup> دختران و پسران کلاس هفتمی را مقایسه کردند و با استفاده از «پرسشنامه مدرسه‌ای مختصر عزت نفس کوپر اسمیت» (SEI)<sup>(۳)</sup> به عنوان میزانی برای متغیر وابسته، چنین یافتند که عزت نفس پسران به صورت قابل ملاحظه‌ای بالاتر از دختران است. سپس آزمودنیها را از نظر پیشرفت تحصیلی به سه گروه زیر متوسط، متوسط و بالای متوسط، تقسیم کردند و چنین یافتند که با افزایش پیشرفت تحصیلی، میزان عزت نفس بالا می‌رود. سپس محققان دوباره SEI را به کار بردند و از آزمودنیها خواستند چنان پاسخ دهند که فکر می‌کنند یک همسن و هم کلاستان از جنس مخالف پاسخ می‌دهد. پسران در هر سه سطح پیشرفت تحصیلی عزت نفس پایین‌تری را در مقابل دختران نشان دادند. دختران متعلق به گروههای پایین متوسط و متوسط از

1- Robinson - Awana, P., T.J. Kehle, and W.R. Jensen (1986). But what about smart girls? Adolescent self-esteem and sex role perceptions as a function of academic achievement. *Journal of Educational Psychology*, 78 (3), 179-183.

2- Self - Esteem

3- Cooper Smith Self - Esteem Inventory (SEI)

نظر پیشرفت تحصیلی، عزت نفس بالاتری را در مقابل پسران و دختران متعلق به گروه بالاتر از متوسط، عزت نفس پایین‌تری را در مقابل پسران نشان دادند.

در برخی مواقع متغیر مستقل به گونه‌ای است که می‌توان آن را دستکاری کرد، ولی عملاً به دلایل علمی و یا اخلاقی از کنترل محقق خارج است. برای مثال، اخلاقی نیست که به صورت تصادفی گروهی از دانش آموزان را در یک کلاس نگاه داریم و به گروه دیگر اجازه دهیم که به کلاس بالاتر بروند، برای آنکه اثرات مردود شدن در یک کلاس را بر موفقیت‌های بعدی بررسی کنیم. لکن، از طریق هم‌تا کردن شاگردان مردود شده و مردود نشده بر حسب تمام متغیرهای مربوط ممکن در مورد مسئله، بینشی کسب می‌کنیم. پیترسون، دوگراسی و آیایی<sup>(۱)</sup> دانش آموزان مردود کلاسهای اول و دوم و سوم را با دانش آموزان همسال مردود نشده در یک نظام آموزشی هم‌تا کردند. آزمودنیها بر حسب جنس، سن تقویمی، نمرات کلی خواندن، زبان و ریاضی در آزمون «پیشرفت کالیفرنیا»، هم‌تا شدند. در سالهای بعد از مردود شدن، دانش آموزان مردود شده را بر مبنای رتبه نسبی آنها در کلاس با دانش آموزان مردود نشده هم‌تایشان مقایسه کردند. رتبه نسبی دانش آموزان مردود کلاس اول و دوم در دروسهای قرائت و ریاضی، نسبت به هم‌تایان غیر مردودشان، به صورت معنی داری بالاتر بود، البته این امر در مورد زبان صدق نمی‌کرد. تفاوت بین مردودین کلاس سوم و هم‌تایان آنها که در زمان بررسی در کلاس چهارم بودند، معنی دار نبود.

محققان همچنین عملکرد دانش آموزان مردود شده را در کلاس بالاتر با هم‌تایان مردود نشده آنها در سال قبل مقایسه کردند و چنین نتیجه گرفتند که دانش آموزان مردود شده عملکرد بهتری از هم‌تایان مردود نشده خود در کلاسهای دوم و سوم، ولی نه کلاس اول، داشته‌اند.

در روال دوم تحقیق علمی - مقایسه‌ای، دو یا چند گروه از نظر متغیر وابسته متفاوت هستند و فرضیه‌ها بر مبنای متغیرهای مستقل احتمالی، آزمون می‌شوند. ده عضو یک تیم کشتی را در نظر بگیرید که سه نفر از آنها در بازگشت از یک مسابقه قهرمانی بیمار می‌شوند. برای بررسی علت بیماری، پزشک از آنها سؤال می‌کند که در راه بازگشت به شهرشان، وقتی برای صرف غذا توقف کردند، چه خوردند. او در می‌یابد که هر سه نفر تمام غذاهایشان متفاوت بوده است، به

1- Peterson, S. E., J. S. DeGracie, and C. R. Ayabe (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal*, 24, 107 - 118

استثنای یک نوع خوراک مرغ باخامه که هر سه نفر از آن خورده بودند. هشت کشتی گیری که مریض نشده بودند، خوراک مرغ با خامه نخورده بودند. او چنین نتیجه گیری می کند که علت ابتلای به بیماری، خوراک مرغ با خامه بوده است. در واقع پزشک، روش مشترک توافقی و تفاوت فیلسوف معروف، جان استوارت میل را به کار می گیرد که می گوید:

اگر دو یا چند مصداق که پدیده در آنها رخ می دهد، تنها یک وجه را به طور مشترک دارا باشند، حال آنکه دو یا چند مصداق دیگر که پدیده در آنها رخ نمی دهد، به غیر از فقدان آن وجه هیچ اشتراکی نداشته باشند، آن وجه که تنها تفاوت این دو مجموعه از مصداقهاست، معلول، یا علت، و یا بخش لازم علت آن پدیده است.<sup>(۱)</sup>

به عنوان مثالی از روال دوم تحقیق علی - مقایسه ای می توان به بررسی بن<sup>(۲)</sup> در مورد متغیرهای مرتبط با احساس ایمنی و وابستگی پسران ۱۸ ماهه نسبت به مادران شاغلشان، اشاره کرد. کیفیت دلبستگی مادر - فرزند (متغیر وابسته) از طریق تحلیل نوارهایی تصویری از پسران، که قبلاً تهیه و ضبط شده بود، اندازه گیری شد و از این طریق، توازن دلبستگی و رفتار جستجوگرانه کودکان در حضور و غیاب مادر و یک شخص ناآشنا مورد ارزشیابی قرار گرفت. نتایج این بررسی نشان داد که در بین پسرانی که احساس دلبستگی مطمئنی داشتند، رابطه مادر - پسر دارای ویژگیهای ملایمت، پذیرش و آزادی ابراز عواطف بود. مادران پسرانی که از نظر دلبستگی احساس امنیت می کردند، در طول سال اول زندگی فرزندشان زودتر از مادران پسرانی که از نظر دلبستگی احساس امنیت نمی کردند، سرکار خود حاضر می شدند. طبقه اجتماعی - اقتصادی و شیوه بچه داری، رابطه ای با دلبستگی مادر - پسر نداشت.

### /// نقش تحقیق علی - مقایسه ای

با توجه به مخاطراتی که در تفسیر نتایج تحقیقات علی - مقایسه ای وجود دارد، بسیاری از صاحب نظران امر تحقیق عقیده دارند که اساساً نباید به این گونه تحقیقات پیردازیم. اصولاً آنها چنین بحث می کنند که بهتر است که ما نادان باشیم، ولی خطر نتیجه گیریهای نادرست را نپذیریم.

1- Mill, J.S. (1846). A system of logic (p.229). New York: Harper & Brothers.

2- Benn, R. K. (1986). Factors promoting secure attachment relationships between employed mothers and their sons. *Child Development*, 57, 1224-1231.

از سوی دیگر، عده‌ای چنین دلیل می‌آورند که بسیاری از متغیرهایی که برای ما حائز اهمیت هستند، از طریق تحقیق آزمایشی، قابل بررسی نیستند. ما نمی‌توانیم کودکان را به‌طور تصادفی در گروه‌های متعلق به خانواده‌های از هم گسیخته یا سالم و طبقه اجتماعی بالا و پایین تقسیم کنیم و یا آنها را در گروه‌های همسالان موفقیت طلب یا غیرموفقیت طلب و غیره قرار دهیم. بنابراین چنانچه بخواهیم هرگونه اطلاعی را در مورد روابط بین این گونه متغیرهای خصیصه‌ای و دیگر متغیرها به دست آوریم، به کارگیری روش علی - مقایسه‌ای تنها راه چاره است. در صورتی که روش‌های مناسب کنترل محدود را به کار بریم و فرضیه‌های جایگزین را در نظر داشته باشیم، احتمالاً بیشتر از آنکه اشتباه کنیم، به نتایج درست دست خواهیم یافت.

مسئله بررسی‌های بسیار معتبر علی - مقایسه‌ای انجام گرفته است. بررسی در مورد رابطه بین کشیدن سیگار و سرطان ریه، نمونه‌ای شناخته شده از این نوع تحقیقات است. نمی‌توان گروهی از مردم را وادار کرد که برای چندین سال سیگار بکشند و گروهی دیگر را از کشیدن سیگار باز داشت، در نتیجه تنها راه انجام تحقیق در این مورد، به کارگیری روش علی - مقایسه‌ای است. فرضیه علیت معکوس که می‌گوید سرطان ریه مردم را به سیگار کشیدن وامی‌دارد، مسلماً قابل قبول نیست. هیچ یک از فرضیه‌های پیشنهاد شده عامل مشترک، مثلاً اینکه آیا افراد عصبی هم در معرض ابتلای سرطان و هم سیگاری شدن هستند و یا پاره‌ای جابجایی‌های ژنتیکی عامل هر دو است و غیره، محتمل نیستند.

پزشکان متغیرهای مستقل جایگزین بسیاری را از طریق انتخاب نمونه‌هایی از مناطقی که هوای آلوده و هوای پاک دارند، کنترل کردند. شواهد آزمایشی در مورد حیوانات که وادار به استنشاق دود سیگار شدند و گروه دیگری که در معرض دود سیگار قرار داده نشدند، دلایلی را برای وجود یک رابطه علی - تأثیری فراهم آورد. با توجه به تمام مباحث فوق و به رغم خطر ذاتی تحقیقات علی - مقایسه‌ای، اکثریت ما به این نتیجه می‌رسیم که بهتر است بر این امر شرط بندی کنیم که بین سیگار کشیدن و سرطان ریه رابطه‌ای علی - تأثیری وجود دارد، تا اینکه بگوییم بین این دو رابطه‌ای وجود ندارد.

همگی ما احتمالاً متأسف خواهیم شد از اینکه در بسیاری موارد که روش آزمایشی واقعی و یا حداقل طرح‌های شبه آزمایشی می‌توانستند به کار گرفته شوند، از روش تحقیق علی - مقایسه‌ای استفاده شده است. این واقعیتی مأیوس کننده است که بارها سازمان‌های دولتی و یا

محلی و یا دولت فدرال، با آوردن آزمودنیهایی که همگی انتخابی بوده‌اند، برنامه‌های جدیدی را آزمون و سعی کرده‌اند اثرات برنامه را بعد از واقعه<sup>(۱)</sup> ارزیابی کنند. بسیار اتفاق می‌افتد که سازمانهای دولتی، از جمله مدارس، در پاسخ به یک فشار اجتماعی، که مثلاً در مورد یک مشکل خاص، کاری صورت دهند به امر سازماندهی یک برنامه جدید برای کلیه آزمودنیهای منتخب پرداخته‌اند. اوانس<sup>(۲)</sup> (عضو دفتر آموزش و پرورش ایالات متحده آمریکا) و شیلر<sup>(۳)</sup> (عضو دفتر موقیتهای اقتصادی)، چگونگی پاسخگویی سازمانهای دولتی را به این‌گونه فشارها، شرح می‌دهند:

کوششهایی که در جهت فراهم آوردن موقعیت لازم برای تخصیص تصادفی آزمودنی‌ها صورت می‌گیرد، همواره مورد اعتراض متقاضیان برنامه قرار می‌گیرد؛ به این ترتیب که چنین روشی برای افرادی که به عنوان گروه گواه انتخاب می‌شوند، محرومیت اختیاری از برنامه را به دنبال دارد. در بین ناراضیان، معترضان بر خروشان، به مقامات رسمی و نمایندگان سنا شکایت می‌برند. در نتیجه مدیران برنامه که می‌خواهند از مواجهه با این امر اجتناب کنند، جانب مخالفان را می‌گیرند.

اگرچه مایل نیستیم، اما تجربه ما را به چنین نتیجه‌گیریهایی وا می‌دارد که ماهیت فرایندهای سیاسی و دولتی باعث می‌شود در مواردی که از نظر زمان محدودیت داریم، با بار سنگینی وجود دارد که بی شک در آنها برنامه‌های مربوط به فعالیتهای اجتماعی بروز خواهد کرد و یا در مواردی که عمل تخصیص تصادفی به ندرت امکان پذیر است، این وضعیت ادامه داشته باشد. متأسفانه فرایند سیاسی، منظم، با برنامه و منطقی نیست. نقطه اوج پشتیبانی عمومی و کنگره از برنامه‌های فعالیت اجتماعی معمولاً بسرعت و بدون مقدمه بروز می‌کند. به محض آنکه قانونی تصویب شد، فشار بر مدیران، برای اجرای سریع برنامه، شدت می‌یابد. در چنین شرایطی، که قاعده است و نه استثنا، منازعات در مورد آنکه برنامه‌ها باید با دقت و در خط یک آزمایش واقعی با تخصیص تصادفی آزمودنیها، به گونه‌ای انجام شود که بتوانیم با اطمینان اثربخش بودن برنامه را ارزیابی کنیم، محکوم به فراموشی است.<sup>(۴)</sup>

1- Ex - Post Facto

2- Evans

3- Schiller

4- Evans, V. W., and J. Schiller (1970). How preoccupation with possible regression artifacts can lead to a faulty strategy for the evaluation of social action programs: A reply to Campbell and Erlebacher. In V. Hellmuth (ed.), *Compensatory Education: A National Debate*, vol. 3: *Disadvantaged Child*, New York: Brunner/Mazel.

به نظر ما نارضایتی متقاضیانی که از یک برنامه تازه «محروم» شده‌اند، تنها قطره‌ای است در مقابل سیل نارضایتی پرداخت کنندگان مالیات، که در می‌یابند میلیونها دلار صرف تولید برنامه‌ای شده است که فاقد روش برنامه ریزی مناسب جهت تعیین این امر است که آیا این برنامه‌ها عملاً چیزی را تکامل بخشیده‌اند یا خیر. باید راهی برای برطرف کردن مشکلات روابط عمومی در یک تخصیص تصادفی وجود داشته باشد. آیا دولت نمی‌تواند برنامه‌ای را با تأمین کل هزینه‌اش به چند مدرسه پیشنهاد کند، با این شرط که در آن نیمی از دانش‌آموزان به صورت تصادفی انتخاب شوند؟ چنانچه برنامه به اندازه کافی جذاب باشد، می‌توان تصور کرد که بسیاری از جوامع مایل خواهند بود در آن شرکت کنند، زیرا حساب می‌کنند که نیمی تصادفی از یک قرص نان، بهتر از هیچ است. پس از به دست آوردن شواهد، دولت می‌تواند برنامه‌های با ارزش را به تمامی دانش‌آموزان ارائه دهد و به تدریج برنامه‌های بی‌ارزش را رها کند.

### /// خلاصه

از طرح‌های علی - مقایسه‌ای زمانی استفاده می‌شود که محقق در موقعیتی نباشد که بتواند فرضیه‌هایش را با تخصیص آزمودنیها به موقعیتهای مختلف، که امکان دستکاری مستقیم در آنها باشد، آزمون نماید. در تحقیق علی - مقایسه‌ای، تغییرات در متغیر مستقل قبلاً رخ داده‌اند و محقق باید آنها را با بازنگری تأثیرات احتمالی آنها بر روی یک متغیر وابسته مشاهده شده، مطالعه کند.

اگرچه نقاط ضعف زیادی در رویکرد علی - مقایسه‌ای وجود دارد، اما معمولاً تنها روشی است که توسط آن محققان تعلیم و تربیت می‌توانند اطلاعات لازم را در مورد ویژگیهای گروههای تعریف شده دانش‌آموزان و یا جهت طرح‌ریزی هوشمندانه برنامه‌های مدرسه، فراهم آورند. این رویکرد به محقق اجازه می‌دهد که موقعیتهایی را بررسی کند که در آنها ارائه متغیر کنترل شده امکان پذیر نیست. متغیرهایی مانند هوش، خلاقیت، موقعیتهای اجتماعی - اقتصادی و شخصیت معلم را نمی‌توان تغییر داد و در نتیجه باید به جای رویکرد آزمایشی دقیق، از تحقیق علی - مقایسه‌ای برای بررسی آنها استفاده کرد.

بسیاری از تحقیقات علی - مقایسه‌ای از نظر عملکردشان در زمینه تعلیم و تربیت قابل ملاحظه هستند. متغیرهایی مانند زمینه‌های خانوادگی، زمینه‌های ژنتیکی، ضربه‌های مغزی و

تجربیات اولیه، هرچند از کنترل علمای تعلیم و تربیت خارج هستند، اما متغیرهای بسیار مهمی در زمینه تعلیم و تربیت هستند.

البته در تحقیق علی - مقایسه‌ای احتمال رسیدن به روابط کاذب همواره وجود دارد. توجه به احتمال وجود عامل مشترک، علیت معکوس و متغیرهای مستقل ممکن دیگر به ما کمک می‌کند که نتایج این گونه تحقیقات را با واقع بینی بیشتر ارزیابی کنیم. چندین راهبرد کنترل محدود، به ما کمک می‌کند که از ارتکاب اشتباهات بزرگ، در طرحهای علی - مقایسه‌ای اجتناب کنیم، لکن هیچ یک از این راهبردها نمی‌توانند کاملاً مشکلات موجود در این گونه طرحها را حل کنند. باید همواره در تفسیر نتایج تحقیقات علی - مقایسه‌ای جانب احتیاط را رعایت کنیم.

### // مفاهیم اصلی

تحقیق قبل از واقعه	تحلیل کوواریانس (ANCOVA)
دیگر متغیرهای مستقل	رابطه علی
علیت معکوس	کنترل محدود
گروههای همسان	متغیرهای فعال در مقابل متغیرهای خصیصه‌ای
نتایج کاذب	نمره های تغییر
وارد کردن متغیرهای بیرونی در طرح	ویژگیهای تحقیق علی - مقایسه‌ای
همتاسازی	

### /// تمرینها

- ۱- تفاوت بین متغیرهای فعال و متغیرهای خصیصه‌ای در چیست؟
- ۲- تحت چه شرایطی از روش تحقیق علی - مقایسه‌ای استفاده می‌کنیم؟ نقاط ضعف اساسی طرح علی - مقایسه‌ای چیست؟
- ۳- چه شرایطی لازم است تا بتوانیم نتیجه‌گیری کنیم که  $X$  عامل بروز  $Y$  است؟
- ۴- کدامیک از فرضیات تحقیقاتی زیر به روش آزمایشی و کدام به تحقیق علی - مقایسه‌ای نیاز دارند؟

الف - آن دسته از کودکانی که مادرشان قصه‌ای برای آنها تعریف می‌کند، به مدت طولانی‌تری



آن را در خاطر حفظ می‌کنند، تا کودکانی که همان قصه را از فرد بیگانه‌ای می‌شنوند.

ب - دانش آموزان خلاق کلاس پنجم انگیزه بیشتری نسبت به دانش‌آموزان غیر خلاق کلاس پنجم دارند.

ج - نوجوانانی که تنها با یکی از والدینشان زندگی می‌کنند، در مقایسه با نوجوانانی که با پدر و مادر خود زندگی می‌کنند، سوابق بیشتری در اداره پلیس دارند.

د - نمرات آزمون هجی کردن دانش آموزان کلاس اولی که هجی کردن را با شیوه آشناسی یاد می‌گیرند، بالاتر از دانش آموزانی است که هجی کردن را با استفاده از روش کل می‌آموزند.

ه - دانشجویان دانشگاه که در آزمون ورودی بعد از لیسانس<sup>(۱)</sup> (GRE) نمره بالاتر از ۱۲۰۰ می‌گیرند، نمراتشان در دانشکده علوم تربیتی بالاتر از دانشجویانی است که نمره پایین‌تر از ۱۲۰۰ می‌گیرند.

و - کودکان معلول نسبت به کودکان سالم دارای سطح پایین‌تری از خودپنداره هستند.

۵- هریک از واژه‌های زیر را تعریف کنید و برای هریک مثالی بزنید.

الف - عامل مشترک

ب - علیت معکوس

ج - دیگر متغیرهای مستقل

۶- به کار بردن نمره‌های تغییر چه نواقصی را جبران می‌کند و چه نواقصی را جبران نمی‌کند؟

۷- نقاط قوت و ضعف هم‌تاسازی چیست؟

۸- محقق با یک متغیر بیرونی مرتبط، که از طریق هم‌تاسازی و دیگر روشها قابل کنترل

نیست، چه باید بکند؟

### /// پاسخها

۱- متغیر خصیصه‌ای، مشخصه‌ای است که آزمودنی قبل از بررسی دارد و در نتیجه، مستقیماً

قابل دستکاری نیست. متغیر فعال نیز یک متغیر مستقل است، ولی محقق می‌تواند آن را

مستقیماً دستکاری کند.

۲- از روش تحقیق علی - مقایسه‌ای در مواردی استفاده می‌شود که متغیرهای خصیصه‌ای

وجود دارند، یعنی آزمودنیها قبل از شروع بررسی دارای آن متغیر مستقل هستند و در نتیجه این متغیرها قابل دستکاری نیستند. مهمترین محدودیت طرحهای علی - مقایسه‌ای، کنترل است. از آنجا که تخصیص تصادفی و دستکاری متغیرهای مستقل امکان پذیر نیست، متغیرهای کنترل نشده می‌توانند عامل به وجود آمدن تغییر در متغیر وابسته باشند.

۳- الف - باید یک رابطه آماری بین X و Y برقرار شود.

ب - X باید از نظر زمانی جلوتر از Y باشد.

ج - عوامل دیگر، تعیین کننده Y نباشند.

۴- الف - آزمایشی

ب - علی - مقایسه‌ای

ج - علی - مقایسه‌ای

د - آزمایشی

ه - علی - مقایسه‌ای

و - علی - مقایسه‌ای

۵- عامل مشترک بدین معناست که اگر متغیرهای A و B با هم مرتبط هستند، این امکان وجود دارد که هیچ یک از آنها علت دیگری نباشد، بلکه هر دوی آنها معلول متغیر سومی (C) باشند. برای مثال، چنانچه مشاهده شود که نمرات بالا در آزمون پیشرفت و نمرات بالای تحصیلی با هم رابطه دارند، این احتمال وجود دارد که هر دو آنها معلول عامل سومی، یعنی هوش عمومی باشند.

علیت معکوس بدین معناست که عکس فرضیه پیشنهاد شده در مورد رابطه مورد نظر صادق باشد. برای مثال، محقق ممکن است مشاهده کند که رابطه‌ای بین از دست دادن کار و نوشیدن نوشابه‌های الکلی به حد افراط، وجود دارد و فرضیه‌اش را چنین تبیین کند که نوشیدن نوشابه‌های الکلی به حد افراط باعث می‌شود که یک کارگر کارش را از دست بدهد، در حالی که در واقع چنین باشد که از دست دادن شغل عامل روی آوردن به الکل باشد.

دیگر متغیرهای مستقل عبارتند از متغیرهایی که سوای متغیر مشاهده شده، عامل رابطه موجود باشند. برای مثال، چنین گفته می‌شود که نسبت جمعیت مراجعه کننده به بیمارستانهای روانی برای درمان اسکیزوفرنی در آمریکا بیشتر از انگلستان است، در حالی که نسبت جمعیت

مبتلا به افسردگی در انگلستان بیشتر از آمریکا است. این، لزوماً بدین معنا نیست که شرایط محیطی آمریکا زمینه‌ساز ابتلا به اسکیزوفرنی و شرایط محیطی انگلستان زمینه‌ساز ابتلا به افسردگی است. ممکن است انگلیسی‌ها در برابر عوارض بیماری اسکیزوفرنی تحمل پذیرتر باشند و احتمالاً با چنین عوارضی کمتر در بیمارستانها بستری می‌شوند، در حالی که آمریکایی‌ها در برابر بروز عوارض افسردگی بردبارتر باشند.

۶- نمره‌های تغییر، تفاوت‌های اولیه در بین گروهها را تعدیل می‌کند. این نمرات الگوهای تغییر افتراقی را تعدیل نمی‌کند.

۷- از طریق هم‌تاسازی، گروههای یکسانی را که از نظر متغیر یا متغیرهای خاص قبل از تیمار همانند بوده‌اند، شکل می‌دهند. در صورتی که گروههای هم‌تاسازی شده از جوامع مختلف باشند، رگرسیون آماری می‌تواند نتایج را تحریف کند. همچنین هم‌تاسازی از تعداد آزمودنیهای موجود می‌کاهد.

۸- محقق می‌تواند متغیر بیرونی مرتبط را در یک طرح علی - مقایسه‌ای وارد و از روش تحلیل واریانس استفاده کند. این رویکرد نیاز به تحلیل اثرات تعاملی اصلی دارد.



# فصل یازدهم

## تحقیق توصیفی

### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- تفاوت بین تحقیق توصیفی را با تحقیق آزمایشی و علی-مقایسه‌ای، توضیح دهد.
- ۲- انواع مختلف تحقیقاتی را که زیر عنوان توصیفی طبقه بندی می‌شوند، نام ببرد.
- ۳- موقعیتهایی را که روش طولی در یک بررسی رشد ارجحیت دارد و مواردی که روش مقطعی ارجحیت دارد، تشخیص دهد.
- ۴- روشها و کاربردهای تحلیل روند و بررسیهای پیگیری را توصیف کند.
- ۵- تحلیل اسنادی را توصیف کند و در مورد این که چه زمانی باید از آن استفاده کرد، مثالهایی بیاورد.
- ۶- هدف عمومی تحقیق همبستگی را بیان کند.
- ۷- طرح تحقیق همبستگی را بیان کند.
- ۸- برخی عوامل را که بر روی بزرگی ضریب همبستگی اثر می‌گذارند، تشخیص دهد.
- ۹- فرضیه‌ای را در مورد ضریب همبستگی  $r$ ، آزمون کند.

- ۱۰- دو هدف اصلی تحقیق همبستگی را توضیح دهد.
- ۱۱- دو واژه پیش بین و ملاک را تعریف کند.
- ۱۲- معادله رگرسیون را برای پیش بینی یک متغیر (Y) از متغیر دیگر (X) بنویسد.
- ۱۳- رگرسیون چند متغیری را تعریف کند و مورد استفاده آن را توضیح دهد.
- ۱۴- پدیده انقباض (کاهش) را در رگرسیون چند متغیری توضیح دهد.
- ۱۵- تحلیل افتراقی را تعریف کند.
- ۱۶- تحلیل عاملی را تعریف کند.
- ۱۷- نقش همبستگی را در تحلیل عاملی توضیح دهد.
- ۱۸- همبستگی جزئی را تعریف کند و هدف آن را توضیح دهد.

مطالعات توصیفی جهت کسب اطلاعات در مورد وضعیت کنونی پدیده، طرح ریزی شده است. این تحقیقات به گونه‌ای اجرا می‌شوند که ماهیت یک موقعیت را به صورتی که در زمان بررسی وجود دارد، تعیین کنند. در این نوع تحقیق، همانند تحقیق علی - مقایسه‌ای، ارائه و یا کنترل تیمار، چنانچه در تحقیقات آزمایشی دیده می‌شود، وجود ندارد. بر خلاف تحقیق از نوع علی - مقایسه‌ای، تحقیق توصیفی به سوی آزمون فرضیه هدایت نمی‌شود. هدف این است که با در نظر گرفتن متغیرها یا شرایط، بگویند که «چه وجود دارد». به عنوان مثال، مدیریت مدرسه مایل است بداند که در سال آینده احتمالاً چند دانش آموز کلاس اول در مدرسه او نام نویسی خواهند کرد تا بتواند برای استفاده مؤثرتر از تسهیلات مدرسه و کارکنان آن، برنامه ریزی کند. او نمی‌خواهد نام نویسی دانش آموزان کلاس اول را به عنوان یک متغیر مرتبط با متغیرهای دیگر بررسی کند. به عبارت دیگر، او فرضیه‌ای را آزمون نمی‌کند، بلکه در پی اطلاعاتی است که به تصمیم‌گیری کمک کند.

ساده‌ترین شکل تحقیقات توصیفی به اندازه‌گیری یک متغیر منفرد می‌پردازد. این امر ممکن است از گزارش فراوانی یک رویداد، همانند مثال فوق، بیشتر نباشد. با این همه، اغلب، بیش از یک متغیر وجود دارد و کوشش در جهت روشن کردن آن است که آیا این متغیرها با هم ربط دارند یا خیر. یک محقق ممکن است به بررسی افت تحصیلی در سال اول دانشکده بپردازد و سپس بخواهد این افت تحصیلی را با جنسیت، نژاد، عملکرد دوره دبیرستان و غیره، ربط دهد. محقق می‌کند که به بررسی پیشرفت تحصیلی دانش آموزان دبیرستانی با استفاده از آزمونهای استاندارد شده ریاضیات و همچنین بخش سؤالات ریاضی آزمون استعداد تحصیلی (SAT)، اشتغال دارد، احتمالاً به بررسی تفاوت عملکرد دختران و پسران نیز می‌پردازد.

چندین نوع بررسی وجود دارند که تحت عنوان تحقیق توصیفی طبقه بندی می‌شوند. این تحقیقات عبارتند از: (۱) تحقیقات پیمایشی (رجوع شود به فصل ۱۲)، (۲) مطالعات رشد، (۳) بررسیهای پیگیری، (۴) تحلیل‌های اسنادی، (۵) تحلیل روند و (۶) بررسیهای همبستگی.

### /// مراحل تحقیق توصیفی

یک تحقیق توصیفی به طور خلاصه مراحل زیر را طی می‌کند:

۱- بیان مسئله:

بیان مسئله، متغیرهای دخیل در بررسی را شناسایی می‌کند و مشخص می‌سازد که بررسی صرفاً در پی تعیین وضعیت این متغیرها است یا روابط بین آنها را نیز مورد مطالعه قرار می‌دهد. همانند تحقیقات آزمایشی، در این نوع تحقیق نیز محقق باید کار را با بیان واضح مسئله شروع کند.

#### ۲- شناسایی اطلاعات مورد نیاز برای حل مسئله:

محقق فهرستی از اطلاعات مورد نیاز را تهیه و مشخص می‌کند که این اطلاعات ماهیت کمی یا کیفی خواهند داشت و در چه قالبی ارائه خواهند شد (اعداد، نمرات آزمون، پاسخهای به پرسشنامه ها و مصاحبه ها و غیره).

#### ۳- انتخاب و یا تهیه وسایل گردآوری داده‌ها:

متداولترین وسایل مورد استفاده جهت تحقیقات توصیفی عبارتند از: پرسشنامه ها، مصاحبه ها، آزمونها و انواع مختلف مقیاسها. چنانچه محقق از یک وسیله از پیش موجود استفاده کند، پایایی و اعتبار آن برای اندازه‌گیری متغیرهای مورد نظر و متناسب بودن آن برای جامعه مورد نظر باید مورد بررسی قرار گیرد. با مطالعه بررسی‌های قبلی که در آنها از این وسایل استفاده شده است، می‌توان شناخت بیشتری درباره کیفیت آنها کسب کرد.

چنانچه محقق، خود، وسیله اندازه‌گیری را بسازد، عاقلانه‌تر است که آن را، برای ارزیابی، ابتدا در یک آزمون مقدماتی با گروهی کوچک به کار برد و اصلاحات لازم را در آن به عمل آورد. چنانچه به خاطر دارید، وسیله‌ها به صورت علمی متغیرهای بررسی را تعریف می‌کنند. در نتیجه، قبل از آنکه محقق قدمی به جلو بگذارد، باید مطمئن شود که داده‌هایی که با استفاده از این وسایل به دست می‌آید، در واقع همان اطلاعاتی هستند که برای حل مسئله مورد نیاز است.

#### ۴- انتخاب نمونه:

محقق گروهی را که قرار است در مورد آنها اطلاعات گردآوری شود، تعیین می‌کند. او در مورد مناسب‌ترین شیوه نمونه‌گیری تصمیم می‌گیرد. شناسایی جامعه آماری مورد نظر و انتخاب نمونه‌هایی که معرف آن جامعه باشند، مراحل حیاتی در اجرای یک تحقیق توصیفی هستند. تعمیم‌پذیری نتایج این‌گونه بررسیها، مستقیماً به این بستگی دارد که نمونه انتخاب شده تا چه میزان معرف جامعه مورد نظر است. معمولاً انتخاب تصادفی آزمودنیها ضروری است.

#### ۵- طراحی شیوه گردآوری داده‌ها:



- محقق راهبرد و برنامه زمانبندی معقول دستیابی به نمونه و استفاده از وسایل گردآوری داده ها را طرح ریزی می کند.
- ۶- گردآوری داده ها
- ۷- تحلیل داده ها
- ۸- تهیه گزارش

### // مطالعات رشد

برای مقوله‌ای به نام تعلیم و تربیت، دستیابی به اطلاعات پایایی در زمینه این که کودکان در سنین مختلف به چه مانند هستند، در یک سطح سنی چه تفاوتی با یکدیگر دارند و چگونه رشد و نمو می‌یابند، بسیار اهمیت دارد. آگاهی از رشد فیزیولوژیکی (بدنی)، ذهنی و عاطفی برای بسیاری از سوالات کاربردی و نظریه‌ای مهم است. همانند غالب نظریه‌های جامع یادگیری، استقرار فیزیکی، برنامه درسی و روشهای تدریس، باید ویژگیهای فراگیر را در نظر بگیرند. دو روش مکمل برای بررسی ویژگیهای کودکان و راههایی که از طریق آنها این ویژگیها همراه با رشد تغییر می‌یابند، به وجود آمده‌اند. این روشها معمولاً به عنوان روش طولی و روش مقطعی خوانده می‌شوند.

### / روش طولی

در روش طولی، یک گروه نمونه واحد از آزمودنیها، در یک دوره زمانی طولانی مورد بررسی قرار می‌گیرد. برای مثال، محقق که بر روی رشد مفاهیم کمی، در بین دانش آموزان دبستانی تحقیق می‌کند، از اندازه گیری مهارتهای کمی کلاس اولیها شروع و در سالهای بعد هم مهارتهای آنها را سالیانه اندازه گیری می‌کند. در نتیجه او می‌تواند چگونگی رشد این مهارتها را در یک دوره زمانی، برای این گروه، ارزیابی کند. از آنجا که او با همان افراد سر و کار دارد، عواملی چون توانایی ذاتی ثابت می‌ماند و تفاوتی که او بین دو کلاس مشاهده می‌کند، می‌تواند به عنوان تغییرات مهارتهای مربوط به رشد آزمودنیها توجیه شود.

مک کینی<sup>(۱)</sup>، با استفاده از روش طولی، ویژگیهای رفتاری کودکانی را که ضعف یادگیری

1- McKinney, J. d. (1989). Longitudinal research on the behavioral characteristics of children with learning disabilities. Journal of Learning Disabilities, 22, 141-150.

داشتند، در طول یک دوره ۳ ساله مورد مطالعه قرار داد. این بررسی از کلاسهای اول و دوم شروع و پیشرفت آنها با نمونه‌ای دیگر از کودکانی که توانایی یادگیری متوسط داشتند، مقایسه شد. متعاقب آن یک بررسی پیگیری در زمینه بازده تحصیلی کودکان متعلق به نمونه اولیه مطالعه طولی که در کلاسهای پنجم و ششم به آنها دسترسی بود، صورت گرفت. او دریافت که کودکانی که ضعف یادگیری داشتند، در مقایسه با کودکان همتایشان که از نظر یادگیری در حد متوسط بودند، به طور قابل ملاحظه‌ای در کلاس رفتارهای ناسازگار بیشتری نشان می‌دادند و در نهایت هم، این رفتارها با عدم پیشرفت تحصیلی مداوم مرتبط شد. یافته‌های این بررسی مؤید شواهد به دست آمده از تحقیقات قبلی است، که اختلاف بین اکثر کودکان با ضعف یادگیری، و کودکان دارای توانایی یادگیری در حد متوسط، با بالا رفتن سن، فاحش‌تر می‌شود. در نتیجه، کودکان دارای ضعف یادگیری به گروه‌های فرعی مختلفی تقسیم شدند که در آنها دسته‌ای مشکل توجه به کلاس را داشتند، دسته‌ای دیگر در اداره و کنترل کلاس درس مسئله داشتند، گروهی رفتار حاکی از انزواگزینی - وابستگی را نشان می‌دادند و بالاخره دسته‌ای رفتار طبیعی داشتند هرچند در ابتدای امر تفاوتی از نظر پیشرفت تحصیلی بین گروه‌های طبقه بندی شده از نظر رفتار مشاهده نشد، اما سه سال بعد مشاهده شد که در مقایسه با گروه فرعی انزواطلب و آنهایی که مشکل رفتاری خاصی نداشتند، گروه فرعی که مشکلات توجه و اداره شدن در کلاس را داشتند، عملکردی ضعیف‌تر را نشان دادند.

مک‌کپنی نتیجه گرفت که تمرکز بیشتر بر مداخلات رفتاری در آموزش و پرورش استثنایی و تلاشهای پیشگیرانه در کلاسهای ابتدایی (سالهای اولیه آموزش) ضرورت دارد.

روش طولی، بررسی فشرده افراد را امکان پذیر می‌سازد، زیرا محقق داده‌ها را براساس سطوح متغیر از آزمونهای واحد، گردآوری می‌کند. به هر حال، بررسیهای طولی نیز مشکلاتی عملی ذاتی دارند. برای شروع مطالعه، نیاز به احساس تعهد شدید فرد و یا مؤسسه‌ای وجود دارد که بخواهد وقت، پول و دیگر منابع را برای چندین سال صرف کند تا پروژه به اتمام برسد. چنانچه یک نمونه انتخاب شده ضعیف باشد، کاری برای بهبود آن نمی‌توان کرد. همچنین نمی‌توان بعد از مدتی که از شروع طرح گذشت، متغیرهای طولی جدید برای بررسی را در مطالعه دخالت داد. در تماس بودن با آزمودنیهایی که تغییر مکان می‌دهند، ممکن است بشدت مشکل باشد. در پاره‌ای موارد، ادامه همکاری آزمودنیها برای یک دوره طولانی مشکل بوده است.

/ روش مقطعی

روش مقطعی فاقد بسیاری از مشکلات عملی روش طولی است. این رویکرد به بررسی آزمودنیها در سطوح مختلف سنی، در یک مقطع زمانی می‌پردازد. به عنوان مثال، یک بررسی مقطعی بر روی رشد مهارتهای کمی، نمونه های مختلفی را از هریک از سطوح کلاسی به کار می‌گیرد، آمارهای به دست آمده از نمونه را مقایسه می‌کند و از طریق آنها در مورد رشد کودکان براساس این مهارتها، نتیجه گیری می‌کند.

نقطه ضعف اصلی روش مقطعی در این است که امکان دارد نتایج مربوط به تفاوتهای بین نمونه ها، به صورت مغرضانه ارائه شود. ممکن است محقق تصادفاً گروهی از کلاس اولیها را به عنوان گروه نمونه انتخاب کند که بلوغ بالاتر از متوسط داشته باشند و برعکس گروه نمونه انتخابی کلاس دوم از نظر بلوغ پایین تر از متوسط باشند و در نتیجه اختلاف دو گروه خیلی کمتر از میزان واقعی آن به نظر می‌رسد. به هر حال، معمولاً در بررسیهای مقطعی امکان به دست آوردن نمونه بزرگتر بیشتر از بررسیهای طولی است و همین مزیت دستیابی به نمونه بزرگ، در بسیاری موارد کمبودهای روش مقطعی را جبران می‌کند.

نقطه ضعف دیگر روش مقطعی احتمال وجود متغیرهای نامربوط است که باعث به وجود آوردن اختلاف بین جامعه هایی می‌شود که نمونه گیری از آنها انجام شده است. برای مثال، فرض کنید محققان رشد گنجینه لغات دانش آموزان دبیرستانی را بررسی می‌کنند. احتمالاً دانش آموزان سال آخر نسبت به دانش آموزان سال اول از تنوع کمتری برخوردار خواهند بود، زیرا دانش آموزان ناتوانتر، به نسبت بیشتری از دانش آموزان تواناتر، تمایل به ترک تحصیل در دوره دبیرستان دارند. در نتیجه، مقایسه گنجینه لغات، انعکاسی از این عامل انتخاب و نیز رشد گنجینه لغات در دوره دبیرستان خواهد بود. البته در بررسیهای طولی، به دلایل مختلف، آزمودنیها می‌توانند از دست بروند و در نتیجه قابلیت مقایسه بین داده های سطوح مختلف کاهش می‌یابد. هرچند محقق که داده های طولی را در دست دارد، خواهد دانست که چه آزمودنیهایی از دست رفته‌اند و آنها تا چه حد با بقیه نمونه متفاوت بودند و سپس در تفسیر نتایج خود می‌تواند این تفاوتها را مد نظر داشته باشد.

چنانچه بخواهیم ویژگی کودکان خاصی را در مراحل مختلف بدانیم، روش مقطعی ارجحیت دارد، زیرا در این روش امکان دسترسی به نمونه های بزرگتر بیشتر است. همچنین در موردی که بخواهیم بر روی تفاوت کلاس اولی ها و کلاس ششمی های کنونی بررسی کنیم، این روش از

ارجحیت برخوردار است؛ زیرا با یک روش طولی، کلاس اولیهای ۵ سال قبل را با کلاس ششمی های امروز مقایسه می‌کنیم.

چنانچه بخواهیم تغییر را به خودی خود بررسی کنیم، روش طولی ارجحیت دارد، زیرا همان آزمودنیها را در طول مراحل تکاملشان دنبال می‌کند. مراحل جهشی و یکنواخت رشد، که در یک بررسی مقطعی نمی‌توان آنها را مشاهده کرد، در بررسی طولی مشاهده می‌شوند. به عنوان مثال، نمرات فرضی سن عقلی سه پسر را که سالیانه اندازه‌گیری شده است، در نظر بگیرید.

باب	جو	پل	$\bar{X}$
۶	۶	۶	۶
۶/۵	۸	۶/۵	۷
۷	۸/۵	۸/۵	۸
۹	۹	۹	۹

هر کدام از آنها یک رشد جهشی را در سن متفاوتی داشته‌اند ولی در مورد هر سه نفر آنها نمره میانگین، افزایش پیوسته را نشان می‌دهد.

### // مطالعات پیگیری

مطالعه پیگیری روشی است که به نحوی به روش طولی شباهت دارد. چنین مطالعاتی به بررسی رشد آزمودنی بعد از ارائه تیمار یا شرایط خاص می‌پردازد. از بین معروفترین این نوع بررسیها، می‌توان به بررسی ترمن و همکاران او<sup>(۱)</sup>، اشاره کرد که آزمودنیهای آن بخشی از نمونه اصلی کودکان سرآمد در سالهای ۲۲-۱۹۲۱ بودند. آزمودنیها پس از طی شش سال و سپس در سالهای ۱۹۳۶، ۱۹۴۰ و ۱۹۴۵ مورد مطالعه قرار گرفتند. علاوه بر نتایج دیگر، پژوهشگران دریافته‌اند که این آزمودنیها در بزرگسالی از بهداشت جسمانی و روانی بهتری نسبت به جامعه عمومی برخوردار بودند. تنها ۴ نفر از ۱۴۶۷ آزمودنی سروکارشان به مؤسسات کيفری افتاد.

1- Burks, B. S., D. W. Jensen, and L. M. Terman. (1930). *The Promise of Youth: Follow-up Studies of a Thousand Gifted Children*, vol. 3. Stanford: Stanford University Press. Terman, L. M., and M. H. Oden (1947). *The Gifted Child Grows Up: Twenty - Five Years' Follow-up Studies of a Superior Group*, vol. 4. Stanford: Stanford University Press.

تقریباً ۹۰ درصد از پسرها به دانشگاهها راه یافتند و ۷۰ درصد از این عده فارغ التحصیل شدند، که از بین آنها ۴۰ درصد دارای درجات عالی بودند. ۸۰ درصد دخترها به دانشگاه راه یافتند که ۶۷ درصدشان فارغ التحصیل شدند و از بین این افراد، ۳۲ درصد دارای درجات عالی بودند. به طور کلی کودکان سرآمد، به صورت بزرگسالان موفق و با استعداد درآمدند.

بررسیهای پیگیری معمولاً به منظور ارزشیابی میزان موفقیت برنامه های ویژه اجرا می شوند. عناوین زیر نشان دهنده این گونه بررسیهاست: «خطر خودکشی در بین کودکان معمولی: یک بررسی پیگیری دو ساله» و «افسردگی دوباره رجعت کرده در نوجوانان: یک بررسی پیگیری».

### // تحلیل اسناد

با وجود آنکه اصل تعلیم و تربیت در وهله اول با انسان سر و کار دارد، اما بسیاری از پروژه های تحقیقاتی مفید و جالب در این رشته به اطلاعات حاصله از جستجوی مدارک و اسناد، وابسته بوده اند. روش تحلیل سیستماتیک مطالب نوشته شده، تحلیل اسنادی یا تحلیل محتوا خوانده می شود. محقق برای بررسی موضوعات و مضامین خاصی، مدارک و اسناد را بررسی می کند. برای مثال به بررسی میزان مشکل بودن کتاب، شواهدی دال بر تعصب یا غرض و یا موارد غالب می پردازد. معمولاً اطلاعاتی که به دست می آید، به صورت شمارش فراوانی در طبقه بندیهای مختلف ارائه می شود. ثورندایک<sup>(۱)</sup> از تحلیل اسنادی جهت شناخت متداولترین لغات مورد استفاده در زبان انگلیسی استفاده کرد. کار او وسیله باارزشی برای تهیه کنندگان کتب درسی زبان دبستانها بوده است.

یک نمونه جدید از تحلیل محتوا، بررسی هیچکاک و تامپکینز<sup>(۲)</sup> از جنسیت گزایی در شش مجموعه تازه چاپ از کتابهای قرائت پایه برای مدارس ابتدایی است. این پژوهشگران ۵۵ کتاب را در سطوح کلاس اول، دوم، پنجم و ششم تحلیل کردند. تمام داستانها را خواندند و به شخصیت اصلی کتاب بر حسب مرد، زن و غیره (حیوان، درخت سخنگو و ...) نمره دادند. آنها

1- Thorndike, E. L. (1932). *A Teacher's Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People*. New York: Teachers College Press, Columbia University.

2- Hitchcock, M. E., and G. E. Tompkins (1987). Basal readers: Are they still exist? *Reading Teacher*, 41, 288-292.

همچنین فراوانی مشاغلی را که برای شخصیت‌های اصلی زن داستانها توصیف شده بود، شمارش کردند. آنها نتایج این تحقیق را با نتایج دیگری که در سال ۱۹۸۱ انجام شده بود، مقایسه کردند. محققان نتیجه گرفتند که ناشران در کتابهای قرائت پایه، از جنسیت گرایی کاسته‌اند. مردان در ۱۸ درصد داستانها، زنها در ۱۷ درصد آنها و مقوله‌ای تحت عنوان «سایر» در ۶۵ درصد بقیه موارد، شخصیت اصلی داستان بوده‌اند. از سال ۱۹۸۱ به بعد، درصد شخصیت‌های اصلی مرد کاهش یافته، در حالی که درصد زنانی که شخصیت اصلی داستان بودند، تغییری نکرده است. شخصیت‌های اصلی زن در ۳۷ نقش متفاوت ظاهر شده‌اند که بیش از همه به ترتیب در نقش کودک، دانش آموز، نویسنده، مادربزرگ و هنرمند درآمده‌اند. هیچکاک و تامپکینز چنین نتیجه گیری کردند که به نظر می‌رسد ناشران با خلق شخصیت‌های خنثی (برای مثال حیواناتی که جنسیتشان مشخص نشده است)، از جنسیت گرایی پرهیز می‌کنند.

به عنوان مثال، از دیگر تحلیل محتواهای انجام شده، می‌توان «توصیف کهولت و کهنسالان در مواد آموزشی دوره ابتدایی»، «محتوای ادبی پاسخهای کودکان به شعر» و «تحلیل محتوای کتب دبستانی تاریخ آمریکا» را نام برد. آخرین بررسی بدین منظور طرح ریزی شده بود که چگونگی برخورد باسیاهان و روابط نژادی را در کتابهای تاریخ مشخص کند.

اصولاً تحلیل محتوا نوعی تحقیق کتابخانه‌ای است. این نوع تحقیق نیاز به صرف وقت زیاد و حوصله و توجه به جزئیات دارد. هرچقدر میزان مطالب تحلیل شده بیشتر باشد، اعتبار نتایج بررسی بالاتر است، اما امکان تحلیل تمام کتب و اسناد مربوط، غیرممکن است و در نتیجه دستیابی به نمونه‌ای معرف از مطالب، امری حائز اهمیت می‌باشد.

جهت مطالعه بیشتر در مورد تحلیل اسناد و موارد استفاده آنها، می‌توان به بحثی در این زمینه در کتاب کریپندورف<sup>(۱)</sup> مراجعه کرد.

## // تحلیل روندها

در مطالعه روندها، ابتدایک جامعه آماری مشخص و سپس یک نمونه معرف از آن انتخاب می‌شود و بر روی متغیرهای خاصی اندازه گیری صورت می‌گیرد. بعد، در فواصل زمانی

1- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Beverly Hills: Sage Publications.

مختلف، بار دیگر از همان جامعه نمونه گیری می شود و آن اندازه گیری تکرار می گردد. این طرح امکان شناسایی روندها (تغییرات) را در مورد آن متغیرها فراهم می آورد.

ارزشیابی سراسری پیشرفت آموزشی<sup>(۱)</sup>، برای مطالعه روندهای پیشرفت تحصیلی در زمینه ریاضیات، داده هایی را در سطح کشور گردآوری کرد.<sup>(۲)</sup> این ارزشیابی ها عملکرد دانش آموزان ۱۳ و ۱۷ ساله را در اواخر دهه ۱۹۷۰، در سال ۱۹۸۲ و بار دیگر در سال ۱۹۸۶، اندازه گیری کردند. نتایج بررسی نشان داد که عملکرد دانش آموزان ۱۳ ساله، که در اواخر دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰ بهبود یافته بود، بین سالهای ۱۹۸۲ و ۱۹۸۶ تغییری نکرد. روند کاهشی که مشخصه عملکرد دانش آموزان ۱۷ ساله طی دهه ۱۹۷۰ بود، معکوس شد. آنان بین سالهای ۱۹۸۲ و ۱۹۸۶ پیشرفت قابل ملاحظه ای کردند. همچنین شواهدی دال بر پیشرفت تحصیلی در بین دانش آموزان سیاه و اسپانیولی زبان و کاهش اختلاف بین پیشرفت تحصیلی دانش آموزان دختر و پسر ۱۷ ساله، مشاهده شد.

### // بررسیهای همبستگی

بررسیهای همبستگی با تعیین رابطه(ها)ی موجود بین متغیرها سر و کار دارد. در این روش، محقق در مورد حداقل دو متغیر، با استفاده از یک نمونه یکسان از آزمودنیها، اطلاعاتی را گردآوری و سپس میزان رابطه را با استفاده از ضریب همبستگی تعیین می کند. یک ضریب همبستگی شاخصی برای جهت و مقدار رابطه است.<sup>(۳)</sup> تحقیق همبستگی نوعاً به منظور برقراری رابطه علت و معلولی به کار نمی رود. احتمالاً از فصل ۵ کتاب به خاطر دارید که وجود همبستگی بین  $X$  و  $Y$  می تواند به این معنی باشد که علت  $X$ ،  $Y$  است یا علت  $X$ ،  $Y$  است و یا هر دو متغیر  $X$  و  $Y$  معلول علت دیگری، یعنی متغیر سوم هستند. وجود همبستگی به خودی خود دال بر علیت نیست. روابط علی را با استفاده از تحقیقات آزمایشی و یا علی-مقایسه ای بررسی می کنند.

هدف از تحقیق همبستگی عبارت است از (۱) توصیف روابط موجود بین متغیرها و (۲) استفاده از یک همبستگی معلوم، برای پیش بینی یک متغیر از روی متغیر دیگر.

### / بررسی رابطه ها

1- The National Assessment of Educational Progress (NAEP)

2- silver, E.A., et al. (1988). The fourth NAEP mathematics assessment: Performance trends and results and trends for instructional indicators. Mathematics Teacher, 81, 720-727.

۳- نمونه محاسبه و تفسیر ضرایب همبستگی در فصل ۵ توضیح داده شده است.

یک نوع از بررسیهای همبستگی اساساً جنبهٔ اکتشافی دارد، بدین معنا که محقق سعی دارد الگوی موجود بین دو یا چند متغیر را شناسایی کند. اطلاعات اکتسابی از این گونه بررسیهای همبستگی، بویژه در مواردی که محقق سعی دارد ساختار پیچیده‌ای را بشناسد و یا می‌خواهد بر روی یک پدیدهٔ رفتاری نظریه بسازد، مشمر ثمر است. متغیرهای بررسی باید بر مبنای یک نظریه، یا تحقیق قبلی و یا مشاهدات محقق انتخاب شوند. رویکرد «تیری در تاریکی»<sup>(۱)</sup>، که در آن محقق همبستگی چندین متغیر را با یکدیگر محاسبه می‌کند تا ببیند نتیجه چه می‌شود، توصیه نمی‌شود.

در پاره‌ای بررسیهای همبستگی، محقق ممکن است بتواند در مورد رابطهٔ مورد انتظار، فرضیه بسازد. برای مثال، با استفاده از نظریهٔ پدیدار شناختی، محقق می‌تواند این فرضیه را تبیین کند که بین درک کودکان کلاس اول از خود و پیشرفت در امر خواندن، رابطهٔ مثبتی وجود دارد. در پاره‌ای موارد، محقق فاقد اطلاعات لازم جهت تبیین فرضیه است.

مطالعه‌ای که بورک<sup>(۲)</sup> انجام داده است، کاربرد روش همبستگی را برای درک این که چرا هرچه تعداد دانش آموزان کلاس افزایش یابد، پیشرفت تحصیلی کمتر خواهد شد، نشان می‌دهد. او فرضیه‌اش را چنین تبیین می‌کند که به کارگیری شیوه‌های مختلف تدریس در کلاسهای با تعداد دانش آموزان مختلف منجر به گوناگونی مقدار پیشرفت تحصیلی شده است. اولین اقدام بورک عبارت بود از بررسی همبستگیهای ساده بین تعداد دانش آموزان کلاس و برخی از شیوه‌های تدریس. جدول ۱-۱۱ نشان دهندهٔ متغیرهایی است که با تعداد دانش آموزان کلاس، همبستگی معنی دار داشته‌اند.

جدول ۱-۱۱ نشان می‌دهد که سوالات دانش آموزان در کلاسهای پرتعدادتر بیشتر (معمولاً یا درخواست کمک داشته‌اند، یا توضیح)، و پاسخگویی معلمان به سؤاها در کلاسهای کم تعدادتر، بیشتر بوده است. استفاده از تکالیف درسی، خاص کلاسهای کم تعدادتر بوده است. مقدار زمانی که معلم برای برقراری ارتباط مستقیم با دانش آموز صرف کرده، در کلاسهای کم تعدادتر بیشتر بوده است.

هرچند این بررسی به مسائلی فراتر از آنچه این مثالها نشان می‌دهند، پرداخته است، لکن

1- 'Shotgun' approach

2- Bourke, S. (1966). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal*, 23, 558-571.



می‌توان ملاحظه کرد که چگونه محقق از همبستگیها برای درک یک پدیده پیچیده، که در این مورد تبیین احتمالی رابطه بین تعداد دانش آموزان کلاس و پیشرفت تحصیلی است، یاری گرفته است.

غالباً زمانی که محققان تعدادی روابط دو متغیری را در یک بررسی واحد، مطالعه می‌کنند، این روابط را در یک ماتریس همبستگی خلاصه می‌کنند که ضریب همبستگی احتمالی را بین همه زوجهای ممکن متغیرها نشان دهند.

جدول ۱۱-۲ یک ماتریس همبستگی را در مورد ۲۱ همبستگی فرضی که می‌تواند از اجرای هفت نوع مختلف آزمونهای استعداد با یک گروه از دانش آموزان به دست آید، نشان می‌دهد که در آن همبستگی بین همه زوجهای ممکن متغیرها، محاسبه شده است. توجه داشته باشید که خط تیره (-) برای نشان دادن همبستگی یک متغیر با خودش به کار رفته است که البته برابر ۱ خواهد بود. داده های ماتریس همبستگی می‌تواند برای درک بهتر سازه، که همانا هوش است، به کار رود.

جدول ۱۱-۱ روابط متغیرهای مربوط به شیوه های تدریس با تعداد دانش آموزان کلاس

متغیر	همبستگی با اندازه کلاس
استفاده از تدریس در سطح کل کلاس	۰/۲۴-
تعداد گروههای تشکیل شده در کلاس	۰/۳۳
ارتباط مستقیم معلم با شاگردان	۰/۲۷-
میزان صدای قابل تحمل	۰/۳۴
میزان پاسخگویی معلم به سؤالات	۰/۲۵-
سؤالات دانش آموز	۰/۳۲
منتظر ماندن آموزگار برای پاسخ	۰/۲۰-
ارزیابی تکالیف	۰/۲۸-
مدیریت غیر آکادمیک	۰/۲۹

جدول ۱۱-۲ نمرات استعداد - همبستگیهای داخلی برای دانش آموزان

## دبیرستانی

میزان استعداد تحصیلی	۱	۲	۳	۴	۵	۶	۷
۱- گنجینه واژگان	-	۰/۴۲	۰/۴۰	۰/۶۵	۰/۲۹	۰/۴۵	۰/۷۴
۲- استعداد عمومی	-	۰/۳۴	۰/۲۶	۰/۳۵	۰/۶۶	۰/۲۳	۰/۲۳
۳- استعداد فضایی	-	-	۰/۲۹	۰/۷۴	۰/۸۱	۰/۲۰	۰/۲۰
۴- قیاسها	-	-	-	۰/۱۵	۰/۵۶	۰/۸۵	۰/۸۵
۵- صور نهفته	-	-	-	-	۰/۷۴	۰/۳۵	۰/۳۵
۶- تشابهات	-	-	-	-	-	۰/۳۹	۰/۳۹
۷- مترادفها	-	-	-	-	-	-	-

به طور نوعی، مطالعات همبستگی نیازمند به نمونه‌های خیلی بزرگ نیست. می‌توان فرض کرد که اگر رابطه‌ای وجود داشته باشد، در یک نمونه متوسط مثلاً ۵۰ تا ۱۰۰ نفری هم خود را نشان می‌دهد. خوب است در اینجا به خاطر بیاوریم که تغییر در نمراتی که با هم همبستگی دارند، بر روی اندازه ضریب همبستگی اثر می‌گذارد. یک دامنه محدود از نمرات یک یا دو متغیر، در مقایسه با دامنه‌ای وسیع از نمرات متغیرها، ضریب کوچکتري را نتیجه می‌دهد.

انتخاب و یا تهیه اندازه‌هایی که شاخصهای مناسبی از متغیرهایی باشند که می‌خواهیم همبستگی آنها را محاسبه کنیم، امر مهمی است. بخصوص این امر حائز اهمیت است که اینها وسایلی باشند که از پایایی و اعتبار برای اندازه‌گیری متغیرهای مورد بررسی، برخوردار باشند. اندازه ضریب همبستگی از بسندگی وسیله اندازه‌گیری برای مقصود مورد نظر، تأثیر می‌گیرد. برای مثال، یک وسیله که برای آزمودنیها زیاد آسان یا سخت باشد، بین آنها تمیز قایل نمی‌شود و منتج به یک ضریب همبستگی کوچک می‌شود. بررسیهایی که از وسایلی، با پایایی کم و اعتبار مشکوک برخوردار است، احتمالاً نتایج قابل قبول را برای محققان به دست نمی‌دهد.

## / پیش‌بینی

همبستگی، مبنای پیش‌بینی است. چنانچه محقق بداند که همبستگی بین دو متغیر وجود دارد، می‌تواند از یک متغیر، متغیر دیگر را پیش‌بینی کند. از آنجا که می‌دانیم بین هوشبهر و معدل نمرات همبستگی مثبت وجود دارد، می‌توان با درجه‌ای از اطمینان پیش‌بینی کرد که

دانش آموز با هوشبهر بالا، معدل نمرات بالایی خواهد داشت. همبستگی بین دو متغیر باید به صورت قابل توجهی بالا باشد تا از نظر پیش بینی باارزش تلقی شود و البته هرچه همبستگی بالاتر باشد، پیش بینی دقیق تر خواهد بود.

حال با یک دسته داده های بسیار ساده، پیش بینی را نشان می دهیم. شکل ۱۱-۱ نموداری از داده های دارای همبستگی را نشان می دهد. در چنین نمودارهایی پیش بین (X)، همواره بر روی محور افقی، و ملاک (Y)، همواره بر روی محور عمودی ترسیم می شود. در این مثال متغیرهای X و Y کاملاً همبستگی دارند (r=۱). در نتیجه تمامی نقاط دقیقاً بر روی خط قطر قرار دارند و به عنوان خط رگرسیون تلقی می شوند. سپس ما خواهیم توانست با استفاده از این خط بر مبنای مقادیر معین X، مقادیر معین Y را پیش بینی کنیم. برای مثال، ملاحظه می کنیم که برای افرادی که نمره X آنها ۶ است، پیش بینی خواهد شد که در متغیر Y، نمره ۱۲ را به دست آورند. به جای آنکه برای پیش بینی یک نمودار ترسیم کنیم، سهل تر است که یک معادله پیش بینی بسازیم. این معادله یک حالت خاص معادله ریاضی برای خط مستقیم است.

$$Y' = bX + a \quad (11-1)$$

نمره پیش بینی شده برای Y = Y'

شیب خط رگرسیون (مقدار تغییر Y برای هر واحد متغیر در X) = b

مقدار ثابت (نقطه ای که خط، محور Y را قطع می کند؛ مقدار نمره Y وقتی که X = ۰ باشد) = a

در مثال فوق ملاحظه می شود که b (شیب منحنی) برابر ۲ است. برای هر واحد تغییر X، Y

دو برابر تغییر می کند:  $b = \frac{Y_2 - Y_1}{X_2 - X_1}$ . وقتی X از ۳ به ۴ تغییر می کند، Y از ۶ به ۸ تغییر می کند:  $b = \frac{8 - 6}{4 - 3} = 2$

مقدار ثابت a را می توان با بررسی نمودار تعیین کرد. خط، محور Y را در نقطه صفر قطع

می کند. با جایگذاری در معادله  $Y' = bX + a$ ، می توان برای هر مقدار X، Y مربوط به آن را

محاسبه کرد. اگر X، ۵ باشد، پس Y پیش بینی شده ۱۰ خواهد بود، یعنی:

$$Y' = 2 \times 5 + 0 = 10$$

در مثال فوق، بین X و Y همبستگی کامل وجود دارد و در نتیجه تعیین b و a آسان است،

ولی در غالب وضعیتهای پیش بینی، X و Y همبستگی کامل ندارند. حال مثال ساده دیگری را

در شکل ۱۱-۲ نشان می دهیم. چنانچه از شکل ۱۱-۲ا مشاهده می شود، تمام نقاط بر روی یک

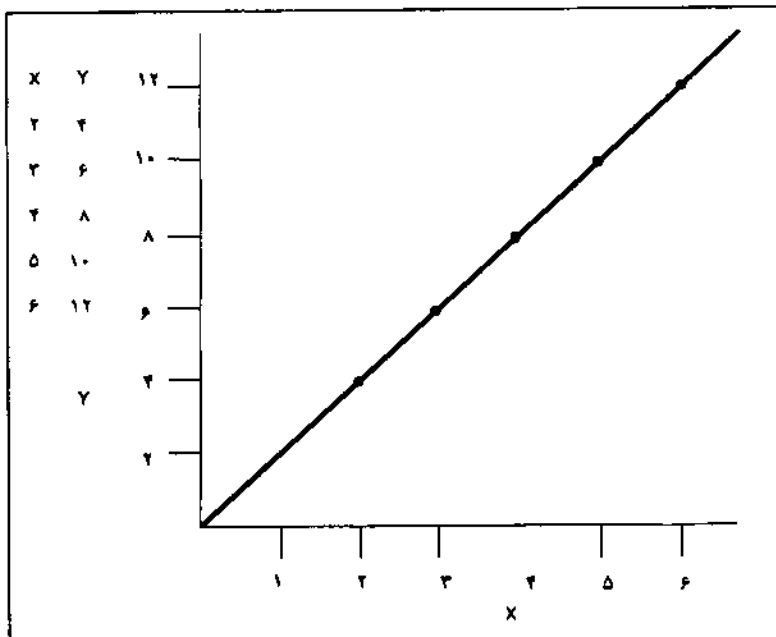
خط مستقیم قرار نمی گیرند، لکن از بین نقاط نشان دهنده بهترین خط مناسب برای این اطلاعات،

می‌توان یک خط رگرسیون رسم کرد. در ابتدای امر باید معادله رگرسیون را بسازیم. مقدار  $a$  و  $b$  در معادله طبق قاعده کوچکترین مربعات تعیین می‌شوند. حاصل این قاعده یک خط رگرسیون است که متوسط مربع انحرافهای نقاط واقعی (نمره های کسب شده) از نمره های پیش بینی شده، بر روی آن خط حداقل مقدار را دارد. در صورتی که از ضریب همبستگی پیرسون ( $r$ ) برای پیش‌بینی استفاده کنیم، شیب خط رگرسیون ( $b$ ) برابر است با ضریب پیرسون ( $r$ ) که در انحراف معیار نمرات پیش‌بینی شده ( $\sigma_y$ ) ضرب شده و بر انحراف استاندارد نمرات پیش‌بینی شده ( $\sigma_x$ ) تقسیم می‌شود.

$$b = r \frac{\sigma_y}{\sigma_x} \quad (11-2)$$

$$b = .784 \left( \frac{1/85}{1/41} \right) = 1/10 \quad \text{در مثال فوق،}$$

شکل ۱۱-۱ داده‌های فرضی نشان دهنده همبستگی کامل



مقدار ثابت (a) عبارت است از میانگین نمرات پیش بینی شده ( $\bar{Y}$ ) منهای حاصلضرب شیب (b) و میانگین پیش‌بین ( $\bar{X}$ ):

$$a = \bar{Y} - \bar{b}X \quad (۱۱-۳)$$

در مثال فوق،

$$a = ۵/۶ - ۱/۱ (۳) = ۲/۳$$

حال می‌توانیم نمرهٔ پیش‌بینی شده ( $\bar{Y}$ ) را برای هر یک از مقادیر X با ضرب کردن شیب (b) در نمرات X و اضافه نمودن ضریب (a) به دست آوریم.

$$Y' = bX + a \quad (۱۱-۴)$$

$$Y' = ۱/۱ X + ۲/۳ \quad \text{در مثال فوق،}$$

اگر X، ۵ باشد

$$Y' = bX + a$$

$$Y' = ۲/۳ + ۱/۱ (۵) = ۷/۸$$

اگر X، ۲ باشد

$$Y' = bX + a$$

$$Y' = ۲/۳ + ۱/۱ (۲) = ۴/۵$$

از آنجا که تمام مقادیر پیش‌بینی شده Y بر خط رگرسیون قرار می‌گیرد، با در نظر گرفتن دو مقدار Y، یعنی ۴/۵ و ۷/۸، می‌توانیم خط رگرسیون را آنچنانکه در شکل ۱۱-۲b نشان داده شده است، رسم کنیم. به خاطر می‌آورید که خط رگرسیون معرف مقادیر پیش‌بینی شده Y برای هر یک از مقادیر X است. می‌توان در شکل ۱۱-۲b مشاهده کرد که عملاً نمرات Y' بر روی خط قرار نمی‌گیرند. این امر همواره در صورتی که همبستگی کامل نباشد، اتفاق می‌افتد.

### / خطای معیار برآورد

اختلاف بین نمرات واقعی Y و نمرات پیش‌بینی شده Y' برای یک مقدار مشخص X، معرف خطای برآورد (e) است، یعنی  $e = Y - Y'$ .

با استفاده از خط رگرسیون، می‌توان برای کسانی که نمرهٔ X آنها ۲ است، نمرهٔ Y برابر با ۴/۵ را پیش‌بینی کرد، ولی مشاهده می‌شود که عملاً نمرهٔ Y، ۵ بوده است. در نتیجه، خطای برآورد عبارت است از  $e = ۵ - ۴/۵ = ۱/۵$ . شاخص خطای متوسط را می‌توان با محاسبه متوسط مربع e حساب کرد. در مورد مثال بالا این شاخص عبارت است از:

$$(4-3/4)^2 + (5-4/5)^2 + (4-5/6)^2 + (6-6/7)^2 + (9-5/8)^2 / 5 = \frac{(5/1)}{5} = 1/0.2$$

با محاسبه جذر این عدد، می‌توان انحراف معیار خطاهای برآورد را محاسبه کرد. این شاخص که خطای معیار برآورد نامیده می‌شود، به عنوان نشانگر میزان دقت پیش بینی به کار می‌رود. فرمول محاسبه را می‌توان به صورت زیر نوشت.

$$\sigma_{est} = \sqrt{\frac{\sum (Y - Y')^2}{n}} \quad (11-5)$$

$\sigma_{est}$  = خطاهای معیار برآورد

$Y$  = نمره‌های عملی  $Y$  در ازای یک نمره مشخص  $X$

$Y'$  = نمره‌های پیش بینی  $Y$  در ازای یک نمره مشخص  $X$

$n$  = تعداد مشاهدات

در مثال فوق این میزان عبارت است از  $1/0.1 = \sqrt{1/0.2}$ . از آنجا که خطای معیار برآورد تابعی از همبستگی بین دو متغیر است، فرمول ساده‌تری برای محاسبه آن وجود دارد که عبارت است از:

$$\sigma_{est} = \sigma_y \sqrt{1-r^2} \quad (11-5a)$$

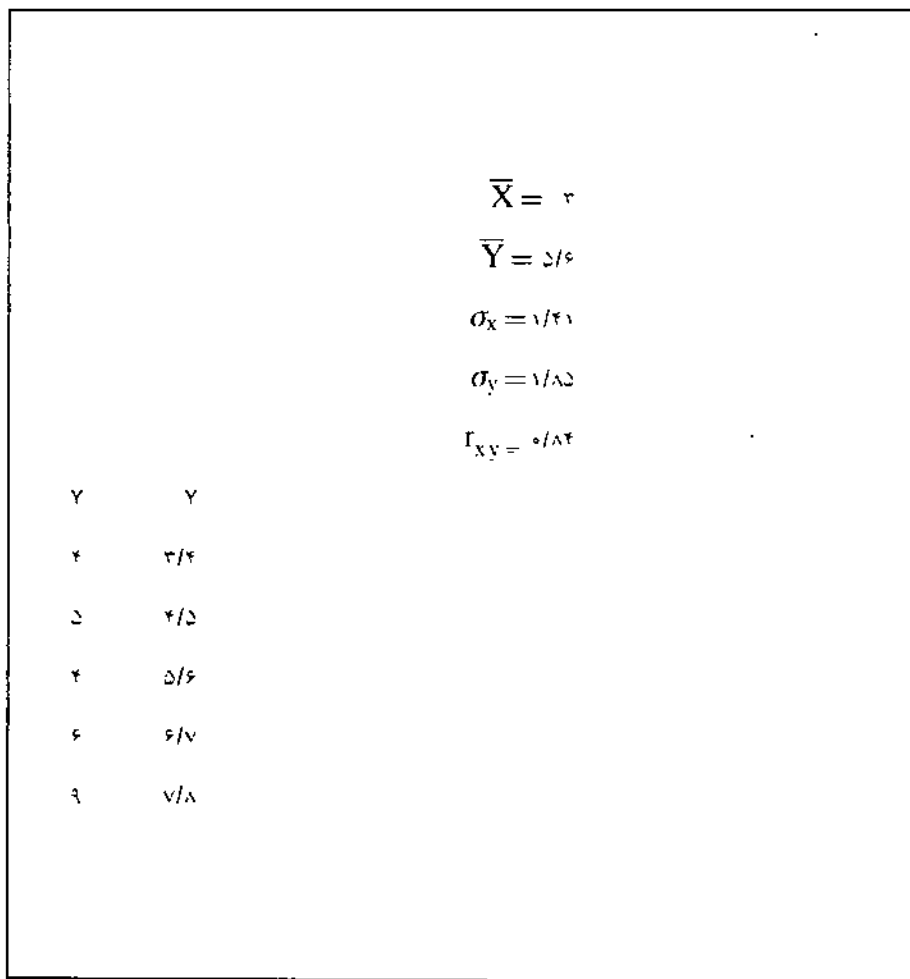
$\sigma_y$  = انحراف معیار نمرات  $Y$

$r^2$  = مربع ضریب همبستگی بین  $X$  و  $Y$

در مثال ما این محاسبه به صورت زیر است:  $1/0.1 = \sqrt{1 - 0.84^2} = 1/0.85$ . حال به مثالی در مورد پیش بینی می‌پردازیم. مدیر گروه مشاوران یک دبیرستان مشاهده کرد که در بین کارمندان، آنتهایی که فاقد اعتماد به نفس بودند، عملکردی ضعیف‌تر از دیگران داشته‌اند. او تصمیم گرفت که میزانی برای عدم اعتماد به نفس بیابد و ببیند که آیا آن میزان می‌تواند به عنوان شناختی برای پیش بینی یک مشاور موفق به کار رود.

او با یاری دیگر مشاوران حرفه‌ای، با شناخت اجزای سازه عدم اعتماد به نفس، بالاخره میزانی برای عدم اعتماد به نفس ساخت که اعتبار و پایایی سازه رضایت بخشی داشت. او این

شکل ۱۱-۲ نمودار پراکندگی داده های فرضی (a) و خط رگرسیون (b)



مقیاس را با نمونه‌ای معرف از مشاوران دبیرستانی به کار برد. او برای هریک از آزمودنیها، میانگین رضایت مراجع را بر روی یک مقیاس ثبت کرد. در این مقیاس نمرات بالاتر، رضایت بیشتری را نشان می‌دهد. این مدیر، ضریب همبستگی پیرسون ۰/۵۰ را بین نمرات مقیاس عدم اعتماد به نفس و رضایت مراجع به دست آورد. جهت ضریب همبستگی به محقق نشان داد که با بالا رفتن نمرات عدم اعتماد به نفس، نمرات رضایت مراجع کاهش می‌یابد. تفاوت بین

همبستگی و صفر نشان‌دهنده قدرت رابطه موجود است.

قدرت همبستگی با ضریب تعیین  $r^2$  نشان داده می‌شود. در این مورد  $r^2$  برابر بود با  $0/25 = 0/5^2$  که نشان دهنده این امر بود که ۲۵ درصد واریانس در نمرات رضایت مراجع با نمرات عدم اعتماد به نفس مشترک بوده است. مدیر می‌تواند با استفاده از نمرات عدم اعتماد به نفس مشاور (X)، نمرات رضایت مراجع را پیش بینی کند. اولین مرحله، محاسبه  $b$  در فرمول پیش بینی است. مقیاس عدم اعتماد به نفس، میانگینی برابر ۲۰ و انحراف معیاری برابر ۴ و نمرات رضایت مراجع میانگینی برابر ۱۶ و انحراف معیاری برابر ۳ داشت. بنابراین:

$$b = r \left( \frac{\sigma_x}{\sigma_y} \right) = -0/5 \left( \frac{3}{4} \right) = -0/375$$

مرحله بعد محاسبه  $a$  است:

$$\begin{aligned} a &= \bar{Y} - \bar{b}X \\ &= 16 - (-0/375) 20 \\ &= 23/5 \end{aligned}$$

بنابراین معادله پیش بینی نمرات Y از نمرات X به صورت زیر خواهد بود:

$$\begin{aligned} Y' &= bX + a \\ &= -0/375X + 23/5 \end{aligned}$$

پیش‌بینی می‌شود مشاوره با نمره مقیاس عدم اعتماد به نفس ۲۴، نمره رضایت مراجع برابر با  $14/5 = 23/5 + (-0/375) 24$  را داشته باشد. برای نمره مقیاس عدم اعتماد به نفس معادل ۱۴، نمره رضایت مراجع،  $(18/25) = 14 + 23/5 = (-0/375)$  پیش‌بینی می‌شود.<sup>(۱)</sup> وسیله‌ای مناسب برای کنترل دقت محاسبه بتا و آلفا عبارت است از محاسبه نمره پیش‌بینی شده Y برای نمره میانگین در X محاسبه صحیح، میانگین Y را به دست می‌دهد. در این مورد،  $16 = 23/5 + (-0/375) 20$ .

۱ - هنگامی که ما با آزمودنی‌های سرکار داریم که داده‌ها را از آنها بدست آورده‌ایم، تمام پیش‌بینی‌ها دقیق هستند. وقتی که این روشها را برای پیش‌بینی در مورد آزمودنی‌هایی که تنها نمره‌های X آنان را در دست داریم، بکار می‌بریم، قابلیت استفاده از این روشها را اعتبار این مفروضه که همبستگی رای آزمودنی‌های جدید همان همبستگی برای آزمودنی‌های اولیه است، تعیین می‌کند. رای مثال، اگر سرپرست مشاوران، از نمره‌های آزمون عدم اعتماد به نفس برای استخدام داوطلبان کار مشاوره استفاده کند، این نمره‌ها تنها در صورتی مفید خواهند بود که همبستگی آنها در بین داوطلبان، مشابه همبستگی ۰/۵ باشد که در بررسی اولیه آمده بود.



از آنجا که همبستگی کامل نیست، برخی از مشاوران، با توجه به نمره مقیاس عدم اعتماد به نفس خود، رضایت مراجع بالاتری را به دست خواهند آورد و برخی دیگر، نمره‌ای پایین تر از آنچه که انتظار می‌رود. خطای معیار برآورد نشان می‌دهد که نمرات واقعی  $Y$  چقدر می‌توانند با نمرات پیش بینی شده برای  $Y$ ، تفاوت داشته باشند. خطای معیار برآورد عبارت است از:

$$.3\sqrt{1-.0/5^2} = 2/06$$

خطای استاندارد برآورد اصولاً یک اندازه تغییر پذیری است و بدین دلیل می‌توان آن را مانند انحراف معیار تفسیر کرد. تقریباً ۶۸ درصد از نمرات واقعی  $Y$  بین  $+1$  تا  $-1$  خطای معیار برآورد  $(2/60)$  و تقریباً ۹۵ درصد بین  $+2$  و  $-2$  خطای معیار برآورد  $(5/20)$  نمرات پیش بینی شده، قرار خواهند گرفت. برای مثال، پیش بینی می‌شود نمره رضایت مراجع مشاوره که نمره مقیاس عدم اعتماد به نفس او ۱۰ است،  $19/75$  باشد. با استفاده از خطای معیار برآورد، انتظار می‌رود ۶۸ درصد از مشاورانی که نمره های رضایت مراجع آنان  $19/75$  پیش بینی می‌شود، نمره‌ای بین  $17/15$  و  $23/35$  را به دست آورند. همچنین انتظار داریم حدود ۹۵ درصد نمره‌ای بین  $\pm 2S_{est}$  یا بین  $14/55$  و  $24/95$  داشته باشند.

#### / رگرسیون چند متغیری

در بسیاری از موقعیتهای زندگی روزمره، بیش از یک متغیر برای پیش بینی یک ملاک، مورد استفاده قرار می‌گیرد. پیش بینی یک ملاک را با استفاده از دو یا چند متغیر پیش بین، رگرسیون چند متغیری می‌نامند. برای مثال، معدل دانشگاه را با استفاده از ترکیبی از متغیرهایی که دانش آموز برای پذیرش به دانشگاه ارائه می‌کند، پیش بینی می‌کنند. معادله رگرسیون مضاعف، شکل گسترده معادله رگرسیون ساده است:

$$Y' = a + b_1X_1 + b_2X_2 + \dots b_nX_n \quad (11-6)$$

$Y'$  = ملاک مورد پیش بینی

$a$  = عدد ثابت

$b_1$  تا  $b_n$  = وزن رگرسیون برای هر پیش بین

$X_1$  تا  $X_n$  = پیش بین‌ها

این شیوه آماری با دادن وزن به هریک از پیش بین ها باعث می شود که متغیرهای پیش بین در ترکیب با هم، به طور بهینه، ملاک Y را پیش بینی کنند. استفاده از رگرسیون چند متغیری را می توان در یک بررسی انجام شده توسط جیکوبز<sup>(۱)</sup>، ملاحظه کرد.

جیکوبز با استفاده از اطلاعات به دست آمده از ۴۱۴۵ دانشجوی سال اول در یکی از دانشگاههای منطقه مرکزی آمریکا، به پیش بینی معدل نمرات نیمسال اول آنها پرداخت. مقادیر پیش بین عبارت بودند از نمرات دانشجویان در آزمونهای فرعی آزمون استعداد تحصیلی (SAT) (آزمون استعداد تحصیلی کلامی و کمی)، بعلاوه رتبه نسبی دانشجویان در دوره دبیرستان (RHSR). رتبه نسبی در دوره دبیرستان با تقسیم رتبه دانش آموز در سال آخر دبیرستان بر تعداد دانش آموزان کلاس به دست آمد و بدین ترتیب مسئله تفاوتی موجود بین تعداد دانش آموزان دبیرستانها حل شد. جدول ۱۱-۳ همبستگیهای ساده بین هر یک از پیش بین ها و ملاک را نشان می دهد.

جدول ۱۱-۳ همبستگیهای هریک از پیش بین ها با ملاک

رتبه نسبی در	آزمون استعداد	آزمون استعداد	
دبیرستان	تحصیلی (کمی)	تحصیلی (کلامی)	
RHSR*	SATM	SATV	
۰/۴۲	۰/۳۹	۰/۳۱	GPA (معدل)

\* دلیل منفی بودن همبستگی بین معدل نمرات و رتبه دبیرستان، نحوه اندازه گیری رتبه در کلاس است. کسی که بالاترین رتبه را در کلاس می آورد، رتبه ۱ می گیرد (پایین ترین نمره) و کسی که پایین ترین رتبه را در کلاس می آورد، رتبه ای معادل تعداد دانش آموزان کلاس که ناعدتاً رتبه بالاتر است، به دست می آورد. از دانش آموزانی که پایین ترین رتبه را داشته اند، انتظار می رود بالاترین نمرات را بیاورند و در نتیجه همبستگی منفی می شود.

برای تعیین همبستگی بهترین ترکیب وزنی سه متغیر پیش بینی با معدل، یک تحلیل رگرسیون چند متغیری صورت گرفت.

همبستگی چندگانه (R) ۰/۵۰ بود. در نتیجه ۲۵ درصد ( $R^2$ ) واریانس معدل نمرات، ناشی

از آزمون کلامی استعداد تحصیلی (SATV)، آزمون کمی استعداد تحصیلی (SATM) و رتبه نسبی در دبیرستان بود. ترکیب متغیرها بالاترین همبستگی را با معدل نمرات داشت و در نتیجه دقیق ترین پیش بینی را ارائه کرد. معادله رگرسیون برای این اطلاعات عبارت بود از:

$$Y'(GPA) = 2/0.813 + (-0/131)RHSR + 0/0014 SATM + 0/0008SATV$$

چنانچه معادله نشان می‌دهد، معدل پیش بینی شده برای هر فرد برابر است با حاصل جمع عدد ثابت (2/0.813) به اضافه حاصلضربهای هر یک از سه پیش بین در وزن رگرسیونی مربوط به خود. فرض کنید دانشجویی این نمرات را به دست آورده است:

$$SATV = 460, SATM = 540, RHSR = 21$$

معدل پیش بینی شده او در دانشگاه 2/93 خواهد بود:

$$Y' = 2/0.813 + (-0/131)(21) + (0/0014)(540) + (0/0008)(460) = 2/93$$

خطای معیار برآورد 0/55 بود. در ساختن رگرسیون چند متغیری و یا معادله پیش‌بینی، در هر حالت باید یک نمره از هر یک از متغیرهای مستقل X (پیش بین‌ها) و متغیر وابسته Y (ملاک) داشت. به هر حال، در آینده معادله را می‌توان در مورد گروهی مشابه، تنها با دانستن متغیرهای مستقل، به کار برد. نمرات آزمون استعداد تحصیلی و رتبه های نسبی را حتی می‌توان برای پیش‌بینی معدل نمرات دانشجوی سال اول، که درخواست پذیرش دانشگاه را دارد، به کار برد. متغیرها باید در مقیاس فاصله‌ای اندازه‌گیری شوند، ولی امکان دارد که متغیرهای مقوله‌ای مثل جنس، طبقه اجتماعی، وضعیت تأهل، ارجحیت سیاسی و از این قبیل را به شرطی که متغیرهای دوتایی ثبت شوند، در معادله رگرسیون قرار دهیم. برای مثال، اگر متغیر جنس را در نظر بگیریم، می‌توان به زن 1 و به مرد صفر داد. در رگرسیون چندمتغیری چنین متغیرهایی را متغیرهای تصنعی می‌گویند.<sup>(۱)</sup>

از آنجا که محاسبات رگرسیون مضاعف بسیار پیچیده است، این گونه محاسبات با کامپیوتر صورت می‌گیرند. برنامه های کامپیوتری موجود نه تنها ضریب همبستگی چندگانه (R) و معادله رگرسیون را محاسبه می‌کنند، بلکه نسبت واریانس ملاک را که ترکیب پیش بین ها (R<sup>2</sup>) بیانگر آن است و آزمونهای معنی دار آماری برای R و برای مشارکت هر یک از پیش بین ها را به

1- J. Cohen and P. Cohen (1983), *Applied Multiple Regression: Correlation Analysis for the Behavioral Sciences* (2d ed), Hilldale, NJ: Erlbaum Associates, for a readable discussion of multiple regression.

ما می‌دهند. برای انتخاب متغیرهای مستقل در معادله رگرسیون چند متغیری، چندین روش وجود دارد. متداولترین روش انتخاب گام به گام<sup>(۱)</sup> خوانده می‌شود. اولین متغیری که مورد بررسی قرار می‌گیرد، متغیری است که با ملاک، بالاترین همبستگی مثبت یا منفی را دارد. به منظور تعیین این امر که آیا در جامعه آماری ما، آن متغیر در پیش بینی ملاک مشارکت مؤثر دارد یا خیر، از آزمون معنی داری  $F^{(۲)}$  استفاده می‌شود.

چنانچه اولین متغیر با ملاک همخوانی داشته باشد، در مرحله بعد متغیری که بیش از همه به پیش بینی می‌افزاید، انتخاب می‌شود. مبنای انتخاب عبارت است از بالاترین همبستگی پس از کنترل آماری همبستگی بین پیش بینی کننده ۱ و ملاک (همبستگی جزئی). سومین متغیری که به بررسی افزوده می‌شود، بالاترین همبستگی بعدی پس از دو پیش بین اول و دوم است که مورد کنترل آماری قرار گرفته‌اند. این انتخاب گام به گام متغیرها تا آنجا ادامه پیدا می‌کند که دیگر هیچ متغیری که با ملاک ها همخوانی داشته باشد، وجود نداشته باشد. در هر مرحله،  $R$  محاسبه می‌شود که به ما می‌گوید آیا متغیر وارد شده به بررسی، به صورت معنی دار به میزان واریانس ملاک که به وسیله متغیرهای قبلی پیش بینی شده است، می‌افزاید یا خیر. ضروری است که محقق از راهنمای بسته نرم افزاری مورد استفاده اش یاری گیرد. برای مثال از SPSS برای توصیف دیگر روشهای انتخاب متغیرها در رگرسیون چند متغیری استفاده کند.

در صورتی که بخواهیم به منظور پیش بینی از معادله رگرسیون چند متغیری استفاده کنیم، باید با نمونه دیگری اعتباریابی کنیم. از آنجا که خطاهای احتمالی در نمونه‌های مختلف، عملکرد متفاوت دارند، وزنه‌های محاسبه شده رگرسیون برای نمونه اصلی، ممکن است برای دیگر نمونه ها یکسان نباشد. بدین معنا که، بهترین پیش بینی کننده برای نمونه اصلی می‌تواند برای یک نمونه دیگر بهترین نباشد. به منظور تعیین اعتبار معادله رگرسیون چند متغیری، باید آن را با به کار بردن در مورد نمونه‌ای دیگر، واری اعتبار کرد. در واری اعتبار معمولاً همبستگی چندگانه عمدتاً پایین تر از نمونه اصلی است که وزنه‌های رگرسیون برای آن محاسبه شده است. این تمایل به کاهش همبستگی چندگانه، که زمانی که تحقیق با نمونه‌ای متفاوت تکرار می‌شود،

#### 1- Stepwise Selection

۲- برای تعیین این که آیا همه متغیرها به حساب آمده‌اند، مقدار  $F$  با یک معیار متداول مقایسه می‌شود. یکی از این معیارها، مقدار حداقل آماره  $F$  است که یک متغیر برای آن که معنادار شود، باید آن را کسب کند. معیار دوم، مشخص کردن احتمال میسبه با آماره  $F$  نظیر  $0.05$  است. یک متغیر، تنها در صورتی وارد مطالعه می‌شود که احتمال همبسته با آزمون  $F$  از سطح معناداری تعیین شده، کمتر یا با آن مساوی باشد.

مشاهده می‌گردد، انقباض خواننده می‌شود. میزان انقباض به اندازه نمونه اصلی و تعداد پیش‌بین‌ها (در عمل نسبت پیش‌بین‌ها به اندازه نمونه) بستگی دارد.<sup>(۱)</sup> هر قدر نمونه‌ای که وزنهای رگرسیون در مورد آنها محاسبه شده است، بزرگتر باشد، انقباض کمتر خواهد بود. توصیه می‌شود، محقق برای دستیابی به R چندگانه پایا، حداقل برای هر یک از پیش‌بین‌ها در معادله رگرسیون، ۳۰ آزمودنی را انتخاب کند؛ اگرچه برخی از دانشمندان عقیده دارند که ۳۰۰ آزمودنی برای هر پیش‌بین، ضروری است. استفاده از رگرسیون چند متغیری تنها به منظور پیش‌بینی صورت نمی‌گیرد و در تحقیقات مختلف کاربرد وسیع دارد. از آنجا که رگرسیون چند متغیری قادر است متغیرهایی را که موجد واریانس در متغیرهای وابسته می‌شوند، شناسایی کند، می‌تواند برای درک محقق از پدیده‌های پیچیده، کاربرد داشته باشد. محقق که می‌خواهد پدیده‌های پیچیده چون پشتکار در دانشگاه را بررسی کند، می‌تواند برای تحلیل تأثیر منفرد و جمعی متغیرهای مستقل در این پدیده، از رگرسیون چند متغیری استفاده کند.

### / تحلیل افتراقی

تحلیل افتراقی شیوه‌ای آماری مربوط به رگرسیون است. در این شیوه با استفاده از تعدادی متغیر، آزمودنیها به دو یا چند گروه متمایز، مانند کسانی که ترک تحصیل می‌کنند در مقابل کسانی که ادامه تحصیل می‌دهند، دانشجویان موفق در مقابل ناموفقها، یزهکاران در مقابل افراد سالم و غیره تقسیم می‌شوند. در تحلیل افتراقی، ملاک، عضویت شخص در یک گروه است. نتیجه این روش یک معادله یا تابع تشخیص است که در آن، نمرات پیش‌بین‌ها در وزنهای هر یک ضرب و بدین وسیله، امکان طبقه‌بندی آزمودنیها در گروهها فراهم می‌شود. در صورتی که تنها دو گروه داشته باشیم، تابع تشخیص اساساً یک معادله رگرسیون چند متغیری با ملاک عضویت گروه است که به صورت ۰ و ۱ رمزگذاری شده است، ولی اگر از نظر ملاک ۳ گروه یا بیشتر داشته باشیم، تحلیل افتراقی فراتر از رگرسیون چند متغیری خواهد بود.

۱- وقتی که N کوچک باشد. توصیه می‌شود برای آن که برآورد بهتری بدست آید، از یک فرمول تصحیح کوتاه شده برای R استفاده شود. این فرمول عبارت است از:

$$R_1^2 = 1 - (1 - R^2) \left( \frac{N-1}{N-n} \right)$$

$R_1^2$  تصحیح شده (کوتاه شده)  $R^2$

N = اندازه نمونه

x = تعداد کل متغیرهای مستقل

واک و پیکو<sup>(۱)</sup> از تحلیل افتراقی برای شناخت پیش بین های موفقیت در دوره دکتری تعلیم و تربیت، استفاده کرده اند. تلاش محققان فوق بر این بود که متغیرهایی را بیابند که مشخص کننده عضویت در یکی از این دو گروه باشند؛ آن دسته از دانشجویانی که با موفقیت دوره دکترا را به اتمام رسانده و گروه دیگر که دوره دکترا را به اتمام نرسانده اند. آنها از تعدادی پیش بینی کننده مانند نمرات آزمون قیاسهای میلر<sup>(۲)</sup> (MAT)، معدل نمرات دوره لیسانس، معدل نمرات دوره فوق لیسانس، فاصله زمانی بین اتمام دوره فوق لیسانس و ورود به دوره دکترا، رشته تخصصی دوره دکترا، سن در بدو ورود به دوره دکترا، وضعیت تأهل و نژاد استفاده کردند. متغیرهایی که پیش بین های موفق در این بررسی بودند، عبارت بودند از نمرات MAT، جنسیت، رشته تخصصی. معادله افتراقی عبارت بود از:

$$(۱۱-۷) \quad Z = C + ۰/۵۷۳۸۲X_{MAT} + ۰/۴۴۸۹۵X_{جنس} + ۰/۶۵۷۲۲X_{رشته\ تخصصی}$$

علامت مثبت ضریب MAT نشان می دهد که دانش آموزانی که در این آزمون نمره بالا گرفته بودند، بیشترین بخت را برای به پایان رساندن دوره دکترا داشتند. از آنجا که نگارندگان مقاله مشخص نکرده اند که متغیر جنسیت چگونه رمزگذاری شده بود، مشخص کردن این امر که آیا مردان بیشتر بخت گذراندن موفقیت آمیز دوره را داشتند یا زنان، با استفاده از معادله، امری غیرممکن است. مؤلفان نتیجه گیری کردند که دانشجویان رشته مدیریت آموزشی بیشترین بخت را برای اتمام دوره دکترا دارند. معادله مشتق شده در مورد شناخت دانشجویان موفق (۹۵/۶ درصد به درستی شناسایی شدند) مؤثرتر از شناسایی دانشجویان ناموفق بوده است؛ نگارندگان معتقدند که تعداد دانشجویان ناموفقی که بتوان تحلیل را بر آن مبنا قرار داد، بسیار کم بوده است. محاسبات پیچیده لازم برای تحلیل افتراقی، نیاز به کامپیوتر دارد. تفسیر نتایج نیز پیچیده است و محقق تازه کار احتمالاً به مشاوره با یک آمارشناس مطلع در این زمینه نیاز دارد.

1- Vacc, N N., and R. Picot (1984), Predicting Success in Doctoral Study. College Student Journal, 18, 113-116

2- Miller Analogies Test (MAT)

## / تحلیل عاملی

یکی دیگر از شیوه‌های آماری که بر مبنای همبستگی استوار است، تحلیل عاملی است. این شیوه به تحلیل همبستگیهای بین مجموعه بزرگی از متغیرها می‌پردازد تا تعداد کمی عوامل مشترک را شناسایی کند. این عوامل عبارتند از سازه‌های فرضی که تصور می‌شود زمینه ساز انواع مختلف اندازه‌های روان‌شناختی از نظر هوش، استعداد، پیشرفت تحصیلی، شخصیت و نگرشها باشند. تحلیل عاملی به آزمونها و دیگر ابزارهای اندازه‌گیری که یک چیز واحد را اندازه‌گیری می‌کنند، اشاره دارد و نیز به این نکته که تا چه حد آن چیز واحد را اندازه می‌گیرند، و از این طریق، موجب می‌شود که محقق با محدودیت‌های کمتری سر و کار داشته باشد. برای مثال، برخی از بررسیهای تحلیل عاملی که بر روی آزمونهای هوش صورت گرفته است، عوامل زیربنایی کلامی، عددی، فضایی، حافظه‌ای و استدلالی هوش را مشخص کرده‌اند.

اولین مرحله در تحلیل عاملی عبارت است از انتخاب متغیرهایی که باید در تحلیل گنجانده شوند و ساختن یک ماتریس همبستگی، که نشان دهنده همبستگی هر اندازه با هر یک از اندازه‌های دیگر باشد. احتمال دارد که تعداد بسیار زیادی همبستگی در ماتریس وجود داشته باشد. سپس ماتریس با استفاده از یک برنامه کامپیوتری تحلیل عاملی مورد محاسبات لازم قرار می‌گیرد و از این طریق مجموعه‌هایی از متغیرها که بین خود همبستگی قوی دارند، ولی با مجموعه متغیرهای دیگر، همبستگی ضعیف دارند، مشخص می‌شوند. این مجموعه متغیرها، عوامل خواننده می‌شوند و هدف بررسی، شناسایی تعداد کمتری از عوامل مجزا است که می‌توانند موجب تغییرات همگام در بین تعداد بیشتری متغیرها شوند. برای مثال، ماتریس همبستگی فرضی زیر را در نظر بگیرید:

۶	۵	۴	۳	۲	۱	
۰/۲۵	۰/۲۲	۰/۲۰	۰/۱۵	۰/۸۰	-	۱- گنجینه وارثان
۰/۲۸	۰/۱۰	۰/۲۵	۰/۱۲	-	-	۲- نیاها
۰/۱۲	۰/۱۵	۰/۷۵	-	-	-	۳- حساب
۰/۲۲	۰/۲۰	-	-	-	-	۴- استدلال عددی
۰/۸۲	-	-	-	-	-	۵- تکمیل تصویر
-	-	-	-	-	-	۶- طراحی با مکعب‌ها

یک تحلیل عاملی «بازنگرانه» سه مجموعه را نشان می‌دهد که در بین خود همبستگی بالا دارند (اندازه‌های ۱ و ۲، ۳ و ۴، ۵ و ۶) و سپس همبستگی ضعیف بین دیگر مجموعه‌ها مشاهده می‌شود. انتظار بر این است که از میان این شش اندازه، سه عامل که عملکرد هوش بر پایه آنها استوار است، یافت شوند. به نظر می‌رسد که این سه عامل عبارت باشند از: کلامی، عددی و فضایی. در یک تحلیل عاملی واقعی، محقق به کامپیوتر ملاکی را می‌دهد که به وسیله آن تعداد نهایی عوامل تعیین می‌شود.

در مرحله بعد، همبستگی بین هریک از متغیرهای اصلی و عوامل مشترک تعیین شده محاسبه می‌شود. این همبستگیها به نام بارهای عامل<sup>(۱)</sup> خوانده می‌شوند. مقدار بارهای عامل نشان دهنده اهمیت در عملکرد هریک از متغیرهاست. محقق امیدوار است که به این نتیجه برسد که هریک از متغیرها، با یک عامل، بار عامل زیاد و یا عوامل دیگر بار عامل کم دارند. این نتیجه نشان دهنده این است که متغیری در آن عامل با متغیرهای دیگر واریانس مشترک دارد، ولی با متغیرهای بار شده بر عوامل دیگر واریانس مشترک ندارد، یا خیلی کم دارد.

مرحله بعدی به نام چرخش عامل<sup>(۲)</sup> خوانده می‌شود. این شیوه در واقع فرایند ساده‌شده‌ای است که برای ارائه یک الگوی روشن‌تر از عوامل و متغیرهایی که بر روی آنها بار شده است، طراحی شده است. در نتیجه، عواملی مستقیماً قابل تفسیر هستند. برنامه کامپیوتری روشهای مناسب چرخش را که هرکدام ملاک سادگی خاص خود را دارند، فراهم می‌کند و این بر عهده استفاده کننده است که در مورد شیوه مورد استفاده تصمیم‌گیری کند. در مرحله بعدی، محقق عوامل حاصله را نام‌گذاری می‌کند. در این مرحله شناسایی متغیرهایی که به صورت معنی‌دار بر روی یک عامل بار می‌شوند، صورت می‌گیرد و سپس نامی از آن منتج می‌شود که به تمام متغیرها اطلاق می‌شود. نتیجه تحلیل می‌تواند یک عامل عددی، یک عامل کلامی یا یک عامل فضایی و غیره باشد. پوسنر و کازس ابزاری را جهت تحلیل عاملی ارائه کردند که به نام پرسشنامه آداب رهبری (LPI)<sup>(۳)</sup> خوانده می‌شود. آنها پاسخهای ۲۸۷۶ نفر آزمودنی را به ۳۰ سؤال پرسشنامه تحلیل کردند و ۵ عاملی را که به وسیله آنها ۳۰ سؤال اندازه‌گیری می‌شد، استخراج کردند. جدول ۴-۱۱ ساختار عاملی را برای LPI نشان می‌دهد. می‌توان مشاهده کرد که

1- Factor Loadings

2- Factor Rotation

3- Posner, B. Z., and J. M. Kouzes (1980). Development and validation of the leadership practices inventory. *Educational and Psychological Measurement*, 48, 483-498.



سؤالهای ۸، ۱۸، ۲۳، ۱۳، ۲۸ و ۳ بارهای عاملی بالایی بر عامل ۱، ولی بارهای پایینی بر دیگر عوامل دارند. به نظر می‌رسد که این موارد، توانایی رهبری را به صورتی که دیگران را وادار به عمل کنند، اندازه‌گیری می‌کنند. از طرف دیگر، سؤالهای ۵، ۲۵، ۱۵، ۲۰، ۱۰ و ۳۰ بارهای عاملی بالایی بر عامل ۲ دارند و ....

شیوه محاسبات تحلیل افتراقی و یا تحلیل عاملی از بحث این گفتار خارج است. برای مطالعه در زمینه‌های فوق، خواننده را به کتابهایی نظیر *مقدمه‌ای بر تحلیل عاملی* (کیم و مولر)<sup>(۱)</sup> و *تحلیل افتراقی* (کلکا)<sup>(۲)</sup> ارجاع می‌دهیم.

---

1- Kim, J. and C.W. Mueler (1978). *Introduction to Factor Analysis: What It Is and How to do it*, Beverly Hills, Sage Publications

2- Kiecka, W.R. (1980). *Discriminant Analysis*, Beverly Hills, Sage Publications

## جدول ۱۱-۴ ساختار عاملی (بارهای عاملی) برای پرسشنامه آداب رهبری

(N=۲۸۷۶)

شماره مورد	عامل ۱ قادر ساختن دیگران به عمل کردن	عامل ۲ تشویق قلبی	عامل ۳ الفا کردن	عامل ۴ مبارزه در عمل	عامل ۵ نمونه سازی
۸	۰/۷۱۹	۰/۱۷۳	۰/۰۹۶	۰/۰۰۸	۰/۰۹۸
۱۸	۰/۶۹۴	۰/۲۰۰	۰/۱۷۶	۰/۰۸۸	۰/۲۱۴
۲۳	۰/۶۸۰	۰/۱۹۸	۰/۱۸۹	۰/۲۳۱	۰/۲۷۳
۱۳	۰/۵۲۶	۰/۱۶۹	۰/۰۹۲	۰/۰۸۵	۰/۰۰۶
۲۸	۰/۵۰۹	۰/۲۸۰	۰/۲۰۶	۰/۱۹۵	۰/۲۹۰
۳	۰/۲۵۹	۰/۲۰۸	۰/۲۳۵	۰/۰۶۹	۰/۲۵۶
۵	۰/۱۱۱	۰/۷۳۱	۰/۲۲۰	۰/۰۹۹	۰/۱۰۹
۲۵	۰/۱۵۲	۰/۷۲۵	۰/۲۵۵	۰/۱۴۳	۰/۱۲۸
۱۵	۰/۴۰۲	۰/۶۸۹	۰/۱۰۲	۰/۱۲۹	۰/۱۱۳
۲۰	۰/۴۵۱	۰/۶۷۳	۰/۱۶۳	۰/۱۴۸	۰/۱۷۲
۱۰	۰/۴۰۰	۰/۶۳۵	۰/۰۷۹	۰/۱۵۴	۰/۱۸۹
۳۰	۰/۲۲۴	۰/۵۳۲	۰/۱۹۴	۰/۲۵۰	۰/۲۴۰
۷	۰/۱۸۵	۰/۲۱۵	۰/۷۰۹	۰/۲۵۱	۰/۱۱۹
۲	۰/۱۵۶	۰/۱۶۵	۰/۶۵۷	۰/۲۷۶	۰/۱۳۶
۲۷	۰/۲۲۳	۰/۲۵۵	۰/۶۲۳	۰/۲۸۴	۰/۲۳۹
۱۷	۰/۱۷۳	۰/۲۲۵	۰/۶۱۵	۰/۲۷۰	۰/۲۴۰
۲۲	۰/۲۲۳	۰/۱۵۱	۰/۵۰۶	۰/۳۶۲	۰/۱۳۶
۱۲	۰/۱۶۶	۰/۱۱۴	۰/۴۸۱	۰/۳۴۵	۰/۱۰۷
۱۶	۰/۱۸۰	۰/۱۶۹	۰/۲۶۶	۰/۳۴۱	۰/۲۳۳
۲۶	۰/۱۶۴	۰/۱۸۵	۰/۲۴۱	۰/۳۲۷	۰/۰۵۷
۱۱	۰/۰۴۳	۰/۰۸۲	۰/۱۸۴	۰/۳۲۲	۰/۱۴۵
۱	۰/۱۸۲	۰/۱۲۸	۰/۲۱۹	۰/۳۴۸	۰/۱۵۳
۲۱	۰/۳۵۴	۰/۱۹۴	۰/۱۷۸	۰/۴۷۳	۰/۱۴۵
۶	۰/۱۷۰	۰/۰۴۹	۰/۱۳۸	۰/۳۹۲	۰/۱۷۳
۲۹	۰/۲۱۸	۰/۱۸۵	۰/۱۴۴	۰/۱۹۲	۰/۶۰۹
۹	۰/۳۴۳	۰/۱۵۸	۰/۰۳۱	۰/۱۰۷	۰/۵۱۲
۱۴	۰/۱۶۴	۰/۱۶۴	۰/۲۳۹	۰/۲۲۸	۰/۵۰۹
۴	۰/۲۳۲	۰/۱۴۲	۰/۳۵۳	۰/۲۳۸	۰/۴۱۱
۱۹	۰/۱۰۹	۰/۱۵۶	۰/۳۲۴	۰/۳۱۵	۰/۴۰۹
۲۴	۰/۳۱۹	۰/۱۲۰	۰/۱۱۵	۰/۲۲۷	۰/۳۷۲

## /// خلاصه

روشهای تحقیق توصیفی، برای به دست آوردن اطلاعات در مورد شرایط موجود، استفاده و در تحقیقات آموزشی به طور وسیعی به کار گرفته می‌شوند. دامنه این روشها از تحقیقات پیمایشی، که وضعیت موجود متغیرهای تعلیم و تربیت را توصیف می‌کنند، تا بررسیهای همبستگی، که روابط بین متغیرها را مطالعه می‌کنند، گسترده شده است. دیگر روشهای توصیفی شامل بررسیهای رشد، بررسی پیگیری، تحلیل اسناد و تحلیل روند می‌باشند. روشهای توصیفی تنها به گردآوری داده‌ها محدود نمی‌شوند. آنها همچنین می‌توانند در مطالعات آزمون فرضیه، به کار گرفته شوند.

بررسیهای همبستگی به محقق اجازه می‌دهد که با استفاده از یک همبستگی مشخص، از یک متغیر، متغیر دیگری را پیش بینی کند. در یک پیش بینی ساده، از یک متغیر  $X$  برای پیش بینی متغیر دیگر  $Y$ ، استفاده می‌شود. در رگرسیون چند متغیری، از بیش از یک متغیر ( $X_1, X_2, \dots, X_p$ ) برای پیش بینی استفاده می‌شود. محقق می‌تواند تعیین کند که دقت پیش بینی چقدر است و چه مقدار از واریانس موجود در  $Y$ ، ناشی از «بهترین» ترکیب خطی پیش بینی‌ها است. شیوه‌های آماری پیچیده‌تر که مبنای آنها همبستگی است، تحلیل افتراقی و تحلیل عاملی می‌باشند.

## // مفاهیم اصلی

بررسی طولی	بررسی پیگیری
بررسی همبستگی	بررسی مقطعی
تحلیل روند	تحلیل افتراقی
تحلیل محتوا	تحلیل عاملی
رگرسیون	خط شیب برآورد
ضریب همبستگی	رگرسیون چند متغیری
ماتریس همبستگی	عامل
مقدار ثابت	متغیر تصنعی (ساختگی)
	وارسی اعتبار

### /// تمرینها

۱- در هر یک از موارد زیر، نوع تحقیق توصیفی مثلاً بررسی رشد، بررسی پیگیری، تحلیل روند را مشخص کنید.

الف - بررسی تومن<sup>(۱)</sup> بر روی بزرگسالانی که در کودکی، استعداد درخشان داشته‌اند.

ب - تحلیل داده‌های گذشته و حال به منظور پیش بینی آینده

۲- در بررسیهای رشد، وقتی که تغییرات نوعی مشخصه‌های عمومی در هر مرحله از رشد کودک مورد نظر محقق نباشد، بلکه محقق بخواهد نحوه تغییرات رشد یک کودک خاص را مطالعه کند، کدامیک از روشهای زیر ارجحیت دارد؟

الف - روش مقطعی

ب - روش طولی

۳- در یک بررسی همبستگی، محقق

الف - تفاوت‌های داخل یک گروه را از نظر یک متغیر با تفاوت‌های داخل همان گروه از نظر متغیر دیگر، مقایسه می‌کند.

ب - کار بررسی را با گروهی شروع می‌کند که از نظر متغیر مورد بررسی اساساً همگن هستند.

ج - عملکرد یک گروه را از نظر یک متغیر با عملکرد گروه قابل مقایسه از نظر همان متغیر، مقایسه می‌کند.

د - تفاوت را در دو گروه ایجاد و سپس آن را اندازه‌گیری می‌کند.

۴- شیوه طولی در بررسیهای رشد از این مزیت برخوردار است که :

الف - بررسی فشرده‌تر فردی است.

ب - در مورد گروههای سنی مختلف در زمان واحد اطلاعاتی را فراهم می‌کند.

ج - گردآوری داده‌ها را موجب می‌شود.

د - خطاهای نمونه‌گیری ندارد.

۵- به منظور محاسبه ضریب همبستگی بین خصیصه‌های A و B، باید:

الف - گروهی از آزمودنیها را داشته باشیم که برخی از آنان دارای ویژگیهای A و بقیه گروه دارای ویژگیهای B باشند.

ب - میزانهایی از خصیصه A را در یک گروه از آزمودنیها و میزانهایی از خصیصه B را در گروه دیگر داشته باشیم.

ج - گروهی آزمودنی داشته باشیم که بعضی از آنان هر دو خصیصه A و B، بعضی دیگر هیچ یک از آنها و گروهی دیگر یکی از دو خصیصه را داشته باشند.

د - دو گروه آزمودنی داشته باشیم که یکی از آنها را به عنوان A و یا غیر A و دیگری را به عنوان B یا غیر B بتوانیم طبقه بندی کنیم.

ه - میزانهایی از خصیصه های A و B را برای هر یک از آزمودنیهای گروه داشته باشیم.

۶- محققى به این نتیجه مى رسد که همبستگی مثبتى بین نمرات هوشبهر و طول فراختای توجه کودکان ده ساله وجود دارد. با این نتیجه مى توان ادعا کرد که:

الف - فراختای توجه طولانی یکی از دلایل هوش است.

ب - هوشبهر بالا یکی از دلایل فراختای توجه طولانی است.

ج - احتمال زیادی وجود دارد که نمونه بزرگی از کودکان ده ساله با هوشبهر بالا، میانگین فراختای توجه کوتاهتری داشته باشند، تا نمونه بزرگی از کودکان ده ساله با هوشبهر پایین.

د - محقق مى تواند پیش بینی کند که فراختای توجه کودکان با هوشبهر بالا طولانی تر از کودکان با هوشبهر پایین باشد.

۷- هریک از عناوین تحقیقی زیر را بیازمایید و تعیین کنید که برای هریک کدامیک از

طرحهای تحقیقی آزمایشی، تحقیق علی - مقایسه‌ای و تحقیق همبستگی مناسب است:

الف - تأثیر طلاق والدین بر انگیزه پیشرفت تحصیلی کودکان

ب - تأثیر یک برنامه خاص آموزش واژگان بر پیشرفت در درس علوم اجتماعی

ج - رابطه بین تعداد دانش آموزان کلاس با رضایت دانش آموزان از کیفیت آموزش

د - تأثیر آموزش آوایی بر سطح نمرات قرائت دانش آموزان کلاس چهارم

۸- محققى در مورد یک نمونه ۱۵۰۰ نفری از دانشجویان سال اول دانشگاه، بین رتبه در

دبیرستان و موفقیت در دانشگاه، ضریب همبستگی معادل ۰/۴۲- را به دست آورد. این ضریب را از نظر جهت، اهمیت و درصد سهم واریانس چگونه تعبیر مى کنید؟

۹- مدیر مدرسه‌ای مى خواهد پیشرفت تحصیلی دانش آموزان را در زبان خارجه پیش بینی

کند. او نمرات یک آزمون هوش، یک آزمون استعداد زبان و یک آزمون قرائت را در دسترس دارد

و بعلاوه جنس دانش آموزان را نیز می‌داند. برای پیش بینی پیشرفت تحصیلی در زبان خارجه، چه پیشنهادی به او می‌کنید؟ نام شیوه پیشنهادی شما چه می‌تواند باشد؟

۱۰- توضیح دهید چگونه می‌توان اندازه‌های دو متغیر، با روشی نسبتاً نظام‌مند به یکدیگر مرتبط شوند، بدون آنکه رابطه علت و معلولی بین آن متغیرها وجود داشته باشد؟

۱۱- محقق اعتبار سازه پرسشنامه‌ای را که برای اندازه‌گیری فشار روانی (تنش) معلم طراحی شده است، بررسی می‌کند. او می‌خواهد بداند که آیا پرسشنامه یک سازه را اندازه‌گیری می‌کند یا چندین سازه را. برای پاسخ به این سؤال چه شیوه‌ای به محقق پیشنهاد می‌کنید؟

۱۲- محقق می‌خواست موفقیت در سال اول دانشکده حقوق را پیش بینی کند. او با نمونه‌ای مشتمل بر ۱۰۰۰ نفر تحلیل رگرسیون چند متغیری را اجرا کرد و از ۶ متغیر پیش بین، معدل نمرات دوره لیسانس، رشته تحصیلی دوره لیسانس، نمره LSAT، نمره یک آزمون کتبی، فاصله زمانی بین فارغ التحصیلی دوره لیسانس و پذیرش در دانشکده حقوق، استفاده کرد. اگر R به دست آمده برابر با ۰/۲ بشود که در سطح ۰/۰۱ معنی دار است، یافته‌های محقق را چگونه تفسیر می‌کنید؟

۱۳- فرض کنید که بین ساعات درس خواندن در روز (X) و معدل نمرات نیمسال اول دانشجویان سال اول (Y) برای یک گروه ۱۰۰ نفری، همبستگی معادل ۰/۹۴+ وجود داشته باشد. با داشتن اطلاعات زیر، معدل دانشجویی را که ۶ ساعت در روز درس می‌خواند، پیش بینی کنید.

$$\bar{X} = 5/74 \quad S_X = 2/31 \quad \bar{Y} = 2/75 \quad S_Y = 0/9$$

/// پاسخها

۱- الف - بررسی پیگیری

ب - تحلیل روند

۲- ب

۳- الف

۴- الف

۵- هـ

۷- الف - تحقیق علی - مقایسه‌ای

ب - تحقیق آزمایشی

ج - تحقیق همبستگی

د - تحقیق آزمایشی

۸-  $r = -0/42$  همبستگی منفی را بین رتبه کلاسی دوره دبیرستان و موفقیت در دانشگاه، نشان می‌دهد، یعنی دانش آموزانی که رتبه عددی پایین دارند، احتمالاً موفقیت بیشتری خواهند داشت. همبستگی منفی ناشی از نحوه رتبه بندی در کلاس است. موفق‌ترین دانش آموز کلاس، رتبه‌ای برابر ۱ (پایین‌ترین عدد) و ناموفق‌ترین دانش آموز کلاس رتبه‌ای معادل با تعداد دانش آموزان کلاس (بالاترین عدد) را دارد. از دانش‌آموزی که در کلاس ۴۰۰ نفری رتبه ۱ دارد، انتظار می‌رود معدل نمرات کالج بالاتری از دانش‌آموز رتبه چهارصدم داشته باشد.

$r = -0/42$  نشان می‌دهد که ۱۸ درصد واریانس نمرات دانشگاه را می‌توان به موفقیت در دبیرستان، که با رتبه مشخص می‌شود، نسبت داد.

۹- مدیر مدرسه، نمونه‌ای از دانش‌آموزان را انتخاب می‌کند و نمرات آنها را در سه آزمون و همچنین نمرات زبان خارجه آنها را به دست می‌آورد.

جنسیت با شماره رمزهای ۰ و ۱ مشخص می‌شود. داده‌ها به صورت آزمون هوش، آزمون استعداد زبان، آزمون قرائت و جنسیت، به عنوان متغیرهای مستقل، و نمرات زبان خارجه، به عنوان متغیر وابسته، در تحلیل رگرسیون چند متغیری وارد می‌شود. به محض آنکه معادله رگرسیون چند متغیری ساخته شود، می‌توان از آن برای پیش‌بینی موفقیت در زبان خارجه برای گروه‌های مشابه دانش‌آموزان، تنها زمانی که متغیرهای مستقل شناخته شده‌اند، استفاده کرد.

۱۰- تنها با همبسته بودن دو متغیر، نمی‌توان نتیجه‌گیری کرد که یکی معلول دیگری است. همبستگی، نشان دهنده وجود رابطه علت و معلولی نیست. احتمال دارد که متغیر سومی که توسط محقق اندازه‌گیری نشده است، عامل رابطه مشاهده شده باشد.

۱۱- محقق باید در مورد پرسشنامه، تحلیل عاملی انجام دهد.

۱۲- با یک نمونه بزرگ، تعجب‌آور نیست که ضریبی به کوچکی  $0/2$  از نظر آماری معنی‌دار شود. با این همه، برای آنکه برای پیش‌بینی مورد استفاده قرار گیرد، به اندازه کافی بزرگ نیست.

$R^2 = 0/04$ ؛ تنها ۴ درصد واریانس موفقیت در دانشکده حقوق را می توان با استفاده از این

ترکیب متغیرهای پیش بین، پیش بینی کرد.

$$b = 0/37 \quad -13$$

$$a = 0/63$$

$$Y' = a + bX$$

$$Y' = 0/63 + 0/37(6)$$

$$Y' = 2/58 \text{ (معدل نمرات پیش بینی شده)}$$



## فصل دوازدهم

### تحقیقات پیمایشی

#### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- هدف تحقیق پیمایشی را بیان کند.
- ۲- طبقات چهارگانه تحقیقات پیمایشی را که برحسب حیطه و موضوع طبقه‌بندی شده‌اند، توصیف کند.
- ۳- مراحل اجرای یک تحقیق پیمایشی را فهرست کند.
- ۴- اهمیت نمونه‌گیری احتمالی را در تحقیق پیمایشی توضیح دهد.
- ۵- از داده‌های نمونه برای تخمین حدود اطمینان پارامتر جمعیت استفاده کند.
- ۶- نقاط قوت و ضعف مصاحبه را به عنوان یک روش گردآوری داده‌ها بیان کند.
- ۷- سوالات باز و بسته را طرح کند.
- ۸- دستور العمل اجرای یک مصاحبه را فهرست کند.
- ۹- نقاط ضعف و قوت استفاده از پرسشنامه را به عنوان شیوه گردآوری داده‌ها بیان کند.
- ۱۰- ده قانون نوشتن سوالات یک پرسشنامه را بیان کند.

- ۱۱- دستور العملهایی را که محقق باید در ارتباط با چهارچوب یک پرسشنامه رعایت کند، فهرست کند.
- ۱۲- نقاط قوت پیش آزمون کردن یک پرسشنامه را توضیح دهد.
- ۱۳- شیوه های پیگیری را که فرد در تحقیق پیمایشی باید دنبال کند، توضیح دهد.
- ۱۴- دستور العمل پرکردن پرسشنامه را بنویسد.
- ۱۵- به طور خلاصه، روش برخورد با کسانی را که پاسخ نداده اند، پس از اجرای طرح بیان کند.
- ۱۶- در مورد شیوه های سنجش اعتبار و پایایی پرسشنامه ها و مصاحبه ها بحث کند.
- ۱۷- روشهای مناسب تحلیل داده های به دست آمده از تحقیقات پیمایشی را توضیح دهد.
- ۱۸- توضیح دهد چگونه محقق می تواند متغیرها را کنترل کند.

پیمایش، روش تحقیقی با اهمیت و بسیار قابل استفاده برای تحقیقات در زمینه‌های روان شناختی، بازرگانی، علوم سیاسی و دولت و همچنین تعلیم و تربیت است. در طول ۵۰ سال اخیر، میزان عناوین و همچنین روشهای مورد استفاده تحقیقات پیمایشی، به صورت قابل ملاحظه‌ای افزایش یافته است. به ندرت پیش می‌آید که در طول یک هفته، از طریق رسانه های خبری در معرض نتایج به دست آمده از نوعی تحقیق پیمایشی قرار نگیریم. در تحقیقات پیمایشی به منظور کشف میزان وقوع و پراکندگی متغیرهای جامعه شناختی، روان شناختی و آموزشی و روابط بین آنها، از جامعه های آماری نمونه گیری می‌شود.<sup>(۱)</sup> داده های به دست آمده از یک تحقیق پیمایشی، عبارت است از پاسخهایی که در جواب به سؤالهای از قبل تعیین شده، از یک نمونه از پاسخگویان دریافت می‌شود. با این همه، محقق مایل است که یافته هایش را به کل گروهی که نمونه از آن انتخاب شده است (جامعه آماری)، تعمیم دهد.

تحقیقات پیمایشی را می‌توان بر مبنای هدفشان به دو گروه گسترده طبقه بندی کرد:  
توضیحی و تبیینی. تحقیقات پیمایشی توصیفی بر تعیین وضعیت یک جامعه آماری مشخص از نظر متغیرهای خاص تأکید دارد. این تحقیقات اصولاً در مورد وضعیت موجود بررسی می‌کنند و در واقع سعی دارند آنچه را که موجود است، اندازه بگیرند، بدون اینکه در مورد چرایی آن سؤال کنند. برخی از سؤالهای نوعی پیمایشی توصیفی عبارتند از: چه درصدی از جوانان آمریکایی از دبیرستان فارغ التحصیل می‌شوند؟ چه درصدی از فارغ التحصیلان دبیرستانی در ایالت ایندیانا به دوره لیسانس وارد می‌شوند؟ کتابخانه های دبیرستانها، به نسبت هر دانش آموز چند کتاب دارند؟ طی سال گذشته، چند مدرسه در کشور در مورد ایدز برنامه‌های آموزشی داشته‌اند؟

تحقیق پیمایشی تبیینی، شکلی از تحقیق علی-مقایسه‌ای است. این نوع تحقیق از توصیف صرف متغیرها فراتر می‌رود. آنها سعی بر آن دارند که تعیین کنند آیا متغیرهای مورد نظر، مرتبط با هم تغییر می‌کنند و یا تحت چه شرایطی با هم تغییر می‌یابند. تحقیق پیمایشی تبیینی سعی بر آن دارد که نگرشها و رفتارها را بر مبنای داده های به دست آمده در یک مقطع زمانی توضیح دهد. تحقیقات پیمایشی تبیینی معمولاً با فرضیه‌ای که گردآوری داده ها را هدایت می‌کند، شروع

1- Kerlinger, F. N (1979). *Behavioral Research: A Conceptual Approach* (p. 151). New York: Holt, Rinehart and Winston.

می‌شود. برای مثال، محققى که به بررسی خودکشى در بین دانشجویان علاقه‌مند است، ممکن است معتقد باشد که وابستگى به گروه، عاملی حائز اهمیت در رویداد خودکشى است. او فرضیه‌اش را چنین تبیین می‌کند که دانشجویانى که وابستگى گروهی دارند، کمتر از دانشجویانى که وابستگى گروهی ندارند، خودکشى می‌کنند. محقق پیمایش را با یک نمونه از دانشگاهها اجرا و داده‌هایی در مورد خودکشى و نوع و میزان وابستگى گروهی قربانیان خودکشى گردآوری می‌کند. چنانچه نتایج پیمایش روشن سازد که در بین دانشجویانى که به گروههای انجمنهای خیریه بانوان و یا خیریه مردان و یا دیگر سازمانهای اجتماعی یا مذهبی دانشگاه وابستگى دارند، کمتر خودکشى اتفاق می‌افتد، محقق ممکن است نتیجه‌گیری کند که فقدان وابستگى گروهی در بین دانشجویان دانشگاهی عاملی مؤثر در خودکشیهای آنان است. در این پیمایش فرضیه وجود رابطه بین وابستگى گروهی و خودکشى تأیید شده است.

همان مشکلات موجود در تفسیر یافته‌های تحقیق علی-مقایسه‌ای در پیمایش تبیینی نیز وجود دارد. قبل از نتیجه‌گیری باید ترتیب زمانی رویدادها و تبیینهای جایگزین در مورد روابط بین متغیرها بررسی شود.

### /// سرشماریها و تحقیقات پیمایشی نمونه‌ای

تحقیقات پیمایشی از نظر حیظه و کانون توجه با هم تفاوت دارند. تحقیق پیمایشی که کل جامعه آماری مورد نظر را در برگیرد، سرشماری خوانده می‌شود، نظیر سرشماریهایی که هر ده سال یک بار توسط دولت آمریکا صورت می‌گیرد. به هر حال در تحقیق، جامعه آماری، کل مردم یک کشور را در بر نمی‌گیرد. مقصود از جامعه آماری، کل گروه افرادی است که یافته‌های یک بررسی در مورد آنها صدق می‌کند. محقق، جامعه آماری خاص مورد بررسی را تعریف می‌کند. اغلب برای محقق بررسی جوامع آماری بسیار بزرگ مشکل و حتی غیر ممکن است. در نتیجه، آنها برای بررسی یک بخش کوچک (یک نمونه) را از جامعه آماری انتخاب می‌کنند. به تحقیق پیمایشی که فقط بخشی از جامعه آماری را مورد مطالعه قرار دهد، تحقیق پیمایشی نمونه‌ای اطلاق می‌شود.

تحقیقات پیمایشی را می‌توان به جدول بندی موضوعهای ملموس، نظیر اینکه چند کودک با اتوبوسهای مدرسه، رفت و آمد می‌کنند، یا تعداد متوسط نام نویسی در کلاسها چقدر است، محدود کرد. مشکلترین نوع تحقیقات پیمایشی، تحقیقاتی است که در آنها محقق در پی

اندازه‌گیری غیر ملموسها، نظیر نگرشها، نظرات، ارزشها و دیگر سازه‌های روان‌شناختی و جامعه‌شناختی است. در چنین تحقیقاتی محقق نه تنها باید مهارت لازم را برای نمونه‌گیری صحیح داشته باشد، بلکه باید از مهارت لازم برای شناختن و ساختن اندازه‌های مناسب و کاربرد نمرات حاصل از این اندازه‌ها در تبیین یا معنای سازه‌های مورد نظر نیز برخوردار باشد. چنانچه تحقیقات پیمایشی را بر مبنای حیطة آنها (سرشماری و یا پیمایش نمونه‌ای) و موضوع آنها (ملموس و غیر ملموس) طبقه‌بندی کنیم، به چهار طبقه دست می‌یابیم: ۱- سرشماری ملموسها ۲- سرشماری غیر ملموسها ۳- تحقیق پیمایشی نمونه‌ای ملموسها ۴- تحقیق پیمایشی نمونه‌ای غیر ملموسها. هر یک از این چهار نوع، مزایا و مشکلات خاص خود را دارد.

### // سرشماری ملموسها

زمانی که فرد به دنبال اطلاعاتی در مورد جامعه‌های کوچک، مانند یک مدرسه، باشد و یا زمانی که متغیرهای مورد بررسی مادی باشند، تلاش کمی برای یافتن سوالات مورد نظر لازم است. اگر مدیر مدرسه بخواهد بداند چند میز در مدرسه‌اش وجود دارد، چند دانش‌آموز از اتوبوس مدرسه استفاده می‌کنند، و یا چند معلم دارای مدرک فوق‌لیسانس هستند، به سادگی آنها را می‌شمارد. از آنجا که او قادر است که جامعه را بررسی کند، می‌تواند تمام ویژگی‌های اطمینان‌را در قیاس کامل داشته باشد. او متغیرهای روشن و خوب تعریف شده را اندازه‌گیری می‌کند، چنانچه شمارش او دقیق و صادقانه باشد، بدون اینکه نگران تضادی باشد، می‌تواند ادعا کند که «در روز اول سپتامبر، تعداد ۶۴۷ میز تحریر، در مدرسه برای کودکان موجود بوده است» یا «۶۵ درصد از مریبان، دارای درجه فوق‌لیسانس هستند». قدرت این نوع سرشماری بستگی به میزان انکار ناپذیری آن دارد. نقطه ضعف آن در منحصر بودن به یک جامعه محدود منفرد و یک مقطع خاص زمانی است. اطلاعات به دست آمده از چنین سرشماریهایی ممکن است اهمیت فوری برای یک گروه محدود داشته باشد، ولی معمولاً چنین پیمایشهایی چیز زیادی به دانش‌تعلیم و تربیت اضافه نمی‌کنند.

### // سرشماری غیر ملموسها

حال فرض کنیم مدیر مدرسه‌ای جویای اطلاعاتی در مورد پیشرفت تحصیلی و یا آرزوهای دانش‌آموزان، روحیهٔ معلمان و یا نگرش والدین نسبت به مدرسه باشد. کار او در این مورد

مشکلتراست؛ زیرا او اکنون با ساختارهایی سر و کار دارد که مستقیماً قابل مشاهده نیستند و باید در مورد آنها از طریق اندازه‌گیریهای غیر مستقیم نتیجه‌گیری کرد. طرح بررسی کلی ارزیابی مدارس (NSSE)<sup>(۱)</sup> یک پرسشنامه نظرخواهی را که برای اندازه‌گیری نگرشها و نظرات نسبت به مدارس طراحی شده است، منتشر می‌سازد. یک پرسشنامه برای دانش‌آموزان، یکی برای معلمان و یک پرسشنامه دیگر برای والدین وجود دارد. ارائه پرسشنامه‌ها به همه دانش‌آموزان، معلمان و یا والدین در نظام آموزشی، مثالی از یک سرشماری غیر ملموس هاست.

به عنوان نمونه ای از این نوع سرشماریها می‌توان از برنامه آزمون پیشرفت تحصیلی که در بیشتر مدارس اجرا می‌شود، نام برد. در این برنامه کلیه کودکان آزمون می‌شوند و عملکرد آنها برای مقایسه با هنجارهای ملی، عملکرد قبلی آنها و غیره، به کار می‌رود. مدیر مدرسه باید نسبت به ماهیت وسایل اندازه‌گیری مورد استفاده و تناسب آنها برای اندازه‌گیری پیشرفت تحصیلی دانش‌آموزان مدرسه خود آگاهی داشته باشد. او باید از خود سؤال کند که آزمونها تا چه حد پایا هستند، آیا همان سازه پیشرفت تحصیلی را که در اهداف مدرسه تعریف شده است، اندازه‌گیری می‌کنند و یا اینکه آن سازه را تا چه حد خوب اندازه‌گیری می‌کنند؟

ارزش سرشماری غیر ملموسها به میزان زیادی بستگی به این دارد که وسیله مورد استفاده واقعاً تا چه حد سازه مورد نظر را اندازه‌گیری می‌کند. وسایل خوبی برای اندازه‌گیری استعداد و پیشرفت تحصیلی در حیطه‌های مختلف علمی در دسترس هستند، اما اندازه‌گیری بسیاری از متغیرهای دیگر با اشکال مواجه می‌شود. در حقیقت به بسیاری از مسائل تعلیم و تربیت به صورت موفقیت آمیز پرداخته نشده است زیرا ما دچار کمبود وسایلی هستیم که به صورت معنی‌دار بتوانند سازه‌های مورد نظر را اندازه‌گیری کنند. متغیرهایی مانند موفقیت معلم، انگیزه دانش‌آموز، سازگاری روانی و رهبری را به سختی توانسته‌اند به طور کاربردی تعریف و اندازه‌گیری کنند.

### // یک تحقیق پیمایشی نمونه‌ای ملموسها

در مواردی که محقق می‌خواهد اطلاعاتی در مورد گروههای بزرگ به دست آورد، مخارج

اجرای سرشماری معمولاً یک عامل بازدارنده است. بنابراین در چنین مواردی از روشهای نمونه گیری استفاده می شود و اطلاعات به دست آمده از نمونه را برای نتیجه گیری در مورد جامعه به صورت یک کل، به کار می برند. دیده ایم که، چنانچه نمونه گیری به خوبی انجام شود، نتیجه گیری در مورد جامعه می تواند کاملاً پایا باشد.

یک مثال معروف از تحقیق پیمایشی نمونه ای، گزارش کلمن<sup>(۱)</sup> است. این بررسی پاسخی بود به بخش ۴۰۲ قرارداد حقوق مدنی سال ۱۹۶۴، که به مسئولان تعلیم و تربیت تکلیف می کرد یک تحقیق پیمایشی در زمینه نابرابری فرصتهای آموزشی در بین گروههای مختلف، در ایالات متحده، انجام دهند. نمونه مورد بررسی ۶۰۰۰۰۰ کودک کلاس اول، سوم، ششم، نهم و دوازدهم را در تقریباً ۴۰۰۰ مدرسه در بر گرفت. این مدرسه ها تقریباً معرف تمام مدارس دولتی ایالات متحده تلقی می شدند، هرچند به عمد سهم بیشتری به مدارس که دارای گروههای اقلیت بودند، داده شد.

از داده های حاصله از این بررسی، چنین نتیجه گیری شد که ۶۵ درصد از سیاهان گروه نمونه به مدارس می رفتند که در آنها ۹۰ درصد دانش آموزان سیاهپوست بودند و ۸۰ درصد سفیدپوستان به مدارس می رفتند که ۹۰ درصد از دانش آموزان خود را از میان سفیدپوستان می پذیرفتند. در مقایسه اندازه کلاس، تسهیلات فیزیکی و شایستگی معلم، تفاوت نسبتاً کمی بین مدارس که به گروههای مختلف نژادی و قومی خدمت می کردند، مشاهده شد؛ هرچند این متغیرها در شهرها و بخشهای روستایی و نواحی مختلف جغرافیایی، متفاوت بودند. با توجه به این متغیرها، محرومان شامل کودکان روستایی و کودکان جنوبی، بدون در نظر گرفتن نژاد، شد.

### // یک تحقیق پیمایشی نمونه ای غیرملموسها

نظرخواهیها مثالهایی از این نوع مطالعات هستند. نظریات مردم مستقیماً قابل مشاهده نیستند، بلکه باید از پاسخهایی که آزمودنیها به پرسشنامه ها و مصاحبه ها می دهند، استنباط شوند. انجام نظرخواهی از دهه ۱۹۳۰ شروع شد و به صورت قابل ملاحظه ای رشد کرد. چنین برآورد می شود که در بین سالهای ۱۹۷۱ تا ۱۹۷۶ در ایالات متحده آمریکا ۱۰۰ میلیون تحقیق پیمایشی با استفاده از مصاحبه انجام شده است؛ در ۱۹۸۰، با ۲۸ میلیون نفر تنها از طریق تلفن

مصاحبه شده بود. در مواردی که پاسخگویان مایل بودند قبل از انتخابات الویتهای خود را آزادانه آشکار سازند، مجریان نظرخواهی در استنباط افکار عمومی کاملاً دقیق بودند و توانستند نتایج انتخابات آتی را پیش بینی کنند. این نظرخواهیها مثالهایی عالی از سودمندی آماره های نمونه در برآورد پارامترهای جامعه آماری فراهم آورده‌اند. اگر کسانی که از یک نامزد انتخاباتی پشتیبانی می‌کنند تمایلی برای ابراز نظر خود نداشته باشند و در مقابل طرفداران نامزد انتخاباتی دیگر نظر خود را آزادانه ابراز کنند، اشتباه قابل ملاحظه‌ای در نتایج نظرخواهی ایجاد خواهد شد. برای مثال، مردم مایلند ابراز کنند که بر علیه کسی که سر کار است، رأی خواهند داد تا بر له او. قبل از انتخابات ریاست جمهوری در سال ۱۹۴۸، نتایج چندین نظرخواهی نشان داد که دیوئی<sup>(۱)</sup> بر ترومن<sup>(۲)</sup> که رئیس جمهور بود، پیروز خواهد شد، ولی بسیاری از مردم ابراز داشتند که هنوز در مورد فرد منتخبشان تصمیم گیری نکرده‌اند. ترومن انتخابات را برد. عملاً اغلب کسانی که ابراز داشته بودند هنوز تصمیم نگرفته‌اند، به ترومن رأی دادند. پاسخگویان همچنین تمایلی ندارند که انتخابی را ابراز کنند که نشانی از علاقه شخصی، تعصب و یا عدم آگاهی آنها از مسائل باشد.

اینکه یک فرد چگونه رأی خواهد داد، غیر ملموس است، ولی آنچه بر روی برگه رأی‌گیری نوشته می‌شود، ملموس است. بخش اخبار شبکه های تلویزیونی در مواردی که تنها با چند منطقه کوچک سر و کار داشتند، به خوبی توانستند نتایج آرای انتخاباتی را پیش‌بینی کنند، ولی آنها قادرند که از اندازه های ملموس از یک نمونه (یعنی، آنچه بر روی برگه های رأی‌گیری نوشته شده است) برای پیش‌بینی آرای جامعه آماری استفاده کنند. در نتیجه، خطر اشتباه تنها به طرحهایی محدود می‌شود که می‌خواهند پارامترهای جامعه را از نمونه های آماری برآورد کنند. از طرفی دست‌اندرکاران نظرخواهیها، که بر مبنای گفته مردم در مورد چگونگی رأی‌دادنشان در آینده آرای یک جامعه آماری را برآورد می‌کنند، از این ضعف مضاعف نیز رنج می‌برند که غیر ملموسها را از زمانی اندازه می‌گیرند که این اندازه‌گیری خود به طور ملموس در پای صندوقهای رأی‌گیری نیز در شرف انجام است. تحقیقات پیمایشی غیر ملموسها به این دلیل که اطلاعات گردآوری شده تنها به صورت غیر مستقیم متغیرهای مورد مطالعه ما را اندازه‌گیری می‌کنند، محدود هستند. شدت این محدودیت به چگونگی اندازه‌گیری متغیر غیرملموس از طریق مشاهده، بستگی دارد.



احتمال دارد که در یک تحقیق پیمایشی واحد، در عین حال ملموسها و غیر ملموسها مورد بررسی قرار گیرند. نگارندگان گزارش کلمن از یک سو از دانشجویان خواستند که به پرسشنامه‌ای پاسخ دهند و از سوی دیگر در مورد آنها آزمونهای هوش و پیشرفت تحصیلی را اجرا کردند تا بتوانند در مورد طبقه اجتماعی، توانایی و پیشرفت تحصیلی و همچنین رابطه بین هریک از این متغیرها با هم و رابطه آنها با متغیرهای ملموس نتیجه‌گیری کنند.

### /// شیوه پیمایش

تحقیق پیمایشی مراحل زیر را شامل می‌شود:

۱- برنامه ریزی: تحقیق پیمایشی با سؤال حایز اهمیتی آغاز می‌شود که محقق معتقد است مناسبترین راه پاسخگویی به آن شیوه پیمایشی است. مسئله تحقیق معمولاً به توزیع و یاروابط بین ویژگیهای رفتاری انسانها، آن گونه که در موقعیت طبیعی رخ می‌دهد، می‌پردازد. وجود فرضیه‌ای که به روشنی بیان شده باشد، در این مرحله از نظر پیش برد بررسی بسیار حایز اهمیت است. مروری بر تحقیقات گذشته ما را قادر می‌سازد که ببینیم محققان دیگر به چه اطلاعاتی در مورد این مسئله دست یافته‌اند. در این مرحله ضروری است که در مورد جامعه‌ای که می‌خواهیم اطلاعات را به آن تعمیم دهیم و همچنین روشها و راههایی که باید برای گردآوری داده‌ها انتخاب کنیم، تصمیم‌گیری شود.

۲- نمونه‌گیری: در تحقیقات پیمایشی، نمونه‌گیری اهمیت خاصی دارد. محقق باید در مورد شیوه نمونه‌گیری مورد استفاده و حجم نمونه تحقیق پیمایشی، تصمیم‌گیری کند. چنانچه محقق بخواهد یافته‌های نمونه را به جامعه آماری تعمیم دهد، ضروری است که نمونه انتخاب شده معرف آن جامعه باشد. شیوه نمونه‌گیری که به احتمال زیاد نمونه‌ای معرف را ارائه می‌کند، شکلی از نمونه‌گیری احتمالاتی است (رجوع شود به فصل ۶). نمونه‌گیری احتمالاتی امکان این برآورد را فراهم می‌آورد که نتایج حاصله از نمونه تا چه حد احتمال دارد از مقادیر جامعه آماری انحراف داشته باشند.

۳- ساختن وسیله و یا وسایل اندازه‌گیری: در مرحله بعدی محقق در مورد شکل سؤالات تصمیم‌گیری می‌کند و سؤالهای وسیله اندازه‌گیری را که برای گردآوری داده‌ها از نمونه، مورد استفاده قرار خواهد گرفت، می‌نویسد. از میان شیوه‌های گردآوری داده‌ها، که در تحقیقات

پیمایشی استفاده می‌شود، می‌توان از مصاحبه حضوری، پرسشنامه و مصاحبه تلفنی نام برد.  
 ۴- اجرای تحقیق پیمایشی: این مرحله شامل اجرای مقدماتی وسیله گردآوری داده‌ها جهت حصول اطمینان از کفایت آن در گردآوری داده‌های مورد نظر، آموزش پرسشگران و مصاحبه‌کنندگان، مصاحبه با آزمودنیها و یا توزیع پرسشنامه‌ها در بین آنها و کاوش در مورد دقت داده‌های گردآوری شده است.

۵- پردازش داده‌ها: مرحله نهایی شامل رمزگذاری داده‌ها، تحلیل آماری (که به احتمال زیاد به وسیله کامپیوتر انجام می‌شود)، تفسیر نتایج و گزارش یافته‌ها می‌شود. مثال زیر مراحل روش پیمایشی را نشان می‌دهد:<sup>(۱)</sup>

الف - هدف: انجام تحقیق پیمایشی در مورد نگرشهای دانشجویان دانشگاه در مورد رفتارهای مربوط به قرار ملاقات گذاردن، که با فعالیتهای جنسی مرتبط باشد و بررسی موارد گزارش شده از خشونت و تجاوز جنسی در ملاقاتهای دانشگاهی.

ب - فرضیه: رفتار قرار ملاقات گذاردن مبهم است و احتمال دارد به رفتار خشونت جنسی از جانب دو طرف منجر شود.

ج - نمونه: تحقیق پیمایشی در یک دانشگاه کوچک با ۶۰۰۰ دانشجو در منطقه غرب مرکزی آمریکا صورت گرفت. شرکت کنندگان در بررسی شامل ۴۰۸ دانشجو (تقریباً نمونه‌ای ۷ درصدی)؛ ۲۴۷ زن و ۱۶۱ مرد بودند که در دوره لیسانس رشته روان شناسی تحصیل می‌کردند. پاسخگویان داوطلبانه و بدون ذکر نام، پرسشنامه را در کلاس پر کردند. میانگین سن پاسخگویان ۲۱/۵ سال بود و اکثریت قریب به اتفاق آنها مجرد بودند.

د - وسیله گردآوری داده‌ها: پرسشنامه‌ای شامل ۱۳ سؤال برای نشان دادن نگرش نسبت به رفتارهای جنسی در قرار ملاقات و همچنین تجربیات پاسخگویان در زمینه رفتار خشونت آمیز جنسی، طراحی شد. پاسخ در مقیاسی شش نقطه‌ای از «هرگز» (۱) تا «بارها» (۶)، نمره گذاری شد. هیچگونه اطلاعاتی در مورد اعتبار و پایایی پرسشنامه داده نشده است.

ه - داده‌ها: داده‌ها به صورت توزیع فراوانی برای کل نمونه و برای هریک از دو گروه زن و مرد، ارائه شد. به منظور آزمون معنی داری تفاوت بین پاسخهای جنسهای مختلف، از آزمون t استفاده شد.

1- Sandberg, G.T, Jackson and P. Petretic Jackson (1987), College Student's Attitudes Regarding Sexual Coercion and Aggression: Developing Educational and Preventive Strategies. Journal of College Student Personal 28, 302-311

و - نتایج<sup>(۱)</sup>: ۵۸ درصد از کل نمونه ابراز کرده‌اند که با فردی که قرار ملاقات گذارده‌اند، تماس جنسی داشته‌اند و این عمل طبق خواست آنها نبوده، بلکه به این دلیل بوده است که آنها فکر می‌کرده‌اند که امتناع از آن مناسب نیست. در پاسخ این سؤال، بین پاسخهای زنان و مردان تفاوت معنی داری مشاهده شد. ۵۸ درصد زنان و ۲۵ درصد مردان ابراز داشتند که در یک قرار ملاقات مورد تجاوز جنسی قرار گرفته‌اند. ده درصد زنان و دو درصد مردان ابراز داشته‌اند که در یک قرار ملاقات، از نظر جنسی مورد سوء استفاده قرار گرفته‌اند. این تفاوت بین زنان و مردان، از نظر آماری، معنی دار بود.

ر - نتیجه‌گیری: با در نظر گرفتن یافته‌های پیمایش، فرضیه بررسی تأیید شد. چنین نتیجه‌گیری شد که به علت ماهیت و عواقب سخت خشونت‌های جنسی، کارکنان دانشگاه در سطوح مختلف، باید نسبت به احتمال وقوع تجاوز و دیگر انواع رفتارهای تجاوزگرانه و قهرآمیز در دانشگاه، آگاهی دهند و سیاست‌هایی اجتماعی اتخاذ و برنامه‌ای برای پیشگیری و ارائه خدمات درمانی در محوطه دانشگاه‌ها، تنظیم کنند.

### // انتخاب نمونه

از خوانندگان خواسته می‌شود که بار دیگر به فصل ۶ در مورد شیوه‌های نمونه‌گیری و عوامل مؤثر در حجم نمونه انتخابی، رجوع کنند. در زمینه حجم نمونه بحث بیشتری در بخش بعد خواهیم داشت.

### / حجم نمونه

حجم نمونه از اولین سؤالهایی است که محقق باید به آنها پاسخ دهد. چگونه در تحقیق پیمایشی، در مورد تعداد افراد مورد بررسی تصمیم‌گیری می‌شود؟ بسیاری از محققان بر این باورند که باید حداقل نمونه‌ای برابر ۱۰ درصد جامعه آماری را انتخاب کرد، لکن این امر، ضروری نیست. برخلاف آنچه عموماً فکر می‌کنند، دقت اطلاعات بستگی به مقدار مطلق<sup>(۲)</sup> نمونه و نه درصد نمونه نسبت به جامعه آماری دارد. احتمالاً توجه کرده‌اید که در نظرخواهیهای

۱ - تمامی نتایج این پیمایش در اینجا گزارش نشده است.

عمومی که در ایالات متحده آمریکا انجام می‌شود، از تعداد افراد زیادی استفاده نمی‌شود. برای مثال، نظرخواهی بخش خبر CBS<sup>(۱)</sup> از رأی‌دهندگان ثبت نام شده، که قبیل از انتخابات ریاست جمهوری ۱۹۸۸ انجام گرفت و از پاسخگویان خواسته شد تا کاندیدای مورد علاقه‌شان را معرفی کنند، تنها با نمونه‌ای برابر ۱۳۴۳ نفر اجرا شد که نسبت به حجم جامعه آماری، بسیار کوچک بود.

مسئله اساسی در تصمیم‌گیری در مورد تعداد افراد نمونه عبارت از درجه دقت موردنظر محقق در برآورد مقادیر جامعه آماری است. محقق تا چه میزان خطا را در تصمیم‌دادن آماده به پارامتر جامعه آماری، می‌پذیرد؟ چنانچه محققان از نمونه‌گیری احتمالی استفاده کنند، مبنایی برای برآورد این که احتمالاً نتایج نمونه چقدر از مقادیر جامعه آماری دور است، در دست دارند، یعنی به عبارت دیگر، می‌توانند برآورد کنند که حاشیه خطا<sup>(۲)</sup> برای یک حجم نمونه چقدر است. آنها حجم نمونه‌ای را انتخاب می‌کنند که مطمئن باشند برآوردهایشان در یک دامنه محدود، در ۹۵ درصد اوقات درست خواهد بود. حال این مسئله را نشان می‌دهیم.

### / برآورد مقادیر جامعه آماری

در مورد متغیرهای دو جمله‌ای (تنها با دو مقدار)، توزیع دو جمله‌ای مبنایی را برای برآورد نسبی از جامعه، که دارای برخی ویژگی‌ها هستند، فراهم می‌آورد. برای مثال چه نسبتی از رأی‌دهندگان ثبت شده با کاستن از بودجه دفاعی کشور موافق هستند یا چه نسبتی از رأی‌دهندگان با حفر چاههای نفت در سواحل دریا برای برآوردن نیاز کشور به انرژی موافق هستند؟ برای برآورد نسبتهای جامعه آماری از پاسخهای نمونه استفاده می‌کنیم. در صورتی که  $n$  بزرگ باشد (۵۰ یا بیشتر)، نسبتهای نمونه تقریباً به صورت نرمال با میانگینی برابر  $P$  (نسبت جامعه) و انحراف معیاری که خطای معیار خوانده می‌شود، توزیع می‌شود. فرمول خطای معیار عبارت است از:

$$\text{خطای معیار} = \sqrt{\frac{Pq}{n}}$$

۱ - نظرخواهی از ۱۳۴۳ نفر در مورد اخبار که توسط CBS انجام و گزارش آن در ۴ نوامبر ۱۹۸۸ ارائه شد و نشان داد که دوکاکس (Dukakis) ۷ درصد بیشتر رأی خواهد آورد. حاشیه خطا ۳٪ بود.

$p$  = نسبتی که یک نوع پاسخ را می‌دهند (مثلاً بلی)

$q$  = نسبتی که پاسخ نوع دیگر را می‌دهند (خیر)

$pq$  = واریانس نمونه

$n$  = حجم نمونه

به خاطر دارید که خطای معیار اندازه‌ای است از داده‌های نمونه به عنوان برآورد مقدار جامعه آماری. هر قدر معیار کوچکتر باشد، احتمال اینکه نمونه معرف جامعه آماری باشد، بالاتر است. فرض کنید که تحقیقی پیمایشی را با نمونه‌ای تصادفی از ۵۰۰ رأی دهنده انجام داده‌ایم و از آنها سؤال شده است که آیا با کاهش بودجه دفاعی کشور موافق هستند؟ ۳۷۵ نفر، یا ۷۵ درصد، پاسخ آری و ۱۲۵ نفر، یا ۲۵ درصد، پاسخ خیر داده‌اند. آیا می‌توانیم ۰/۷۵ را به عنوان برآوردی از نسبت واقعی کلیه رأی دهندگانی که با کاهش بودجه دفاعی کشور موافقت می‌کنند، بدانیم؟ میزان مطلوب بودن برآورد را می‌توانیم با تعیین حدود اطمینان برای نسبت واقعی، تعیین کنیم. هر قدر این حدود به ۰/۷۵ نزدیکتر باشد، دقت برآورد بالاتر است. اولین قدم محاسبه خطای معیار است:

$$\text{خطای معیار} = \sqrt{\frac{(0/75)(0/25)}{500}} = 0/194$$

ما از خطای معیار برای تعیین فاصله‌ای که احتمالاً حاوی مقدار واقعی جامعه آماری در درصد معینی از موارد است، استفاده می‌کنیم. این فاصله در سطح اطمینان ۹۵ درصد، که سطحی متداول است، عبارت است از مقدار نمونه بعلاوه یا منهای ۱/۹۶ خطای معیار.

از فصل ۵ به خاطر دارید که ۹۵ درصد ناحیه زیر منحنی بین ۱/۹۶ + انحراف معیار و ۱/۹۶ - انحراف معیار از میانگین قرار می‌گیرد. در این مورد، فاصله برای درصد افرادی از جامعه که با کاهش بودجه دفاعی موافق هستند، برابر خواهد بود با  $0/75 \pm 0/37$  یا  $0/194(0/194) \pm 0/75$ ، یعنی احتمالی برابر ۰/۹۵ وجود دارد که نسبت واقعی در جامعه که موافق اختصاص بودجه کمتری برای مقاصد دفاعی است، بین ۷۱ و ۷۹ درصد ( $0/75 \pm 0/04$ ) باشد. در نتیجه حاشیه خطای این تحقیق پیمایشی ۴ درصد است. چنانچه محقق حاشیه خطای کوچکتری را طلب کند، نمونه بزرگتری مورد نیاز است. از فرمول محاسبه خطای معیار، ملاحظه می‌شود که با

افزایش  $n$  میزان خطای معیار کاهش می‌یابد. چنانچه در تحقیق پیمایشی فوق به جای ۵۰۰ نفر، از ۱۵۰۰ نفر نظرخواهی می‌شد، خطای معیار به صورت زیر کاهش می‌یافت:

$$\sqrt{\frac{(0/75)(0/25)}{1500}} = 0/11$$

بنابر این حدود اطمینان ۹۵ درصد تنها  $[0/75 \pm (1/96 \times 0/11) \pm 0/2]$  بسط می‌یافت و در نتیجه ۷۳ تا ۷۷ به عنوان فاصله اطمینانی که حاوی مقدار جامعه آماری است، به دست می‌آمد. با افزایش  $n$  دقت برآورد افزایش می‌یابد و حاشیه خطا از نقطه ۴ درصد، به نقطه ۲ درصد می‌رسد. برای دستیابی به دقت مورد نظر، تا چه میزان باید به حجم  $n$  افزوده شود؟ فرض کنید که یک مدیر مدرسه مایل است بداند دانش آموزان در مورد داشتن یک برنامه آموزش ایدز در مدرسه چه احساسی دارند. مدیر می‌خواهد که دقت برآوردش در سطح اطمینان ۹۵ درصد، در داخل نقطه ۴ درصدی باشد. از فرمول تعیین فاصله اطمینان  $(1/96 \times s.e. = 0/04)$  می‌توان برای یافتن  $n$  استفاده کرد. تنها مشکل این است که مدیر در فرمول خطای معیار، مقدار  $p$  را در دست ندارد. مدیر می‌تواند یکی از این دو کار را بکند: اول آنکه، می‌تواند یک بررسی مقدماتی را با  $n=25$  به منظور به دست آوردن برآورد تقریبی  $p$  انجام دهد. فرض کنید هفت دانش آموز از ۲۵ دانش آموز به داشتن چنین برنامه‌ای اظهار علاقه کنند  $(p = \frac{7}{25} = 0/28)$ . اگر  $p=0/28$  باشد، پس  $q = 0/72$ .

$$1/96 \sqrt{\frac{pq}{n}} = 0/4$$

مقدار  $n$  را می‌توان به طریق زیر به دست آورد:

$$1/96 \sqrt{\frac{(0/28)(0/72)}{n}} = 0/4$$

اگر هر دو طرف معادله را بر ۱/۹۶ تقسیم کنیم، خواهیم داشت:

$$\sqrt{\frac{(0/28)(0/72)}{n}} = \frac{0/4}{1/96}$$

$$\frac{(0/28)(0/72)}{n} = \left(\frac{0/04}{1/96}\right)^2$$

$$\frac{0/2016}{n} = 0/00041164$$

$$n = 484$$

مدیر به نمونه‌ای برابر ۴۸۴ دانش‌آموز نیاز دارد تا بتواند فاصله اطمینان ۹۵ درصدی  $(0/28 \pm 0/04)$  را بسازد. فرض کنید که مدیر مدرسه تحقیق پیمایشی را با ۴۸۴ دانش‌آموز انجام دهد و نتیجه بگیرد که تنها ۱۲۱ دانش‌آموز خواهان این برنامه هستند، نسبت نمونه در اینجا  $0/25 \left[ \frac{121}{484} \right]$  می‌شود. او می‌تواند  $p = 0/25$  را به کار برد و دوباره  $n$  را محاسبه کند.

$$1/96 \sqrt{\frac{(0/25)(72)}{n}} = 0/04$$

در اینجا  $n=450$  به مدیر مدرسه اطمینان می‌دهد که نمونه اولیه ۴۸۴ دانش‌آموزی در برآورد  $p=0/25$  حتی بیش از کفایت بوده است. مدیر می‌تواند ادعا کند که برای جامعه دانش‌آموزان او  $p = 0/25 \pm 0/04$  است و ۹۵ درصد احتمال وجود دارد که فاصله اطمینان واقعی جامعه آماری را در برگیرد. در نتیجه، مدیر توانست تعداد نمونه را جهت دستیابی به حدود اطمینان و حدود خطای مورد نظر، محاسبه کند. چنانچه در محاسبه دوم این نتیجه به دست می‌آمد که مقدار  $n$  باید افزایش یابد، لازم بود که مدیر مدرسه تعداد دانش‌آموزان شرکت‌کننده در پیمایش را افزایش دهد و محاسبه را دوباره تکرار کند.

راه دیگری که مدیر می‌توانست در پیش گیرد و از اجرای پیمایش مقدماتی نیز اجتناب ورزد، بدین صورت بود که بسادگی فرض کند  $p = 0/50$  است و معادله را حل کند.

$$1/96 \sqrt{\frac{(0/50)(0/50)}{n}} = 0/04$$

بنابراین  $n = 600$  را بدست می‌آورد. به کار بردن  $p = 0/50$  همواره بی‌خطرترین راه برای محاسبه تعداد نمونه است، زیرا  $q = p$  و وقتی  $q = p = 0/50$  باشد، بالاترین مقدار را خواهد داشت و در نتیجه بالاترین برآورد ممکن را برای تعداد نمونه مورد نیاز به دست می‌دهد. سپس، اگر محقق نتیجه بگیرد که  $p$  در عمل بزرگتر یا کوچکتر از  $0/50$  است، با داشتن تعداد نمونه‌ای

زیادتر که با فرض  $p = 0/50$  محاسبه شده است، برآوردی دقیق تر از آنچه مورد انتظار بود، به دست خواهد آورد. این روش دستیابی به تعداد نمونه، اگر نه اقتصادی ترین، بلکه محافظه کارانه ترین شیوه است. جدول ۱-۱۲ تعداد نمونه مورد لزوم را برای حاشیه خطای معین با  $p$  مشخص ارائه می دهد.

در نظر داشته باشید که در صورتی که حاشیه خطای قابل اغماض، پایین ترین و  $p = 0/50$  باشد، بزرگترین نمونه ضروری است. چنانچه حاشیه خطای پذیرفتنی افزایش یابد و مقدار  $p$  از  $0/50$  تغییر کند، تعداد نمونه مورد نیاز کاهش خواهد یافت.

### / برآورد میانگین جامعه آماری

فرض کنید داده های پیمایش به صورت میانگینها باشد. برای مثال، مدیر دبیرستانی مایل است در مورد سطح هوشی دانش آموزان سال اول دبیرستان، اطلاعاتی داشته باشد. این مدیر ۱۰۰ دانش آموز را به صورت تصادفی انتخاب و نمرات آنها را در یک آزمون استاندارد هوش، که سال قبل به تمام دانش آموزان شهر داده شده بود، پیدا می کند. میانگین نمرات به دست آمده برابر با ۱۱۲ بوده است. انحراف معیار جامعه آماری  $\sigma$  در این آزمون استاندارد شده ۱۵ بوده است. فرمول خطای معیار میانگین عبارت است از:

$$\text{خطای معیار} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1/5$$

جدول ۱-۱۲ حداقل تعداد نمونه لازم برای حاشیه های مختلف خطا حول برآورد پارامتر در سطح اطمینان ۹۵ درصد. تعداد نمونه به صورت تابعی از نسبت های قابل انتظار نمونه نشان داده شده اند.

حداکثر حاشیه خطا	۰/۱۰ یا ۰/۹۰ <sup>(۱)</sup>	مقدار $p$ ۰/۲۵ یا ۰/۷۵	۰/۵۰
٪۱	۳۴۶۲	۷۲۱۲	۹۶۱۶
٪۲	۸۶۶	۱۸۰۳	۲۴۰۴
٪۳	۳۸۵	۸۰۲	۱۰۶۹
٪۵	۱۳۹	۲۸۹	۳۸۵
٪۱۰	۳۵	۷۳	۹۷

۱- واریانس اطمینان  $pq$  در حالتی که  $pq = (0/90)(0/10)$  و  $pq = (0/10)(0/90)$  یکسان است.



بار دیگر محقق فاصله اطمینان را در سطح ۹۵ درصد برای مقادیر جامعه آماری، از طریق اضافه کردن و کم کردن  $1/96$  خطای معیار به آماره نمونه، به دست می آورد. بنابراین، فاصله برای میانگین جامعه آماری برابر است با (خطای معیار)  $1/96 \pm \bar{X}$ . مدیر مدرسه با میانگین ۱۱۲ و خطای معیار  $1/5$ ، می تواند با ۹۵ درصد اطمینان بگوید که میانگین هوش جامعه آماری برای کلیه دانش آموزان سال اول دبیرستان ۱۰۹ و  $112 \pm 3$  است.

از فرمول خطای معیار میانگین می توان جهت محاسبه تعداد نمونه مورد نیاز برای درجه مشخصی از دقت در سطح اطمینان ۹۵ درصد استفاده کرد. فرض کنید مدیر مدرسه حاشیه خطایی کمتر از  $3 \pm$  نمره را برای این بررسی در نظر داشته است، مثلاً می توانیم بگوییم  $2 \pm$  مورد نظر او بوده است. تعداد نمونه لازم برای حاشیه خطای ۲، یعنی  $2 =$  (خطای معیار)  $1/96$ ، برابر چه عددی می شد؟

$$1/96 \left( \frac{\sigma}{\sqrt{n}} \right) = 2 \quad \text{مقادیر را در فرمول جایگزین می کنیم:}$$

$$\frac{\sigma}{\sqrt{n}} = \frac{2}{1/96} \quad \text{هر دو طرف معادله را به 1/96 تقسیم می کنیم:}$$

$$\frac{\sigma^2}{n} = \left( \frac{2}{1/96} \right)^2 \quad \text{هر دو طرف معادله را به توان ۲ می رسانیم:}$$

$$\frac{\sigma^2}{n} = 1/04$$

$$n = \frac{\sigma^2}{1/04}$$

$$n = \frac{225}{1/04}$$

$$n = 216$$

حال معادله محاسبه  $n$  را در دست داریم. به منظور دستیابی به حاشیه خطایی برابر ۲ نمره در سطح اطمینان ۹۵ درصد، مدیر باید تعداد نمونه را به ۲۱۶ افزایش دهد. چنانچه دقت بیشتری مورد نظر باشد (برای مثال، حاشیه خطای ۱)، تعداد افراد نمونه باید باز هم بیشتر شود.

افزایش تعداد نمونه می‌تواند پرهزینه باشد. ارسال پستی پرسشنامه‌ها و پیگیریها هزینه در بردارد. مصاحبه حتی از ارسال پستی پرسشنامه‌ها هم گرانتر تمام می‌شود. محقق باید وقت و بودجه موجود برای تحقیق پیمایشی را در نظر گیرد و حجم نمونه‌ای را انتخاب کند که از نظر اقتصادی اجرای آن عملی باشد. به خاطر داشته باشید که تعداد زیاد نمونه به خودی خود ضامن معرف بودن نمونه نیست، شیوه نمونه‌گیری در تعیین اینکه آیا نمونه معرف جامعه است یا خیر، از اهمیت بیشتری برخوردار است. از خواننده خواسته می‌شود که بار دیگر به فصل ۶ در مورد شیوه نمونه‌گیری و عوامل دیگری که در تصمیم‌گیری در مورد تعداد نمونه دخالت دارند، رجوع کند.

### /// ابزار گردآوری داده‌ها

در هر دو شیوه مصاحبه و پرسشنامه از رویکرد طرح سؤال استفاده می‌شود. از این ابزار می‌توان برای به دست آوردن اطلاعات در زمینه حقایق، باورها، احساسها، نیات و غیره، استفاده کرد. هرچند در هر دو شیوه مصاحبه و پرسشنامه از رویکرد طرح سؤال استفاده می‌شود، اما تفاوت‌های قابل ملاحظه‌ای بین دو شیوه وجود دارد.<sup>(۱)</sup>

در مصاحبه، اطلاعات از طریق برخورد رو در رو و با تلفنی بین مصاحبه‌کننده و پاسخگو، گردآوری می‌شود. در پرسشنامه، اطلاعات از طریق پاسخهای کتبی پاسخگویان به مجموعه‌ای از سؤالات به دست می‌آید. هریک از این روشها نقاط ضعف و قوت خاص خود را دارد. محقق باید با در نظر گرفتن نیازها و ویژگیهای بررسی خود، در مورد انتخاب یکی از دو شیوه تصمیم‌گیری کند.

### // مصاحبه‌ها

یکی از مهمترین جنبه‌های مصاحبه، انعطاف پذیری آن است. مصاحبه‌کننده از این فرصت برخوردار است که آزمودنی و کل وضعیتی را که پاسخگو تحت آن پاسخ می‌دهد، مشاهده کند. در صورتی که سوء تفاهمی برای پاسخگو پیش بیاید، مصاحبه‌کننده می‌تواند سؤال را تکرار کند

۱ - بخشی کلی و قابل توجه در مورد ساخت و استفاده از پرسشنامه‌ها و مصاحبه‌ها به عنوان ابزار تحقیق در کتاب زیر آمده است:

D. Warwick and C.Lininger (1975), *The Sempole Survey: Theory and Practice* (ch. 6-7), New York: McGraw- Hill, and D.A.Dillman (1976), *Mail and Telephone Surveys: The Total Design Method*, New York : Wiley

و یا در مورد آن توضیح دهد. همچنین مصاحبه کننده می تواند در صورتی که پاسخ به نظرش ناقص و یا بی ربط بیاید، از پاسخگو توضیح بیشتری بخواهد.

تعداد بیشتر پرسشنامه های تکمیل شده یکی دیگر از مزایای مصاحبه است. تماس حضوری باعث می شود که احتمال همکاری پاسخگو و دادن اطلاعات مورد نظر، افزایش یابد. تعداد کم پرسشنامه های پستی عودت داده شده که خاص شیوه پستی است (معمولاً ۴۰ درصد)، نه تنها تعداد نمونه را کاهش می دهد، بلکه نتایج را جهت دار می کند. بعلاوه، مصاحبه کننده می تواند برای تمام یا غالب سوالات پاسخ گیرد. یکی از مشکلات جدی پرسشنامه پستی، از دست دادن اطلاعات است.

یکی دیگر از مزایای این شیوه، کنترلی است که مصاحبه کننده بر روی ترتیب مطرح کردن سوالات دارد. در بعضی موارد، ضروری است که پاسخگو از ماهیت سوالهای بعدی اطلاع نداشته باشد؛ زیرا پاسخ او به این سوال می تواند بر پاسخهای قبلی، اثر گذارد. این مسئله در یک مصاحبه مشکلی را ایجاد نمی کند، زیرا پاسخگو نمی داند چه سوالی پرسیده می شود و بعلاوه نمی تواند به عقب برگردد و پاسخی را که قبلاً داده است، عوض کند. در مورد پاسخگویی که نمی توانند پرسشنامه کتبی را بخوانند و پاسخ دهند، تنها راه ممکن برای گردآوری اطلاعات، مصاحبه است.

نقطه ضعف اساسی مصاحبه ها، پرهزینه تر و وقت گیرتر بودن آنها نسبت به روش پرسشنامه است.

بر مبنای ماهیت پاسخی که محقق از پاسخگو می طلبد، دو نوع اصلی سوال (باز و بسته) در مصاحبه به کار برده می شود. سوال باز، به پاسخگویان این امکان را می دهد که به جای انتخاب از بین چند گزینه، آزادانه پاسخ دهند. فرد آزاد است تا پاسخش را در چهارچوب دیدگاه خود بدهد. مصاحبه کننده بسادگی سوال را می خواند و پاسخ تحت اللفظی پاسخگو را ضبط می کند. مثالی از سوال باز عبارت است از: بارزترین مشکل خواندن که در بین دانش آموزان سال اولتان مشاهده کردید، چه بود؟

در چهارچوب سوال بسته، پرسشگر سوال را می خواند و پاسخهای انتخابی را به پاسخگو ارائه می دهد. از بین فهرست نظرات ارائه شده، پاسخگو جوابی را که بیشتر با باورها و نظراتش مطابقت دارد، انتخاب می کند. نمونه ای از سوال بسته در زیر آمده است:

معمولاً چه تکلیف نوشتنی برای واحد درسی شما ضروری است؟

۱) گزارشها (۲) مقالات و یادداشتها (۳) مقالات تحقیقاتی (۴) امتحاناتی که باید در منزل انجام شود (۵) سایر (لطفاً مشخص کنید)

چهارچوب پاسخهای باز آزادی بیشتری را برای پاسخگو جهت اظهار نظر فراهم می‌کند و معمولاً پاسخهایی در حوزه وسیعتر ارائه می‌شود، ولی سؤالات باز، هم برای مصاحبه‌کننده و هم برای پاسخگو، وقت گیرتر است. برای مصاحبه‌کننده ضابط پاسخهای تحت اللفظی پاسخگویان و طبقه‌بندی و رمزگذاری اطلاعات جهت تحلیل، امری دشوار است. پاسخگویی به سؤالات بسته برای آزمودنی ساده تر و سریعتر و برای مصاحبه‌کننده تحلیل و رمزگذاری آن سهل تر است. چهارچوب سؤالات بسته یکسان بودن موقعیت پاسخگویی را تضمین می‌کند و پاسخگویی به سؤالاتی که ماهیت احساسی یا خصوصی دارند، در صورت بسته بودن آنها، آسانتر است.

سؤالات بسته دچار این محدودیت هستند که بینشی واقعی از اینکه آیا پاسخگویان واقعاً اطلاعاتی و یا نظریه‌ای مشخصاً شکل گرفته درباره موضوع دارند یا خیر، به وجود نمی‌آورند. برای پاسخگوی ناآگاه آسانتر است که یکی از پاسخهای ارائه شده را انتخاب کند، تا اینکه عدم آگاهی خود را از موضوع ابراز دارد. برای مثال، در پاسخ به این سؤال «کدام جنبه از برنامه اقتصادی رئیس جمهور را مثرمترین می‌دانید؟» پاسخگویی که اطلاعاتی از برنامه اقتصادی رئیس جمهور ندارد، بسادگی می‌تواند پاسخی منطقی را از بین پاسخهای ارائه شده، انتخاب کند. از طرف دیگر، پاسخگویی که اطلاعات و یا نظریاتی شکل یافته در مورد موضوع دارند، از اینکه به سؤالات طبقه بندی شده ساده‌ای پاسخ دهند که به آنها اجازه ارائه پاسخهای مکفی را نمی‌دهد، خرسند نخواهند بود.

البته می‌توان در یک مصاحبه از هر دو سؤالات باز و بسته استفاده کرد. سؤالات بسته برای کسب اطلاعات واقعی یا نقطه نظرات پاسخگو در مورد مقولاتی که امکان شناسایی تعداد محدودی پاسخهای جایگزین برای آنها وجود دارد، توصیه می‌شود. برای موضوعات پیچیده‌تر که محقق می‌خواهد در مورد آنها درک آزمودنی، چهارچوب مرجع مورد استفاده برای پاسخگویی یا انگیزشهای زیربنایی پاسخ را تشخیص دهد، سؤالات باز ترجیح دارد.

در عمل، محقق در فرایند ساخت یک برنامه مصاحبه، اغلب کار خود را با تعداد محدودی

آزمودنی شروع می‌کند و با آنها مصاحبه‌ای شامل سؤالات باز انجام می‌دهد تا به دقت پاسخهای جایگزینۀ احتمالی به پرسشها را شناسایی کند سپس می‌تواند از این نمونه کوچک، برنامه مصاحبه‌ای با سؤالات بسته را بر مبنای نتایج به دست آمده تهیه کند.

### / اجرای مصاحبه

وظیفۀ اصلی مصاحبه‌کننده این است که سؤالات را به گونه‌ای پرسد که پاسخهای معتبری را به دست آورد و پاسخها را دقیق و کامل ضبط کند. کار اصلی مصاحبه‌کننده ایجاد فضایی است که پاسخگو در آن احساس راحتی کند. مصاحبه‌کننده باید بعد از معرفی دوستانۀ خود، توضیح مختصری در مورد هدف مصاحبه بدهد، لکن از دادن اطلاعات بیش از حد لازم درباره بررسی پرهیز کند؛ زیرا می‌تواند پاسخها را جهت دار کند.

مصاحبه‌کننده همچنین وظیفه دارد که توجه پاسخگو را نسبت به مصاحبه جلب کند و مصاحبه را به آرامی پیش برد. این کار وقتی انجام پذیر است که مصاحبه‌کننده کاملاً با سؤالات و توالی آنها آشنا باشد تا بتواند سؤال را به صورت محاوره‌ای مطرح سازد و برای یافتن سؤال بعدی مرتب مکث نکند. مسلم است که مصاحبه‌کننده باید از ابراز هرگونه تأیید، تعجب و یا هراس در مقابل هرگونه پاسخ پاسخگو اجتناب ورزد.

چنانچه داده‌های تطبیقی مورد نظر باشد، مصاحبه‌کننده باید فرایند را استاندارد کند. سؤالات باید به گونه‌ای یکسان و با لغات یکسان برای همه پاسخگویان خوانده شوند. در صورتی که پاسخگو مکث کند یا منحرف شود و یا پاسخ نامربوط دهد یا چنانکه به صورت بارز سؤال را بد تعبیر کند، مصاحبه‌کننده می‌تواند با گفتن «لطفاً در مورد پاسختان بیشتر توضیح دهید» یا «ممکن است کمی بیشتر توضیح دهید»، دوباره سؤال کند. مسلماً مصاحبه‌کننده باید دقت کند که پیشنهاد و یا علامتی برای پاسخ احتمالی ابراز نکند.

پاسخها باید به طور کامل و دقیق ثبت شود. در مورد سؤالات باز، لغات دقیق پاسخگو باید به صورت تحت اللفظی در حین پاسخگویی، ضبط شود. کار ثبت پاسخها را می‌توان با به کار بردن کلمات اختصاری و جمله‌ها و یا به وسیلۀ یک ضبط صوت، آسانتر کرد. ضبط صدای پاسخگو از این مزیت برخوردار است که کار ثبت پاسخهای تحت اللفظی پاسخگو و همچنین شرکت آزاد مصاحبه‌کننده در محاوره را بدون آنکه ناگزیر باشد توجه خود را بر نوشتن پاسخها

متمرکز کند، امکان پذیر می‌سازد. لکن، بسیاری از مردم از اینکه صدایشان ضبط شود، احساس ناراحتی می‌کنند و بشدت در مورد آنچه می‌گویند، محافظه کار می‌شوند. اصول اخلاقی ایجاب می‌کند که قبل از ضبط صدای پاسخگو، از او برای این کار اجازه گرفته شود.

### / آموزش مصاحبه کننده

آموزش مصاحبه کنندگان بالقوه قبل از آنکه از این شیوه استفاده کنند، بسیار ضروری است. در ابتدا مصاحبه کننده تحت تعلیم، باید نمونه هایی از مصاحبه هایی را که توسط افراد آموزش دیده صورت می‌گیرد، مشاهده کند و سپس او را وادارند که تحت نظارت مستقیم آموزش دهنده، مصاحبه را انجام دهد. در مصاحبه های تمرینی، پاسخگویان باید از همان جامعه آماری انتخاب شوند که نمونه بررسی را تشکیل خواهند داد.

### // پرسشنامه ها

تماس مستقیم تک به تک با آزمودنیها، که لازمه مصاحبه است، وقت گیر و پرهزینه است. در غالب موارد همان اطلاعات می‌تواند به وسیله پرسشنامه گردآوری شود. از آنجا که نحوه اجرای پرسشنامه به گونه‌ای است که خود فرد آن را تکمیل می‌کند و معمولاً هم به وسیله پست فرستاده می‌شود، نسبت به مصاحبه امکان بیشتری را برای گرفتن نمونه زیادتر و همچنین از نواحی پراکنده تر، فراهم می‌کند.

مزیت دیگر این شیوه آن است که اگر پرسشنامه‌ای بتواند محرمانه بودن را تضمین کند، قادر خواهد بود که پاسخهای صادقانه تری را نسبت به مصاحبه حضوری دریافت کند. در مصاحبه، احتمال دارد آزمودنیها نسبت به ابراز نظرانی که مورد پسند عام نیست و یا اطلاعاتی که فکر می‌کنند زمانی بر علیه شان مورد استفاده قرار خواهد گرفت، بی میلی نشان دهند. بعلاوه، مصاحبه کننده‌ای که ظاهر، خلق و خو و یا شیوه هدایتش بتواند بر نتایج مصاحبه اثر گذارد، در زمان تکمیل پرسشنامه حضور ندارد و در نتیجه این مسائل بالقوه دیگر وجود نخواهند داشت. یکی از نقاط ضعف پرسشنامه امکان تعبیر غلط سوالات توسط پاسخگو است. احتمال دارد محقق دقیقاً بداند که منظورش از سؤال چیست، ولی پاسخگو به علت انشای ضعیف و معانی مختلف یک واژه، از آن تعبیری کاملاً متفاوت را درک کند. بعلاوه احتمال دارد، بخش بزرگی از

جامعه توانایی خواندن و پاسخگویی به پرسشنامه فرستاده شده توسط پست را نداشته باشند. تنها افراد برخوردار از تحصیلات بالا می‌توانند یک پرسشنامه پیچیده را تکمیل کنند. همانطور که قبلاً اشاره شد، تعداد پرسشنامه های تکمیل شده‌ای که برگشت داده می‌شوند، کمتر از مصاحبه های تکمیل شده است. برای فردی که پرسشنامه را دریافت می‌کند، بسیار ساده است که آن را کنار گذارد و پر کردن و پس فرستادنش را فراموش کند. میزان کم پرسشنامه های برگشتی، تعمیم پذیری نتایج بررسی انجام شده با پرسشنامه را محدود می‌سازد. نمی‌توان فرض کرد که آن دسته از افراد نمونه که پاسخ نداده‌اند، به صورت تصادفی در کل گروه توزیع شده‌اند. بررسیهای انجام شده نشان داده است که بین ویژگیهای آن دسته از افراد نمونه که پاسخ می‌دهند و آن گروه که به این گونه بررسیها پاسخ نمی‌دهند، معمولاً تفاوت‌هایی نظام‌مند وجود دارد.<sup>(۱)</sup> معمولاً میزان پاسخگویی در بین افراد باهوش‌تر، تحصیل کرده‌تر، آگاه‌تر و علاقه‌مندتر و یا عموماً کسانی که بیشتر طرفدار مسئله مورد بررسی پرسشنامه هستند، بالاتر است. اگرچه در بررسی با پرسشنامه، هدف، برگشت ۱۰۰ درصد پرسشنامه‌هاست، لکن انتظار منطقی‌تر، بازگشت ۷۵-۹۰ درصد پرسشنامه‌هاست. عواملی در میزان برگشت پرسشنامه های ارسالی توسط پست، مؤثرند. از بین آنها می‌توان این موارد را نام برد: (۱) تعداد سوالات پرسشنامه (۲) مقدمه پرسشنامه (۳) مرجع ارسال پرسشنامه (۴) جذابیت پرسشنامه (۵) آسان بودن پر کردن و برگرداندن پرسشنامه (۶) علاقه‌ای که محتوای آن ایجاد می‌کند و (۷) شیوه‌های پیگیری مورد استفاده. این عوامل در بخشهای بعدی به‌طور مشروح مورد بحث قرار خواهند گرفت.

### / ساختن پرسشنامه

ساختن یک پرسشنامه خوب امری دشوار و وقت گیر است، ولی به خوبی می‌دانیم که احتمال اینکه یک پرسشنامه خوش ساخت، جواب خوب بگیرد، بیشتر از یک پرسشنامه ضعیف است. در زیر پیشنهادهایی برای نوشتن سوالات در یک پرسشنامه پستی ارائه می‌شود.

۱- ابزار اندازه گیری را به گونه‌ای بسازید که کیفیت بالای آن را نشان دهد. میزان برگشت پرسشنامه‌ای که به نظر عجولانه سرهم بندی شده است، زیاد نخواهد بود. در طول فرایند ساختن پرسشنامه، تجدید نظرهای متعدد جهت حذف هرگونه سؤال مبهم و یا غیر ضروری، لازم است.

۲- تا آنجا که ممکن است، پرسشنامه را کوتاه کنید. تا حداقل وقت پاسخگو را بگیرد. احتمال اینکه پاسخگویان پرسشنامه های کوتاه را تکمیل کنند و پس فرستند، بسیار بیشتر است. محقق باید کوشش کند سؤالی غیر ضروری، بخصوص آنهایی را که جوابشان را می توان از منابع دیگر به دست آورد، حذف کند. کلیه سؤالات یک پرسشنامه باید در خدمت حل مسئله تحقیق باشد؛ بدین معنی که آنها باید اطلاعاتی را ارائه دهند که برای آزمون فرضیه و یا پاسخگویی به سؤالی مطرح شده در بررسی، استفاده شوند. برای مثال، در صورتی که در یک بررسی اطلاع از سن پاسخگو برای تحلیل اطلاعات ضروری نباشد، می توان سؤال مربوط به سن پاسخگو را حذف کرد.

۳- مطمئن شوید که پاسخگویان اطلاعات لازم را برای پاسخگویی به سؤالات دارند. از طرح سؤالاتی که به تجارب و موضوعهایی می پردازد که می دانید برای نمونه شما ناآشناست، پرهیزید.

۴- سؤالات را به گونه ای طرح کنید که برای کلیه پاسخگویان قابل فهم باشد. لغات به کار برده شده باید غیر فنی و در حد درک کم سوادترین پاسخگویان باشد. در عین حال، محقق نباید در پرسشنامه خود لحن بنده نوازانه و افاده آمیز به کار برد. ایده خوبی است که از افراد دیگر، ترجیحاً کسانی که سوابقی مشابه افراد نمونه بررسی دارند، خواسته شود که سؤالات را بخوانند و تعبیرشان را از محتوای هر سؤال ارائه دهند. برای مثال سؤالاتی که در آنها از واژه هایی چون کسر بودجه، موازنه تجاری و تولید کلان ملی استفاده می شود، احتمالاً برای پیمایشی که برای عموم طرح ریزی شده است، مناسب نیست. همچنین محقق باید مراقب باشد که از اختصارات و کلماتی که از حروف اول کلمات دیگر ساخته شده و برای همگان آشنا نیستند، استفاده نکند.

۵- هر یک از سؤالات را تا آنجا که ممکن است کوتاه و ساده کنید. هر لغت و جمله ای را که برای روشن کردن مفهوم سؤال ضروری نیست، حذف کنید. انتخاب دقیق لغات، بویژه در پرسشنامه هایی که برای گروههای فرهنگی یا اجتماعی - اقتصادی سواي گروهی که محقق به آن تعلق دارد، طراحی می شود، از اهمیت بیشتری برخوردار است.

۶- سؤالات پرسشنامه را به نحوی بنویسید که پاسخهای روشن بگیرد. در صورت امکان، پاسخها باید کمی شوند. برای مثال، به جای اینکه از پاسخگو بخواهید از بین پاسخهایی چون بعضی اوقات، اغلب و همیشه، یکی را انتخاب کند، تعداد دفعات در هفته را قرار دهید. کلماتی



چون اغلب و بعضی اوقات برای افراد مختلف، معانی متفاوت دارند.

۷- سؤالات پرسشنامه به گونه‌ای باشد که پاسخهای آزمودنی را جهت دار نکند. یعنی، جمله بندی سؤال نباید به پاسخگو جهت بدهد. به این دلیل، از کلمات دوپهلو مربوط به حیثیت، حاوی مسائل عاطفی یا اغراق آمیز باید اجتناب شود. بعضی از کلمات چنان جاذبه عاطفی در فرهنگ ما دارند که استفاده از آنها به هر نوع، موجب جهت دار شدن پاسخها می‌شود. برای مثال، سؤال «آیا از حقوق آمریکایی خود تا به حال استفاده و برای رأی دادن ثبت نام کرده‌اید؟» مسلماً پاسخ را جهت دار می‌کند. سؤال ساده «آیا رأی داده‌اید؟» ارجحیت دارد. دیلمن عقیده دارد کلماتی چون آزادی، تساوی، سرمایه‌گذاری خصوصی، عدالت و صداقت در فرهنگ ما جاذبه بسیار مثبتی دارند. کلماتی چون بوروکرات، سوسیالیست، رئیس و برنامه ریزی دولت دافعه بسیار دارند. در صورت امکان، این گونه لغات نباید به کار برده شوند.

۸- از طرح سؤالاتی که به علت فقدان مفروضات بیان شده می‌توانند گمراه کننده باشند، بپرهیزید. چهارچوب داوری برای پاسخگویی به سؤالات باید برای همه پاسخگویان یکسان باشد، چنانچه مفروضاتی قبل از پاسخگویی ضروری باشد، سؤالات مطرح شده برای بررسی این مفروضات نیز باید آورده شوند.

برای مثال، یک بررسی پیمایشی که برای دانش آموزان سال آخر دبیرستان این سؤال را مطرح می‌کند: «آیا فکر می‌کنید دبیرستان شما به اندازه کافی، شما را برای دانشگاه آماده می‌کند؟»، فرض را بر این گذارده است که دانش آموز به دانشگاه خواهد رفت و نسبت به آنچه او را برای رفتن به دانشگاه آماده می‌سازد، آگاهی دارد. سؤال «آیا برای رأی دادن در انتخابات آینده نام نویسی کرده‌اید؟» فرض را بر این گذارده که دانش آموز دبیرستانی ۱۸ ساله است، اما احتمال دارد این فرض صحیح نباشد.

۹- مطمئن شوید که شقوق مختلف هر سؤال پرسشنامه جامعیت دارد و تمام حالات ممکن در آن زمینه را در بر می‌گیرد. برای مثال، اگر بخواهید وضعیت تأهل را مشخص کنید، باید نه تنها پاسخهای انتخابی، «متاهل» و «مجرد» را در بر بگیرد، بلکه شامل «مطلقه»، «بیوه» و «جدازندگی می‌کنم» نیز باشد.

برای تهیه شقوق مختلف برای سؤالات پرسشنامه‌ای که برای شناسایی نگرشها و نظرات مردم در مورد یک مسئله طراحی شده است، بهتر است که در ابتدا سؤالاتها به صورت باز به گروه

کوچکی از پاسخگویان ارائه شود. پاسخهای آنها را می توان بعداً به صورت پاسخهای انتخابی در نسخه نهایی پرسشنامه به کار برد. در مورد سؤالاتی که پاسخهای بسیار زیادی می توانند داشته باشند، باید همواره پاسخی تحت عنوان «سایر» آورد و از پاسخگو خواسته شود که در مورد آن توضیح دهد. سؤال «موقعیت شما در سیستم مدرسه چیست؟» می تواند با پاسخهای انتخابی مدیر، آموزگار، کتابدار و غیره (لطفأ مشخص کنید) دنبال شود.

۱۰- از سؤالاتی که می تواند عکس العملهایی نظیر شرمندگی، شک و تردید و یا عداوت را در پاسخگو برانگیزد، اجتناب کنید. سؤالات نباید پاسخگو را در موضع دفاعی قرار دهد. برای مثال، مردم معمولاً نسبت به پاسخگویی به سؤالات در زمینه سن، درآمد، دین و تحصیلات، اکره دارند. محقق به جای سؤال در مورد سن پاسخگو، می تواند سال تولد را بپرسد. به نظر می رسد مردم تمایل بیشتری نسبت به اعلام سال تولدشان داشته باشند تا سن خود. سؤال «آیا دیپلم دبیرستان دارید؟» ممکن است باعث شرمندگی پاسخگویی شود که دبیرستان را به اتمام نرسانده است. سؤال را می توان به این صورت مطرح کرد: «وقتی دبیرستان را ترک کردید، کلاس چندم را به پایان رسانده بودید؟»

۱۱- از طرح سؤالات مضاعف که سعی دارد دو سؤال را در یک سؤال مطرح کند، پرهیزید. برای مثال، سؤال «آیا احساس می کنید که دانشگاه باید دوره های درسی مهارتهای اولیه را برای دانشجویان ارائه کند و این دوره را به عنوان واحد درسی محسوب کند؟»، سؤالی مضاعف است. وقتی پاسخگو به این گونه سؤالات جواب می دهد، محقق نمی داند که آیا پاسخ شامل هر دو بخش سؤال می شود و یا فقط یک بخش. جواب بله به سؤال فوق می تواند به این معنی باشد که پاسخگو عقیده دارد دانشگاه، هم باید دوره های مهارتهای اولیه را ارائه دهد و هم آنها را به عنوان واحد درسی به شمار آورد، یا باید دوره ها را ارائه کند، ولی آنها را به عنوان واحد درسی منظور نکند.

### / انواع سؤال

هر دو نوع سؤالات باز و بسته در پرسشنامه به کار برده می شوند. در مبحث مربوط به مصاحبه، نقاط قوت و ضعف این سؤالات آورده شد. سؤالات را می توان به چند روش ساخت:

۱- سؤالات تکمیلی یا پرکردنی: سؤالاتی بازی هستند که پاسخگو باید با کلام خود به آنها

پاسخ دهد. برای مثال، ضعف اصلی آمادگی دانشجویان شما برای دانشگاه، کدام است؟

۲- سوالات فهرست‌گونه: سوالاتی هستند که تعدادی پاسخ احتمالی را ارائه می‌دهند و پاسخگویان پاسخی را که به آنها مربوط می‌شود، علامت می‌زنند. برای مثال، از کدام وسائل کمک آموزشی در کلاسهایتان استفاده می‌کنید؟ (هرچند پاسخی را که به شما مربوط می‌شود، علامت بزنید)

۱- تخته سیاه

۲- فیلم

۳- نوار تصویری

۴- نوار صوتی

۵- بلندگو

۶- کتاب

۷- سایر (لطفاً مشخص کنید)

طبقه بندی پاسخهای این سوالات نیز یک طبقه بندی اسمی است.

۳- سؤالهای مقیاسی: از پاسخگو می‌خواهد که مفهوم، رویداد و یا موقعیتی را در ابعادی چون (الف) کمیت یا شدت، به صورت «چه قدر» یا (ب) فراوانی به صورت «چند بار»، ارزیابی کند. برای مثال، مهارتهای نگارشی دانشجویان خود را، در این نیمسال، چگونه ارزیابی می‌کنید؟ (دور شماره مورد نظر را دایره بکشید)

۱- بسیار ضعیف

۲- کمتر از حد کفایت

۳- کافی

۴- بیش از حد کفایت

۵- عالی

۶- اطلاعی ندارم

۴- سوالات رتبه‌بندی: از پاسخگویان می‌خواهد که ترتیب ارجحیت خود را در بین تعدادی گزینه مشخص کنند. رتبه بندی نباید بیش از ۶ مورد را در بر گیرد؛ زیرا برای پاسخگویان مقایسه آنها بسیار دشوار می‌شود. نمونه‌ای از سؤالهای رتبه بندی در زیر آمده است.

آیا دانشجویان شما در خواندن برخی انواع مطالب بیش از انواع دیگر مشکل دارند؟ لطفاً ترتیب مشکل بودن مطالب را با ۱ برای دشوارترین و ۴ برای آسانترین، رتبه‌بندی کنید.

- کتابهای درسی

- دیگر کتابهای مرجع

- مقالات مجلات

- سایر (لطفاً مشخص کنید)

۵- سؤالات نوع مقیاس لیکرت: به پاسخگو امکان می‌دهد تا پاسخش را بر روی یک پیوستار از کاملاً موافقم تا کاملاً مخالفم مشخص کند و بسیار در پرسشنامه‌ها مورد استفاده قرار می‌گیرد. مزیت این نوع سؤال در آن است که می‌توان به پاسخهای مختلف آن امتیاز داد و اندازه‌های گرایش مرکزی، تغییر پذیری، همبستگی و اندازه‌های مشابه را محاسبه کرد.

### / ترتیب سؤالات

توصیه می‌شود که به اولین سؤال توجه خاصی شود؛ زیرا این سؤال می‌تواند تعیین‌کننده ادامه پاسخگویی به پرسشنامه و یا کنار گذاردن آن، از جانب پاسخگو باشد. بسیار حائز اهمیت است که سؤال اول برای همه پاسخگویان به میزان کافی جالب و ساده باشد تا بتوانند آن را تفسیر کنند و به آن پاسخ گویند. اینکه پاسخگو بتواند به سؤال اول پاسخ دهد، موجب افزایش انگیزش و اعتماد به نفس پاسخگو می‌شود؛ زیرا احساس می‌کند که توانایی تکمیل پرسشنامه را دارد، ولی آسان بودن بدان معنی نیست که این سؤال بی‌اهمیت است. ضروری است که سؤال اول به دنبال دریافت اطلاعات مفیدی باشد که به روشنی با موضوع بررسی مرتبط است. به این دلیل، نباید پرسشنامه‌ها را با سؤالهای مربوط به سن، جنس، تحصیلات، شغل، اصلیت، وضعیت تأهل و غیره آغاز کرد. پاسخگو احتمال دارد این سؤالات را بی‌ربط بداند و یا آن را تجاوز به حریم خصوصی خود پندارد، و در نتیجه تکمیل پرسشنامه را رها کند. همچنین توصیه می‌شود چند سؤال اول به صورت سؤالهای بسته، و نه باز، باشد تا پاسخگو بتواند سرعت به آنها پاسخ دهد. سؤالهای باز شاید به پاسخ نوشتاری طولانی نیاز داشته باشند.

سؤالهایی را که از نظر محتوا شبیه هم هستند، باید کنار یکدیگر آورد. برای مثال، در پرسشنامه‌ای که از استادان دانشگاه در مورد مهارتهای اولیه دانشجویانشان سؤال می‌شود، باید تمامی سؤالهای مربوط به خواندن را با هم آورد. سپس سؤالهای مربوط به نوشتن کنار یکدیگر

قرار می‌گیرند و به دنبال آن سؤالهای مربوط به مهارتهای ریاضیات آورده می‌شود. در یک زمینه محتوایی، سؤالها بر مبنای نوع سؤالات دسته‌بندی می‌شوند. برای مثال، سؤالهایی که جوابشان صرفاً بلی یا خیر است، کنار یکدیگر و سؤالهایی که پاسخشان به رتبه بندی و یا مشخص کردن میزان موافقت و مخالفت از سوی آزمودنی نیاز دارد، در یک جا آورده می‌شوند.

در هر زمینه موضوعی، سؤالها باید به گونه‌ای مرتب شوند که از نظر روان شناختی مناسب باشد. چنانچه ترتیب سؤالها از نظر منطقی و روان شناختی صحیح باشد، باعث می‌شود پاسخگو با تفکر بهتر به سؤال پاسخ دهد. برای مثال، محقق ابتدا باید در مورد اینکه آیا پاسخگویان از شرایط کاریشان راضی هستند یا خیر، مطمئن شود و سپس از آنها در مورد تغییراتی که پیشنهاد می‌کنند، سؤال کند. چنانچه در یک زمینه موضوعی هم سؤالهای کلی و هم سؤالهای خاص پرسیده شود، باید سؤالهای کلی را اول آورد. سؤالهای عینی در مورد یک مسئله باید قبل از سؤالهای ذهنی آورده شود. سؤالاتی که کمتر احتمال دارد ناخوشایند باشند، باید قبل از سؤالهای ناخوشایندتر آورده شوند. برخی مواقع مردم تمایلی به پاسخ دادن به سؤالهایی در زمینه نگرشها، ارجحیتها، انگیزه‌ها، رفتارها و احساسات شخصی و زمینه‌های مشابه ندارند، ولی چنانچه در ابتدا سؤالهای عینی، که مشخص کننده و روشن کننده موقعیت باشند، آورده شوند، احتمالاً پاسخگویی برای پاسخگو آسانتر می‌شود. برای مثال، محقق می‌خواهد تحقیقی پیمایشی را در مورد میزان استعمال ماری جوانا در بین دانش آموزان انجام دهد، می‌تواند ابتدا با سؤالهای عینی تر نظیر «شما استعمال ماری جوانا را در مدرسه‌تان چگونه ارزیابی می‌کنید: مشکلی جدی، مشکلی معمولی، مشکلی جزئی یا هیچ مشکلی؟» آغاز کند. به دنبال این سؤال می‌تواند سؤالهای «آیا فکر می‌کنید میزان استعمال ماری جوانا، امسال افزایش یافته، به همان میزان مانده یا کاهش یافته است؟» یا «آیا دانش آموزانی را می‌شناسید که ماری جوانا استعمال می‌کنند؟» و سپس شاید «آیا شما هرگز ماری جوانا کشیده‌اید؟» آورده شود.

به کارگیری اصل یاد شده در مورد قرار دادن سؤالهای کمتر ناخوشایند قبل از سؤالهای ناخوشایندتر، در مورد سؤالهای مربوط به مشخصات فردی نظیر سن، جنس، شغل نیز صادق است و این گونه سؤالها باید به جای آنکه در ابتدای پرسشنامه آورده شوند، در انتهای آن قرار گیرند. برای پاسخگو، پس از تکمیل بقیه پرسشنامه و بعد از آنکه احساس کرد این گونه اطلاعات می‌تواند با موضوع بررسی ارتباط داشته باشد، پاسخگویی به این سؤالها کمتر شک برانگیز است.

## / شکل پرسشنامه

شکل پرسشنامه باید به گونه‌ای باشد که از جذابیت برخوردار باشد و خواندن و پاسخگویی به آن برای پاسخگو و رمز گذاری و نمره دادن به آن برای محقق آسان باشد.<sup>(۱)</sup>

۱- سؤالات را در تمام طول پرسشنامه به صورت متوالی شماره گذاری کنید، بدون آنکه شماره‌ای تکرار و یا حذف شود. اختصاص یک شماره خاص برای هر سؤال باعث می‌شود که دچار سردرگمی در رمز گذاری پاسخها نشویم.

۲- با استفاده از حروف کوچک برای سؤالاها و حروف بزرگ برای گزینه های پاسخ، می‌توان سؤالاها را از پاسخها متمایز کرد. هر رهنمود خاصی برای پاسخگویی در داخل پرانتز قرار می‌گیرد و با حروف کوچک نوشته می‌شود. برای مثال:

آیا با وضع معیاری برای مهارتهای اولیه، به عنوان یکی از ملزومات فارغ التحصیلی از دبیرستان موافقت می‌کنید؟ (دور شماره را دایره بکشید)

۱- خیر

۲- بلی

۳- از شماره ها برای شناسایی طبقات مختلف پاسخ استفاده کنید. استفاده از شماره ها نسبت به حروف، خط تیره، مربع یا دایره جلوی هر پاسخ ارجحیت دارد؛ زیرا شماره تخصیص یافته به یک گزینه، خود شکلی از رمز گذاری اولیه است که پردازش داده ها را تسهیل می‌کند. برای مثال:

بالاترین سطح تحصیلی که به اتمام رسانده‌اید، چه است؟

۱- ابتدایی

۲- بخشی از دبیرستان

۳- دبیرستان را تمام کرده‌ام.

۴- بخشی از دانشگاه

۵- دانشگاه را تمام کرده‌ام.

۶- تا حدودی در دوره فوق لیسانس کار کرده‌ام.

۷- درجه فوق لیسانس یا دکترا

شماره ۵ نشان می‌دهد که فرد دانشگاه را تمام کرده است و به آسانی می‌تواند در آن طبقه تحصیلی محاسبه شود. از آنجا که طول جملات مربوط به اظهار نظرها متفاوت است، توصیه می‌شود که شماره مربوط به هر گزینه به جای سمت راست در سمت چپ آن (در زبان انگلیسی) قرار داده شود.

۴- در تخصیص نمرات به طبقات مختلف پاسخها، ثبات داشته باشید. یعنی، در تمام پرسشنامه همواره از یک نمره برای همان پاسخ استفاده کنید. معمولاً نمرات پایین را برای پاسخهای منفی و نمرات بالاتر را برای پاسخهای مثبت تخصیص می‌دهند. برای مثال، ۱ برای نه و ۲ برای بلی، ۱ برای غیر مطلوب و ۲ برای مطلوب، به کار برده می‌شود. به هر حال شکلی که انتخاب می‌شود، باید در تمام پرسشنامه ثابت بماند، زیرا اگر در سؤالی ۱ برای خیر و در بخش دیگر پرسشنامه، ۱ برای بلی تخصیص داده شود، پاسخگو دچار سردرگمی خواهد شد.

۵- پاسخها باید به جای آنکه به صورت افقی آورده شوند، به صورت عمودی نوشته شوند. ترتیب پاسخها به صورت عمودی از این مزیت برخوردار است که پرسشنامه به نظر کمتر شلوغ می‌آید و در عین حال خطای متداول علامت‌گذاری کنار پاسخ نیز حذف می‌شود. برای مثال:

وضعیت تأهل شما در حال حاضر چیست؟

۱- هیچوقت ازدواج نکرده‌ام ۲- متأهل ۳- طلاق گرفته‌ام ۴- جدا زندگی می‌کنم ۵- بیوه

ترتیب گمراه‌کننده فوق به صورت زیر تصحیح می‌شود.

وضعیت تأهل شما در حال حاضر چیست؟

۱- هیچوقت ازدواج نکرده‌ام.

۲- متأهل هستم.

۳- طلاق گرفته‌ام.

۴- جدا زندگی می‌کنم.

۵- بیوه هستم.

۶- پرسشنامه را با روشی نظیر افست، که از کیفیت چاپ بهتری برخوردار است، تکثیر

کنید. چاپ با کیفیت بالا به پرسشنامه ظاهری حرفه‌ای‌تر می‌بخشد و اثری مثبت بر پاسخگویان

دارد.

## / انجام پیش آزمون

برای شناسایی ابهامها، سوتعبیرها یا دیگر نارساییهای پرسشنامه، لازم است آن را قبل از چاپ نهایی پیش‌آزمون کرد. ابتدا، بهتر است از همکاری که با بررسی آشنایی دارند، خواسته شود که پیش‌نویس پرسشنامه را بررسی کنند و نظراتشان را در مورد اینکه آیا این وسیله اطلاعات مورد نظر را به دست می‌آورد و یا مشکلی در آن ملاحظه می‌شود که از چشم افتاده است، ارائه دهند.

سپس، پرسشنامه به صورت شخصی و انفرادی با گروهی کوچک که از میان همان جامعه آماری مورد نظر انتخاب شده‌اند، پر شود. پاسخگویان سؤالات را تک به تک پاسخ می‌دهند و به محقق در زمینه مشکلاتی که با هر سؤال دارند، باز خورد می‌دهند. به اظهار نظرهایی نظیر «نمی‌دانم در اینجا منظورتان چیست» و «بیش از یکی از این پاسخها در مورد من صدق می‌کند»، باید توجه خاص شود. محقق باید اطمینان حاصل کند که سؤالاها به صورت یکسان توسط تمامی پاسخگویان تفسیر می‌شوند.

مشاهدهٔ پاسخگویان در حین تکمیل پرسشنامه نیز می‌تواند روشنگر باشد. صرف بیش از حد وقت بر روی یک سؤال و گذشتن از یک سؤال و بازگشت به آن در نوبت بعدی، می‌تواند به عنوان مشکل موجود در آن سؤال تلقی شود.

نتایج پیش‌آزمون می‌تواند در جهت روشن کردن سؤالاها و یا احتمالاً حذف آنها، کاربرد داشته باشد. همچنین مسئله‌ای که شدیداً حائز اهمیت است، این است که تعیین شود آیا افراد متعلق به طبقات مختلف اجتماعی و فرهنگی جامعه آماری مورد بررسی، سؤالاها را یکسان برداشت می‌کنند یا خیر.

سؤالاهاى خاصی که توسط نتایج پیش‌آزمون باید به آنها پاسخ داد، عبارتند از:

- ۱- آیا به نظر می‌رسد پاسخگویان با پرسشنامه راحت هستند و برای تکمیل آن انگیزه دارند؟
- ۲- آیا برخی از سؤالات، گمراه کننده هستند؟
- ۳- آیا بعضی از سؤالات در پاسخگو احساس شرمندگی و یا خشم ایجاد می‌کنند؟
- ۴- آیا دستورالعملها روشن هستند؟
- ۵- پاسخگو در چه مدت پرسشنامه را تکمیل می‌کند؟
- ۶- آیا برداشت همهٔ پاسخگویان از سؤالات، یکسان است؟



## / توزیع پرسشنامه

محققان ممکن است فرستادن یک نامه حاوی معرفی طرح را برای پاسخگویان بالقوه، قبل از ارسال خود پرسشنامه، مفید بیابند. به این طریق آزمودنی مورد بررسی نسبت به آن آگاهی می‌یابد و در نتیجه با دیدن بسته حاوی پرسشنامه گیج نمی‌شود. در هر صورت باید نامه‌ای خطاب به پاسخگو با نام و عنوان، همراه پرسشنامه باشد. شکل ۱-۱۲ نامه پیوست پرسشنامه را که در آن بخشهای حایز اهمیت مشخص شده‌اند، نشان می‌دهد. نامه پیوست پرسشنامه به منظور معرفی پرسشنامه به پاسخگویان بالقوه و وادار کردن آنها به پاسخگویی به آن به کار می‌رود. نامه یادداشت باید حاوی بخشهای زیر باشد:

۱- هدف بررسی: در پاراگراف اول نامه باید هدف بررسی و کاربرد بالقوه آن را توضیح داد. مرتبط ساختن اهمیت بررسی به یک گروه مرجع که افراد نمونه با آنها آشنایی دارند، می‌تواند کمک کننده باشد. نامه پیوست که برای دانشجویان دوره فوق لیسانس یا دکترا فرستاده می‌شود، مسلماً باید بر اهمیت این اطلاعات در بهبود بخشیدن به آموزش دانشگاهی دوره های فوق لیسانس یا دکترا، تأکید کند.

۲- تقاضای همکاری: نامه باید توضیح دهد که چرا پاسخگویی بالقوه در نمونه بررسی گنجانده شده است و همچنین باید جاذبه‌ای برای او ایجاد کند و او را به همکاری وادارد. باید این احساس را در پاسخگویان ایجاد کرد که سهم مهمی در بررسی دارند.

۳- دادن احساس امنیت به پاسخگو: نامه نه تنها باید این اطمینان را به پاسخگویان بدهد که پاسخهایشان محرمانه تلقی خواهد شد، بلکه باید چگونگی محرمانه نگاهداشتن پاسخها را برایشان توضیح دهد. به منظور تسهیل امر پیگیری که برای بالابردن میزان برگشت پرسشنامه ضروری است، توصیه می‌شود بر روی پرسشنامه ها از شماره شناسایی استفاده شود. در صورتی که روی پرسشنامه، شماره شناسایی نباشد، مسئله جهت دار شدن بررسی به لحاظ عدم پاسخگویی همه افراد نمونه، غامض تر می‌شود؛ زیرا راهی وجود ندارد که بفهمیم چه کسی پاسخ داده و چه کسی پاسخ نداده است و لذا امر پیگیری بسیار پیچیده می‌شود. در صورتی که از شماره های شناسایی استفاده شود، ضروری است به پاسخگویان گفته شود که این شماره تنها به این دلیل به کار برده شده است که محقق بتواند در زمان برگشت پرسشنامه، نام پاسخگو را از فهرست افرادی که پرسشنامه را برایشان فرستاده است، حذف کند. باید به پاسخگو اطمینان داد

که نام آنها هرگز بر روی خود پرسشنامه ها نخواهد آمد و در نتیجه راهی وجود نخواهد داشت که بتوان پاسخها را به افراد مربوط ساخت. در صورتی که محقق تصمیم داشته باشد بلافاصله بعد از آنکه پاسخها را ثبت کرد، پرسشنامه ها را نابود سازد، این موضوع باید در نامه بیاید تا به پاسخگویان اطمینان داده شود که گمنام باقی خواهند ماند.

بعضی از محققان ترجیح می دهند از به کار بردن هرگونه سیستم شناسایی، بویژه در مورد موضوعهای حساس، حذر کنند. در این مورد، ضروری است در بسته ارسالی، کارت پستی نیز گذارده شود تا پاسخگو آن را به صورت جداگانه پست کند و معلوم شود پرسشنامه برگشت داده شده است. کارت پستال حاوی متنی تایپ شده دال بر آن است که پرسشنامه عودت داده شده است. جایی نیز در آن برای نام پاسخگو در نظر گرفته می شود. به این طریق، سابقه‌ای از پرسشنامه‌های برگردانده شده، باقی می ماند.

۴- کفالت بررسی: امضای نامه حایز اهمیت است و بر برگرداندن پرسشنامه اثر می گذارد. چنانچه بررسی بخشی از یک تز دکترا باشد، اگر فردی که برای پاسخگویان بخوبی شناخته شده است (مثل رئیس بخش و یا رئیس دانشکده)، نامه را امضا کند و یا آن را ظهرنویسی کند، بسیار کمک کننده خواهد بود. چنین امضایی احتمالاً مؤثرتر از امضای یک دانشجوی ناشناس دوره دکترا خواهد بود. چنانچه کفالت بررسی بر عهده بنیاد و یا سازمانی است، این امر باید ذکر شود. باید از نامه های با عنوان مربوط به دانشگاه و یا سازمان استفاده کرد.

۵- وعده فرستادن نتایج: می توان به پاسخگویان پیشنهاد کرد که در صورتی که علاقه مند باشند، نتایج بررسی برایشان ارسال شود. باید برای پاسخگویان توضیح داده شود که چگونه تقاضای دریافت نتایج را به اطلاع محقق برسانند. یک روش عبارت است از مشخص کردن محلی بر روی پاکت مخصوص برگرداندن پرسشنامه تا پاسخگو آن را علامت بزند و همچنین محلی برای نام و آدرس پاسخگو.

۶- تشکر: باید برای کمک و همکاری در اجرای بررسی ابراز تشکر کرد.

۷- تاریخ روز روی نامه: تاریخ روی نامه پوست باید به زمان پست کردن آن نزدیک باشد. نامه‌ای که تاریخش چندین هفته قبل از زمان دریافتش باشد، پاسخگوی بالقوه را تحت تأثیر قرار نمی دهد.

۸- درخواست برگرداندن فوری: تقاضای برگرداندن پرسشنامه در اسرع وقت، حایز اهمیت

است. چنانچه زمانی حدود ۲ هفته یا یک ماه تعیین شود، پاسخگو ممکن است پرسشنامه را کنار گذارد و به جای آنکه انگیزه پرکردنش را پیدا کند، آن را فراموش نماید. در صورتی که پرسشنامه‌ای توجه پاسخگو را در یک هفته اول جلب نکند، احتمالاً هیچوقت برگردانده نخواهد شد.

شکل ۱-۱۲ نمونه‌ای از یک نامه پیوست پرسشنامه برای یک تحقیق پیمایشی

<p>سرنامه ←</p> <p>تاریخ ارسال ←</p> <p>هدف بررسی ←</p> <p>اهمیت پاسخگر ←</p> <p>درخواست همکاری ←</p> <p>مهلت ارسال پاسخ ←</p> <p>ایجاد اعتماد در مورد محرمانه بودن پاسخ ←</p> <p>اراز تشکر ←</p> <p>امضای مدیر پروژه ←</p>	<p style="text-align: center;">دانشگاه ایندیانا</p> <p style="text-align: center;">دفتر مطالعات ارزشیابی و انجام آزمونها تقاطع خیابان سوم و خیابان جوردن بلومینگتون، ایندیانا ۴۷۴۰۵</p> <p style="text-align: center;">۱۵ فوریه، ۱۹۸۹ فارع‌التحصیل عزیز</p> <p>دفتر مطالعات ارزشیابی و انجام آزمونها قصد دارد برای آگاهی از نگرشها و عقاید فارع‌التحصیلان جدید دانشگاه ایندیانا در بлумینگتون، دربارهٔ تجربیات آموزشی از آنان نظرخواهی کند. ما علاقه‌مندیم بدانیم که این دانشگاه، نیازهای آموزشی شما را چگونه برآورده ساخته است. نتایج این نظرخواهی در نقد و بررسی و تقویت برنامه‌های آموزشی دانشجویان فعلی و آتی مورد استفاده قرار خواهد گرفت.</p> <p>نام شما از فهرست فارع‌التحصیلان دانشگاه ایندیانا از سال ۱۹۸۲ تا ۱۹۸۷ به طور تصادفی انتخاب شده است. برای آن که نتایج این نظرخواهی معرف تمام دانشجویان جدید باشند، تکمیل و برگرداندن هر پرسشنامه از اهمیت برخوردار است. پاسخ به پرسشنامه حتی ده دقیقه از وقت شما را خواهد گرفت. اما برای موفقیت این بررسی بسیار حیاتی است. من از شما خواهش می‌کنم تا ۲۴ فوریه، ۱۹۸۹، این پرسشنامه را تکمیل و آن را در پاکتی که به همراه پرسشنامه است، برای ما ارسال کنید. شما باید مطمئن باشید که پاسخهای شما کاملاً محرمانه باقی خواهد ماند. پاکت برگشت پرسشنامه دارای یک شماره شناسایی است که به ما امکان می‌دهد نام شما را با فهرست دریافت کنندگان پرسشنامه مطابقت دهیم. پس از انجام این کار، پاکت را دور خواهیم انداخت. نام شما هرگز بر روی برگه پاسخها یا پرسشنامه فید نخواهد شد. اگر به دریافت خلاصه‌ای از نتایج نظرخواهی علاقه‌مند هستید لطفاً در پشت پاکت در مربع مربوطه علامت بگذارید. تا نتایج بررسی تا اواسط تابستان برای شما ارسال شود.</p> <p>اگر در مورد این نظرخواهی سؤالی دارید، لطفاً به وسیله نامه یا تلفن با ما تماس بگیرید.</p> <p style="text-align: center;">شمارهٔ تلفن ما ۱۵۹۵ - ۸۵۵ - ۸۱۲ است. از همکاری شما بسیار سپاسگزاریم با احترام لوسی سی. جیکوبز، Ph.D. مدیر پروژه</p>
---	---

کلیه موارد فوق باید در نامه پیوست پرسشنامه گنجانیده شود، ولی در عین حال این نامه باید تا آنجا که ممکن است کوتاه باشد. حداکثر متن پیشنهادی نامه، یک صفحه است. نامه را همراه با پرسشنامه در یک پاکت قرار دهید. همیشه آدرس و پاکت تمبردار را جهت برگرداندن پرسشنامه، برای استفاده پاسخگو ضمیمه کنید. اگر می‌خواهید پرسشنامه‌های زیادی برگردانده شوند، این امر ضروری است.

### / پیگیریها

به منظور دستیابی به حداکثر درصد بازگشت پرسشنامه‌های پستی یک تحقیق پیمایشی، برنامه‌ریزی برای پیگیری پرسشنامه‌ها، ضروری است. معمولاً در پیگیری، چند مرحله طی می‌شود:

**یادآوری اولیه** - چنانچه پرسشنامه ظرف یک هفته تا ۱۰ روز پس از پست کردن آن، عودت داده نشد، باید برای پاسخگو کارت پستالی فرستاده شود. این کارت پستال در واقع یادآوری مؤدبانه‌ای است برای آنکه پرسشنامه‌ای فرستاده شده و پاسخ آن از نظر بررسی بسیار حائز اهمیت است. این کارت پستال پاسخگویان را به تکمیل و برگرداندن پرسشنامه در اسرع وقت (همان روز) ترغیب می‌کند. البته از کسانی که قبلاً پرسشنامه را پست کرده‌اند، باید تشکر شود.

باید به کسانی که پرسشنامه را گم کرده‌اند و یا هرگز دریافت نکرده‌اند، پیشنهاد ارسال نسخه دیگری از آن داده شود. معمولاً کارت پستال یادآوری تعداد زیادی پاسخ به دنبال دارد.

**پیگیری ثانویه** - این مرحله پیگیری، که معمولاً ۳ هفته بعد از تاریخ ارسال پرسشنامه اولیه صورت می‌گیرد، شامل ارسال یک نامه، نسخه دیگری از پرسشنامه و پاکت آدرس‌دار جهت برگرداندن پرسشنامه است. در این نامه ابتدا به کسانی که پاسخ نداده‌اند، گفته می‌شود که پرسشنامه‌های آنها دریافت نشده است و مفید بودن آنها را برای بررسی تصریح می‌کند. در مورد فرستادن یک پرسشنامه اضافی برای جایگزینی پرسشنامه قبلی باید تأکید شود و برای تکمیل و برگرداندن آن شدیداً اصرار شود. باید به پاسخگویان گفته شود که چنانکه قبل از دریافت این نامه، پرسشنامه را پست کرده‌اند، پرسشنامه دوم را پر نکنند.

**پیگیری سوم** - سومین مرحله پیگیری بین ۶ تا ۷ هفته بعد از تاریخ ارسال پرسشنامه اولیه

صورت می‌گیرد. فرایند آن شبیه مرحله دوم و شامل یک نامه و پرسشنامه جایگزین است. بسیاری از محققان این مرحله پیگیری را از طریق پست سفارشی انجام می‌دهند. چنانچه محقق پس از طی مراحل سه گانه پیگیری ۷۵ تا ۹۰ درصد پرسشنامه‌ها را دریافت کند، می‌تواند اجرای تحقیق پیمایشی را خاتمه یافته تلقی کند و باقیمانده آزمودنیها را بدون جواب بخواند. بر محقق است که تصمیم بگیرد آیا پاسخهایی که بر اثر پیگیری اضافی دریافت می‌کند، ارزش وقت و هزینه‌اش را دارد یا خیر.

بعضی اوقات توصیه می‌شود که محقق در مرحله سوم پیگیری کارت پستی ضمیمه کند که در آن آزمودنیها بتوانند مشخص کنند که تمایلی به شرکت در تحقیق پیمایشی ندارند و پرسشنامه را بر نخواهند گردانید. چنین کاری شناسایی کسانی را که جواب نداده‌اند، آسان می‌کند.

#### / طرز برخورد با کسانی که پاسخ نداده‌اند

در تحقیقات پیمایشی، کسانی که پاسخ نمی‌دهند، مشکلی جدی به حساب می‌آیند. چنانچه پس از طی مراحل پیگیری، میزان برگشت پرسشنامه‌ها حدود ۷۵ درصد باشد، محقق باید سعی کند با کسانی که پاسخ نداده‌اند، تماس بگیرد و اطلاعاتی در مورد ویژگیهای آنها به دست آورد و همچنین پاسخهای آنها را بگیرد. نتایج تحقیقات نشان داده‌اند که به نظر می‌رسد ویژگیهایی نظیر تحصیلات، هوش، انگیزش و علاقه مندی به موضوع تحقیق در بین کسانی که پاسخ داده‌اند و گروهی که پاسخ نداده‌اند، متفاوت است. هرچند ممکن است در ابتدای امر، محقق پرسشنامه‌های پستی را برای نمونه‌ای معرف جامعه آماری فرستاده باشد، ولی نتایج تحقیقات پیمایشی که میزان پاسخ آنها کم است، می‌تواند بشدت جهت دار باشد. یک روش متداول به این صورت است که محقق سعی کند شخصاً یا با تلفن یا نمونه‌ای تصادفی و کوچک (شاید ۱۰ درصد) از کسانی که پاسخ نداده‌اند، مصاحبه تلفنی کند. می‌توانیم میانگین پاسخها و یا نسبت پاسخگویان و آن دسته از افراد نمونه را که به سوالات پاسخ نداده‌اند، مقایسه کنیم تا ببینیم آیا دو گروه به صورت معنی داری متفاوت هستند یا خیر. چنانچه در مقایسه بین پاسخهای پاسخگویان اولیه و افراد نمونه مصاحبه شده، تفاوتی معنی داری دیده نشود، می‌توان منطقاً فرض کرد که پاسخگویان، معرف نمونه غیر آریبی از تمام کسانی هستند که

پرسشنامه دریافت کرده‌اند. ولی، محقق برای آنکه بداند آیا پاسخگویان نمونه آریبی از جامعه آماری هستند یا خیر، راهی به جز این ندارد. در صورتی که تفاوت‌های معنی داری مشاهده شود، در تحلیل نهایی داده‌ها، می‌توان در مورد پرسشنامه‌هایی که بازگردانیده نشده‌اند، روش وزن دادن را به کار برد.<sup>(۱)</sup> ایکن<sup>(۲)</sup> نشان داده است که میزان معرف بودن پاسخهای کل نمونه تحقیق پیمایشی، تابعی است از حجم نمونه، نسبت پرسشنامه‌های برگشتی و نسبت پاسخگویی که به سؤالات در جهت خاصی پاسخ داده‌اند. او برای تعیین حداقل نسبت افرادی که باید پرسشنامه تحقیق پیمایشی را برگردانند تا محقق بتواند مطمئن باشد که پاسخهای پاسخگویان معرف کل جامعه آماری است، فرمولی را ارائه می‌کند.

چنانچه نتایج تحقیق نشان دهد که گروه‌های فرعی خاصی که بوضوح قابل شناسایی هستند، پرسشنامه را برگردانده‌اند، ممکن است ضروری باشد که سؤال اصلی تحقیق را تغییر دهند و این گروه‌های فرعی را حذف کنند. برای مثال، چنانچه در یک بررسی، معلمان دبیرستان میزان کمتری پرسشنامه را نسبت به معلمان ابتدایی برگردانده باشند، محقق ممکن است نتیجه‌گیری کند که پرسشنامه ارتباط کمتری با معلمان دبیرستان داشته و یا کمتر مورد علاقه آنها بوده است و تصمیم بگیرد که بررسی را به معلمان ابتدایی محدود سازد. سؤال تحقیق باید دوباره تبیین شود تا تغییر را نشان دهد.

#### / پرسشنامه‌هایی که به صورت دسته جمعی پر می‌شوند

پرسشنامه‌ای که توسط خود فرد پر می‌شود، معمولاً در گروه‌هایی که به لحاظ هدف خاصی گرد هم آمده‌اند (برای مثال، والدین دانشجویان سال اول که در روز قبل از نام‌نویسی در محوطه دانشگاه گرد هم آمده‌اند) یا برای گردآوری اطلاعات در موقعیتی خاص (برای مثال، در خوابگاه)، تکمیل می‌شود. مسئول اجرای تحقیق پیمایشی حضور دارد و می‌تواند در صورت لزوم کمک کند. معمولاً نمونه کاملاً خاص است (والدین سال اولیها) و تنها در مورد آن جامعه آماری قابلیت تعمیم دارد.

1- Hausen, M. H., and W. N. Hurwitz (1948). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.

2- Aiken, L. R. (1981). Proportion of returns in survey research. *Educational and Psychological Measurement*, 41, 1033-1038.

## // مصاحبه تلفنی

مصاحبه تلفنی به عنوان منبعی از اطلاعات مورد توجه قرار گرفته است. مزیت اصلی آن علاوه بر میزان بالای پاسخی که دریافت می‌کند، کم‌هزینه‌تر بودن و سرعت انجام آن است. مصاحبه تلفنی می‌تواند با افرادی که در یک ناحیه وسیع جغرافیایی پراکنده‌اند، در زمانی کوتاه، انجام گیرد. معمولاً سازمانهای نظرخواهی ملی به منظور دستیابی به نظرات رأی دهندگان در سطح کشور، در زمانهای نزدیک به برگزاری انتخابات، از تلفن استفاده می‌کنند. دیگر تحقیقات پیمایشی که در سطح وسیع در شهرهای عمده کشور اجرا می‌شود، به جای فرستادن پرسشگر به محلهایی که می‌تواند برای آنها خطرناک باشد، از تلفن استفاده می‌کنند. دیگر مزیت مصاحبه تلفنی در یک شهر بزرگ، دسترسی به افرادی است که ممکن است درهای خانه‌هایشان را بر روی پرسشگران باز نکنند، ولی تمایل داشته باشند که با تلفن صحبت کنند.

نقطه ضعف اصلی آن فقدان امکان برقراری رابطه با پاسخگو است که در شرایط مصاحبه حضوری وجود دارد. احتمال دارد کنار آمدن با شک و تردید پاسخگویان شگفت زده، بویژه زمانی که سؤلهای شخصی و یا حساس پرسیده می‌شوند، امری دشوار باشد. روشی مناسب برای برخورد با این مشکل عبارت است از فرستادن نامه‌ای که پاسخگویان بالقوه را از تلفنی که به زودی به آنها خواهد شد، آگاه می‌سازد. برای دستیابی به نتایج خوب، برخورداری از مهارت بسیار زیاد برای ساختن یک پرسشنامه تلفنی و انجام مصاحبه، ضروری است. نقطه ضعف دیگر این شیوه حذف خود بخودی خانواده‌هایی است که تلفن ندارند و می‌توانند نتایج را جهت دار کنند. برای اطلاعات بیشتر در مورد مصاحبه تلفنی، از خوانندگان خواسته می‌شود که به کتاب دیلمن<sup>(۱)</sup> مراجعه کنند.

## /// اعتبار

مسئله اعتبار باید مورد توجه قرار گیرد، یعنی باید ببینیم آیا مصاحبه یا پرسشنامه واقعاً آن چیزی را که باید اندازه بگیرد، اندازه می‌گیرد یا خیر. یکی از بارزترین شواهد مورد نیاز اعتبار عبارت است از اعتبار مربوط به محتوا، که این نوع اعتبار را می‌توان از طریق درخواست قضاوت، از همکاران توانایی که با هدف پیمایش مورد نظر آشنایی دارند، به دست آورد. در این شیوه، از

این همکاران خواسته می‌شود که سؤالات تحقیق پیمایشی را بررسی کنند و بگویند آیا این سؤالات برای اندازه‌گیری آنچه باید اندازه بگیرند، کفایت دارند یا خیر و آیا آنچه می‌سنجند یک نمونه معروف از حیطه رفتاری در دست تحقیق هست یا خیر.

در برخی بررسیها از مشاهده مستقیم رفتار برای یافتن شواهد مرتبط با ملاک اعتبار پاسخها، استفاده شده است. پس از آنکه پاسخها گرفته شد، با استفاده از روش مشاهده، بررسی می‌شود که آیا رفتار عملی آزمودنی با آنچه در مورد نگرشها، نظرات و دیگر پاسخها، ابراز کرده است، توافق دارد یا خیر. منابع اطلاعاتی دیگر، نظیر گروههای سوم، می‌تواند به عنوان ملاک مورد استفاده قرار گیرد.

برخی از متغیرهایی که بر اعتبار یک پرسشنامه اثر می‌گذارند، عبارتند از: (۱) تا چه حد موضوع بررسی برای پاسخگو اهمیت دارد؟ می‌توان فرض کرد که پاسخهای معتبرتر را از افرادی بگیریم که به موضوع بررسی علاقه مندتر هستند و یا در مورد آن آگاهی بیشتری دارند و (۲) آیا پرسشنامه، گمنام باقی ماندن پاسخگو را تضمین می‌کند؟ منطقاً می‌توان فرض کرد در صورتی که پاسخگویان گمنام باقی بمانند و بخصوص در مواردی که سؤالات حساس و خصوصی پرسیده می‌شود، پاسخهای صادقانه تری دریافت می‌شود.

### /// پایایی

می‌توان از دو پرسشگر مختلف خواست که با فرد واحدی مصاحبه کنند تا همسانی نتایج کنترل شود. این یکی از روشهای سنجش پایایی مصاحبه‌ها و پرسشنامه هاست. برای کنترل ثبات همسانی درونی می‌توان زوایدی را به ابزار گردآوری داده‌ها اضافه کرد، یعنی می‌توان جمله بندی سؤالات مربوط به همان موضوع را تغییر داد و در پرسشنامه و یا مصاحبه تکرار کرد.

می‌توان پرسشنامه و مصاحبه‌ای را با همان افراد بعد از مدتی تکرار و یا اینکه دو شکل پرسشنامه را با همان افراد تکمیل کرد. این گونه شیوه‌های کنترل معمولاً پرهزینه و وقت‌گیر و به نحوی غیر عملی هستند؛ زیرا یافتن آزمودنیهایی که مایل باشند پرسشنامه یا مصاحبه را برای بار دوم تکرار کنند، آسان نیست. مسئله دیگری که در مورد این روش وجود دارد، این است که برخی از سؤالات به جنبه‌هایی از رفتار مربوط می‌شوند که ثبات کمتری دارد و تغییر آن در طول زمان می‌تواند قابل قبول باشد.



### /// تحلیل آماری در تحقیقات پیمایشی

تحقیقات پیمایشی توصیفی معمولاً نیازی به تحلیل آماری پیچیده ندارند. تحلیل داده ها می تواند به سادگی شامل تعیین فراوانی و یا درصد متغیرهای اصلی بررسی باشد. برای مثال، تحقیق پیمایشی که به منابع کتابخانه ای می پردازد، می تواند تعداد کتابهای داستان و تعداد کتابهای غیر داستانی و غیره را گزارش کند، یعنی یک جدول بندی ساده کافی است. تحقیق پیمایشی که به نگرشهای مردم نسبت به مسئله ای می پردازد، می تواند تعداد و درصد افرادی را که به طبقات مختلف پاسخها تعلق دارند، گزارش کند.

از طرفی، در تحقیقات پیمایشی اکتشافی، تمایل به یافتن رابطه بین متغیرهای بررسی وجود دارد، یعنی علاقه مندیم بدانیم که آیا  $X$  و  $Y$  با یکدیگر تغییر می کنند و تحت چه شرایطی با یکدیگر تغییر می کنند. به طور کلی، رابطه را با ترتیب دادن توزیعهای فراوانی یک متغیر در مقابل متغیر دیگر، از طریق جدول بندی تقاطعی، مورد بررسی قرار می دهند. ساده ترین جدول تقاطعی شامل دو متغیر با دو طبقه بندی برای هر یک از متغیرهاست. صور پیچیده تر نیز مانند  $3 \times 4$  و  $2 \times 4$  و  $2 \times 3$  و غیره، می تواند وجود داشته باشد، ولی تحلیل و تفسیر فراوانیها با اضافه شدن تعداد متغیرها، مشکلتر می شود. یک جدول تقاطعی نشان می دهد که ترکیبهای متفاوت متغیرها، چندبار تکرار می شود، که محقق می تواند رابطه بین متغیرها را از آن میان مشاهده کند. به منظور معتبر نشان دادن قدرت رابطه، معمولاً فراوانیها به صورت درصد در می آید. جداول تقاطعی را می توان تقریباً برای هر نوع داده ای به کار برد، ولی معمولاً این گونه جداول برای داده های طبقه بندی یا اسمی به کار برده می شوند.

نتایج یک تحقیق پیمایشی فرضی را در زمینه نگرشها نسبت به مالیات جدید برای کتابخانه ها، که به وسیله آن کتابخانه های کشور وسعت داده می شوند و بهبود می یابند، در نظر بگیرید. جدول تقاطعی فراوانیهای دو متغیر، در جدول ۲-۱۲ آمده است.

جدول ۲-۱۲ نشان می دهد که  $62/5$  درصد  $(\frac{150}{240})$  ساکنان شهرها با مالیات کتابخانه موافقت و در مقابل  $36/9$  درصد  $(\frac{96}{260})$  ساکنان روستاها با آن موافق هستند. قاعدتاً محقق باید از داده های این جدول چنین نتیجه گیری کند که رابطه مستقیم بین نگرش نسبت به مالیات کتابخانه و محل سکونت وجود دارد.

جدول ۲-۱۲ نگرش نسبت به مالیات کتابخانه به تفکیک محل سکونت

شهر	روستا	کل	
۱۵۰	۹۶	۲۴۶	موافق
۹۰	۱۶۴	۲۵۴	مخالف
۲۴۰	۲۶۰	۵۰۰	کل

اکثریت ساکنان شهرها با مالیات کتابخانه موافق و اکثریت ساکنان روستاها با آن مخالف هستند، لکن مشاهده‌گر آگاهتر، ممکن است به این نکته اشاره کند که شهر (محل یک دانشگاه ایالتی مهم) درصد بالاتری افراد تحصیل کرده دارد و اینکه ممکن است سطح تحصیل و نه محل سکونت، به خودی خود عامل ابراز نگرش مثبت نسبت به مالیات کتابخانه باشد. به عبارت دیگر، احتمال دارد متغیر سومی که همان سطح تحصیل است، وجود داشته باشد؛ زیرا احتمالاً افراد با سطح تحصیل بالاتر در شهرها و نزدیک دانشگاه‌های محل کارشان زندگی می‌کنند و همین افراد هم مدافع مالیات کتابخانه هستند.

به منظور تحقیق در مورد این تبیین، نیاز به کنترل متغیر سوم داریم. برای کنترل متغیر سومی که در یک پیمایش با دو متغیر اصلی رابطه دارد، ضروری است که آن متغیر را ثابت نگاهداریم و ببینیم آیا بین دو متغیر اول و دوم، در صورتی که متغیر سوم تغییر نکند، باز هم رابطه‌ای وجود دارد یا خیر.

ساده‌ترین راه ثابت نگاهداشتن یک متغیر این است که آزمودنیها را به دو گروه مختلف تقسیم کنیم؛ به صورتی که هر یک مقدار متفاوتی از آن متغیر را دارا باشند و همان تحلیل را بر روی هر یک از این گروهها انجام دهیم. یعنی، محقق می‌بیند که آیا در داخل هر یک از گروهها، بین دو متغیر اول و دوم رابطه وجود دارد یا خیر. چنانچه هنوز بین دو متغیر رابطه وجود داشته باشد، دیگر نمی‌توان آن را ناشی از تغییرات متغیر سوم دانست؛ زیرا در داخل گروهها، متغیر سوم ثابت نگاهداشته شده است. برای کنترل متغیر سوم (سطح تحصیل) در مثال فوق، پاسخگویان را در گروههای با تحصیلات دانشگاهی و بدون تحصیلات دانشگاهی (با فرض اینکه چنین اطلاعی در دست باشد) طبقه بندی و رابطه بین این دو گروه مجزا را بررسی می‌کنیم. جدول ۳-۱۲ این اطلاعات را نشان می‌دهد.

## جدول ۱۲-۳ نگرش نسبت به مالیات کتابخانه، نامرتبط با تحصیلات

## افراد با تحصیلات دانشگاهی

شهر	روستا	کل	
۱۴۴	۸۰	۲۲۴	موافق
۳۶	۲۰	۵۶	مخالف
۱۸۰	۱۰۰	۲۸۰	کل

## افراد بدون تحصیلات دانشگاهی

شهر	روستا	کل	
۶	۱۶	۲۲	موافق
۵۴	۱۴۴	۱۹۸	مخالف
۶۰	۱۶۰	۲۲۰	کل

## جدول ۱۲-۴ نگرش نسبت به مالیات کتابخانه، مرتبط با تحصیلات

## افراد با تحصیلات دانشگاهی

شهر	روستا	کل	
۸۰	۲۰	۱۰۰	موافق
۴۰	۴۰	۸۰	مخالف
۱۲۰	۶۰	۱۸۰	کل

## افراد بدون تحصیلات دانشگاهی

شهر	روستا	کل	
۷۰	۷۶	۱۴۶	موافق
۵۰	۱۲۴	۱۷۴	مخالف
۱۲۰	۲۰۰	۳۲۰	کل

چندین برنامه کامپیوتری که جداول تقاطعی را ارائه می‌دهند، آماده‌هایی را به این منظور ارائه می‌کنند.

اطلاعات مندرج در جدول ۳-۱۲ نشان می‌دهند که تبیین، صحیح است؛ بدین معنا که در صورت کنترل تحصیلات دانشگاهی با ثابت نگاهداشتن آن، رابطه‌ای بین متغیرهای «محل سکونت» و «نگرش نسبت به مالیات» وجود ندارد. در بین تحصیلکردگان دانشگاه، ۸۰ درصد ( $\frac{144}{180}$ ) ساکنان شهرها و به همین ترتیب ۸۰ درصد ( $\frac{8}{110}$ ) ساکنان روستاها، با مالیات موافق هستند. در بین افرادی که تحصیلات دانشگاهی ندارند، ۱۰ درصد ( $\frac{6}{60}$ ) ساکنان شهرها و همچنین ۱۰ درصد ( $\frac{16}{160}$ ) ساکنان روستاها، با مالیات موافقت می‌کنند.

فرض کنید داده‌ها به صورتی باشد که در جدول ۴-۱۲ نشان داده شده است. هرچند سطح تحصیلات در هر یک از جداول ثابت نگاهداشته شده است، رابطه بین متغیرهای محل سکونت و نگرش، مشهود است. در بین افراد با تحصیلات دانشگاهی که در شهرها سکونت دارند،  $\frac{66}{7}$  درصد ( $\frac{8}{110}$ ) و در بین این گروه افراد که در روستاها سکونت دارند  $\frac{33}{3}$  درصد ( $\frac{4}{60}$ )، با مالیات موافق هستند. در بین افرادی که تحصیلات دانشگاهی ندارند،  $\frac{58}{3}$  درصد ساکنان شهرها و ۳۸ درصد ساکنان روستاها، با مالیات موافق هستند. در این مورد، آنچه موجب می‌شود که پاسخگویان احساس متفاوتی نسبت به مالیات کتابخانه داشته باشند، شهری و یا روستایی بودن است، نه تحصیلات دانشگاهی.

ممکن است لازم باشد متغیرهای بسیاری را در یک تحقیق پیمایشی کنترل کنیم. برای مثال، جنسیت باید قبل از هر توصیفی برای یک رابطه، کنترل شده باشد. در این مورد، محقق می‌تواند آزمودنیها را به دو گروه مرد و زن، تقسیم و داده‌های مربوط به هر گروه را جداگانه بررسی کند. طبقه اجتماعی را می‌توان با طبقه بندی آزمودنیها در گروههای مختلف که بر مبنای معیارهای مناسب شکل گرفته‌اند، کنترل و اطلاعات را در داخل هر یک از گروهها تحلیل کرد. در زمان طرح‌ریزی یک تحقیق پیمایشی، باید تا آنجا که ممکن است کنترل متغیرها پیش‌بینی و سوالات لازم برای چنین تحلیلهایی آورده شود.

### // آماره‌های جداول تقاطعی

محقق می‌خواهد رابطه مشهود در جداول تقاطعی را با اندازه‌ای که نمایانگر مقدار پیوستگی بین متغیرهاست و یا با یک آزمون معنی داری آماری خلاصه کند. انتخاب اندازه پیوستگی، به سطحی که دو متغیر مندرج در جدول تقاطعی در آن سطح اندازه‌گیری شده‌اند، بستگی دارد.

در صورتی که هر دو متغیر جدول تقاطعی در سطح اسمی اندازه گیری شده باشد، آزمون خی دو را می توان برای تعیین اینکه آیا رابطه نظام مند بین دو متغیر وجود دارد یا خیر، به کار برد (برای بحث خی دو به فصل ۶ مراجعه کنید). لکن، خی دو تنها نشان می دهد که آیا دو متغیر به یکدیگر مربوط و یا مستقل از یکدیگر هستند و میزان رابطه آنها را مشخص نمی کند. برای مثال مقدار  $\chi^2$  برای داده های جدول ۴-۱۲ برابر با ۱۶/۶۷ است که از نظر آماری در سطح ۰/۰۱ معنی دار است.

در نتیجه، می دانیم که در بین افراد با تحصیلات دانشگاهی، رابطه معنی داری بین محل سکونت و نگرش نسبت به مالیات وجود دارد، ولی نمی دانیم که دو متغیر با چه قدرتی با یکدیگر رابطه دارند. در مورد جداول  $2 \times 2$  نظیر مثالهایی که در این فصل آمده است، آماره فی ( $\phi$ ) می تواند به عنوان اندازه ای برای قدرت رابطه، به کار برده شود. در صورتی که هیچ رابطه ای وجود نداشته باشد، مقدار فی برابر صفر و وقتی متغیرها کاملاً با یکدیگر رابطه دارند، مقدار آن ۱+ خواهد بود. برای جداول بزرگتر از  $2 \times 2$ ، اندازه مناسب رابطه، ضریب وابستگی (C)<sup>(۱)</sup> است. حداقل مقدار آن صفر است، ولی حداکثر مقدار آن می تواند هر مقداری باشد و بستگی به اندازه جدول دارد. به این دلیل، C تنها برای مقایسه جداولی می تواند مورد استفاده قرار گیرد که تعداد ستونها و ردیفهای آنها برابرند.<sup>(۲)</sup> در صورتی که جداول تقاطعی دو بعدی بزرگتر از  $2 \times 2$  باشند و دو متغیر، اسمی باشند، می توان از آماره کرامر<sup>(۳)</sup> که تعمیم ریاضی فی است، استفاده کرد. در صورتی که دو متغیر جدول تقاطعی، رتبه ای باشند، می توان از دیگر آماره ها - مانند گاما، Tau B و Tau C، کندال یا D سامر - برای نشان دادن میزان رابطه استفاده کرد. برای مطالعه بیشتر در زمینه آماره کرامر و اندازه های فوق، خواننده به کتابهای درسی آمار ارجاع داده می شود.<sup>(۴)</sup>

### /// خلاصه

تحقیق پیمایشی روش تحقیقی برای گردآوری اطلاعات در زمینه های مختلف، از شمارش

1- Contingency Coefficient (C)

۲- برای مطالعه بیشتر در زمینه محاسبه فی و ضریب وابستگی مراجعه شود به کتاب:

D. Ary and L.C. Jacobs (1976), Introduction to Stabcs (ch.9), New York, Holt, Rinehart and Winston, for a Discussion of the Computation of phi and the Contingency Coefficient. 3- Cramer,s statistic

4- Hays, W.L. (1988), Statistics (4th ed), New York: Hold, Rinehart and Winston

عینی و فراوانی تا نگرشها و نظرات است. تحقیقات پیمایشی بر مبنای هدف، کانون توجه و حیطة آنها طبقه بندی می‌شوند. این گونه تحقیقات نیاز به برنامه‌ریزی دقیق، نمونه‌گیری نا اریب از جامعه آماری، ساختن ابزار گردآوری داده‌ها و تحلیل دقیق نتایج دارد.

چنانچه محققان از نوع نمونه‌گیری احتمالاتی استفاده کنند، می‌توانند از نتایج مربوط به نمونه، مقادیر مربوط به جامعه آماری را استنتاج کنند. شیوه معمول عبارت است از تعیین فاصله اطمینانی که مقادیر جامعه آماری با بیشترین احتمال در آن قرار می‌گیرد. پهنای فاصله، تابعی از میزان خطای مورد پذیرش محقق و حجم نمونه است. با کاهش احتمال درست بودن و افزایش تعداد نمونه، فاصله اطمینان، کوچک می‌شود. محقق همچنین می‌تواند از این شیوه برای تخمین زدن تعداد نمونه مورد نیاز برای سطح مطلوب دقت، استفاده کند.

وسایل اصلی گردآوری داده‌ها در تحقیقات پیمایشی، مصاحبه‌ها و پرسشنامه‌ها هستند. در هر دو شیوه از آزمودنیهای منتخب سؤال می‌کنند، ولی هر یک نقاط قوت و ضعف خود را دارند. پایا بودن وسیله‌های گردآوری داده‌ها امری حائز اهمیت است. این امر ثابت شده است که راههای مختلف پیگیری، برگشت پرسشنامه‌های پستی را افزایش می‌دهد.

ساختن جداول تقاطعی راهی عالی برای نشان دادن رابطه موجود بین متغیرهای یک پیمایش است.

### // مفاهیم اصلی

پیش‌آزمون کردن	افرادى که پاسخ نداده‌اند (بی‌جوابها)
پیمایش نمونه‌ای	پیگیری
تحقیق پیمایشی توصیفی	تحقیق پیمایشی تبیینی
جهت دار بودن مصاحبه‌کننده	جداول تقاطعی
سرشماری	حاشیه خطا
سوالات باز	سوالات از نوع لیکرت
سوالاتی که رتبه بندی می‌کنند	سوالات مقیاسی
فهرست واریسی	سوالات دو وجهی
مصاحبه	مجموعه پاسخ

### /// تمرینها

۱- یک سؤال تحقیقاتی را پیشنهاد کنید که بهترین راه جواب دادن به آن، تحقیق پیمایشی باشد.

۲- تفاوت بین تحقیقات پیمایشی توصیفی و تبیینی را شرح دهید.

۳- بعضی از عواملی که بر تعداد نمونه اثر می‌گذارند، کدامند؟

۴- خبر شب تلویزیون گزارش کرد: در یک تحقیق پیمایشی که اخیراً در جامعه آمریکا انجام شد، ۴۵ درصد پاسخگویان ابراز داشتند که عملکرد رئیس جمهور را تأیید می‌کنند. قبل از آنکه بتوانید بر روی این گزارش تفسیری بدهید، به دانستن چه چیز دیگری نیاز دارید؟

۵- گزارش زیر را چگونه تفسیر می‌کنید؟ «نظرخواهی از نمونه‌ای ۱۰۰۰ نفری که به صورت تصادفی از بین رأی دهندگان ثبت شده در ایندیانا انجام شد، نشان داد که ۳۷ درصد با بخت آزمایی ایالتی موافقتند. ارقام این نظرخواهی در معرض خطای نمونه‌گیری معادل  $\pm 3$  درصد است. سطح اطمینان ۹۵ درصد است.»

۶- فاصله اطمینان را برای نسبت جامعه در سطح ۹۵ درصد برای هریک از سه نمونه زیر با تعداد متفاوت محاسبه کنید. فرض کنید که نسبت نمونه برای هریک  $0/40$  باشد. افزایش نمونه چه تأثیری بر عریض شدن فاصله دارد؟ چرا؟

تعداد افراد نمونه A، ۱۰۰ نفر، تعداد افراد نمونه B، ۱۰۰۰ نفر و تعداد افراد نمونه C، ۱۰۰۰۰ نفر است.

۷- شما مقیاسی را برای اندازه‌گیری «خستگی مفرط» معلمان ساخته‌اید. نمونه‌ای ۱۰۰ نفری از معلمان که در سیستم آموزشی شهرهای بزرگ شاغل هستند، به صورت تصادفی انتخاب شده‌اند و نمره میانگین آنها  $10/5$  و انحراف استانداردشان  $2/3$  است. برآورد شما از میانگین نمره «خستگی مفرط» کل معلمان در سیستم آموزشی چیست؟ از سطح اطمینان ۹۵ درصد استفاده کنید.

۸- یک سازمان نظرسنجی ملی می‌خواهد نتیجه انتخابات ریاست جمهوری را در محدوده  $\pm 5$  درصد، پیش بینی کند. تعداد نمونه چقدر باید باشد تا بتوان این سطح دقت را به دست آورد؟ سطح اطمینان ۹۵ درصد را در نظر بگیرید.

۹- یک تحقیق پیمایشی میزان پاسخ اولیه‌ای برابر ۵۱ درصد داشت. به محقق چه

پیشنهادهایی برای رفع اشکال درصد پایین پاسخ می‌دهید؟

۱۰- در دانشگاه ایالتی علاقه به انتخاب رشته بازرگانی طی چند سال اخیر رو به کاهش گذارده است. بخش بازرگانی می‌خواهد بداند که آیا دانشجویان سال اول علاقه‌ای به انتخاب این رشته دارند یا خیر. از آنجا که امکانات مالی برای انجام پیمایش بر روی ۴۵۰۰ دانشجوی سال اول وجود ندارد، آنها پیمایش را با نمونه‌ای ۵۰۰ نفری و تصادفی انجام می‌دهند و به این نتیجه می‌رسند که براساس اعلام دانشجویان، ۱۱۰ دانشجو به این رشته علاقه مندند. در سطح اطمینان ۹۵ درصد، برآورد شما از تعداد دانشجویان سال اولی که به ادامه تحصیل در رشته بازرگانی علاقه مند هستند، چیست؟

۱۱- کدام یک از موارد زیر نمونه‌هایی اریبی از جامعه دانشجویان دانشکده‌های دانشگاههای بزرگ است؟

الف - نمونه‌ای تصادفی از دانشجویانی که غروب جمعه به کتابخانه وارد می‌شوند.  
 ب - نمونه‌ای تصادفی از دانشجویانی که برای کلاسها نام نویسی کرده‌اند.  
 ج - نمونه‌ای تصادفی از دانشجویانی که برای مسابقات فصلی، بلیط خریده‌اند.  
 د - نمونه مرکب از دانشجویانی که بعد از دیدن یک آگهی در روزنامه دانشگاه، داوطلب شده‌اند.  
 ۱۲- به کتابخانه بروید و مقاله تحقیقاتی زیر را که فن تحقیق پیمایشی را نشان می‌دهد، پیدا کنید: آگاهی دانشجویان نسبت به مسئله استفاده از الکل و رفتارهای افتراقی مشروبخواری، استرنج و اشعیت، نشریه دانشکده.

الف - مسئله تحقیق چیست؟

ب - تعداد افراد نمونه چه بوده و نمونه چگونه انتخاب شده است؟ زمینه انجام پیمایش چه بوده است؟

ج - وسیله گردآوری داده‌ها را توصیف کنید. این وسیله برای گردآوری چه نوع داده‌هایی طراحی شده بود؟

د - داده‌ها به چه صورتی گزارش شده‌اند؟ چه تحلیل آماری بر روی داده‌ها انجام شده است؟  
 ه - یافته‌های اصلی پیمایش را بنویسید.

و - آیا نقایصی در طرح یا روش شناسی این پیمایش وجود دارد که ممکن است بر اعتبار یافته‌ها، اثر گذارد؟



### /// پاسخها

- ۱- پاسخها می تواند متنوع باشد.
  - ۲- پیمایشهای توصیفی سعی بر آن دارند آنچه را که موجود است، بدون چرایی آن تبیین کنند. پیمایشهای تبیینی در پی تبیین یافته هاست.
  - ۳- برخی از عوامل عبارتند از: همگنی جامعه آماری، تعداد طبقات مشاهده‌ای که انجام خواهد شد، نوع طرح نمونه‌گیری مورد استفاده و درجه دقت مورد نظر محقق برای برآورد پارامتر جامعه آماری.
  - ۴- فرد می خواهد تعداد افراد نمونه، چگونگی انتخاب آنها، پهنای فاصله حول پارامتر برآورد شده جامعه آماری (حاشیه خطا) و سطح اطمینان را بداند.
  - ۵- احتمال اینکه میزان رأی دهندگان ثبت شده، که با یک بخت آزمایشی ایالتی موافقت، بین ۳۴ و ۴۰ درصد باشد، ۰/۹۵ است.
- ۶-

نمونه C

نمونه B

نمونه A

$$\begin{aligned} \text{خطای معیار} &= \sqrt{\frac{(0/4)(0/6)}{10,000}} = 0/005 & \text{خطای معیار} &= \sqrt{\frac{(0/4)(0/6)}{1000}} = 0/0155 & \text{خطای معیار} &= \sqrt{\frac{(0/4)(0/6)}{100}} = 0/049 \\ \text{فاصله} &= 0/40 \pm 1/96(0/005) & \text{فاصله} &= 0/40 \pm 1/96(0/0155) & \text{فاصله} &= \pm 1/96(0/005) \\ &= 0/40 \pm 0/01 & &= 0/40 \pm 0/03 & &= 0/40 \pm 0/10 \\ &= 0/39 \text{ تا } 0/41 & &= 0/37 \text{ تا } 0/43 & &= 0/3 \text{ تا } 0/5 \end{aligned}$$

افزایش تعداد افراد نمونه، پهنای فاصله اطمینان را کاهش می دهد. با افزایش تعداد افراد نمونه، خطای نمونه و در نتیجه مقدار فاصله اطمینان کاهش می یابد.

۷-

$$\text{خطای معیار} = \frac{S}{\sqrt{n}} = \frac{2/3}{\sqrt{100}} = \frac{2/3}{10} = 0/23$$

$$\bar{X} \pm 1/96(0/23) = 10/5 \pm 0/45$$

یا بین ۱۰/۰۵ و ۱۰/۹۵

۸- فرض کنید  $p = q = 0.5$ 

$$1/96 \sqrt{\frac{(0.50)(0.50)}{n}} = 0.05$$

$$\sqrt{\frac{0.25}{n}} = \frac{0.05}{1/96}$$

$$\frac{0.25}{n} = \left(\frac{0.05}{1/96}\right)^2$$

$$\frac{0.25}{n} = 0.0065$$

یا ۳۸۵ نفر،  $n = 384/6$ 

۹- محقق باید با ارسال یک کارت پستال یادآوری، پیگیری و سپس یک یا دو پرسشنامه دیگر را ارسال کند. بعد از خاتمه مرحله پیگیری، محقق باید سعی کند با عده‌ای از کسانی که پرسشنامه را بدون پاسخ گذاشته‌اند، مصاحبه کند و اطلاعاتی را در مورد ویژگیهایشان به دست آورد و پاسخهایشان را به منظور تعیین اینکه آیا به صورت معنی‌دار با پاسخگویان تفاوت دارند یا خیر، دریافت دارد.

-۱۰

$$p = 0.22 \left(\frac{110}{500}\right)$$

$$q = 0.78$$

$$s.e. = \sqrt{\frac{(0.22)(0.78)}{500}} = 0.0185$$

$$\text{فاصله} = 0.22 \pm 1/96 [0.0185] = 0.22 \pm 0.4$$

می‌توان انتظار داشت ۱۸ تا ۲۶ درصد، یا ۸۱۰ تا ۱۱۷۰ نفر از دانشجویان، به ادامه تحصیل در رشته بازرگانی علاقه‌مند باشند.

۱۱- نمونه‌های (الف)، (ج) و (د) معرف جامعه دانشجویان دانشکده در یک دانشگاه بزرگ نیستند.

۱۲- الف - نگرش دانشجویان نسبت به استفاده خودشان از الکل و درک خودشان از رفتار مشروبخواری در میان دانشجویان همکلاسی چیست؟

ب - یک نمونه ۵ درصدی (تعداد افراد نمونه = ۱۱۵۰ نفر) از دانشجویان دانشگاه آیووا<sup>(۱)</sup> به صورت تصادفی و بر مبنای چهار شماره آخر کارت شناسایی دانشجویی خود انتخاب شدند.

ج - پرسشنامه‌ای پستی به منظور گردآوری داده‌ها در زمینه میزان استفاده از الکل، دلایل استفاده از الکل، نگرشهای پاسخگویان نسبت به عواقب مشروبخواری، آگاهی نسبت به مسئله تخدیر با الکل و آگاهی دانشجویان نسبت به منابع یاری‌دهنده در جهت سوء استفاده از الکل، تهیه شد.

د - تعداد و درصد برای کل گروه و برای زنان و مردان به صورت جداگانه محاسبه شد. به منظور آزمون رابطه بین پاسخها و جنسیت برخی متغیرهای خاص دیگر، از آزمون خی‌دو استفاده شد.

ه - اکثریت بالایی از دانشجویان به نحوی مشروب الکلی می‌نوشیدند، که از بین آنها ۵۵ درصد حداقل یک بار در هفته به این کار مبادرت می‌کردند. اکثریت در هر بار بین ۱ تا ۳ لیوان و ۲۹ درصد ۴ تا ۶ لیوان مشروب می‌نوشیدند. بین زنان و مردان، تفاوت معنی‌داری در نوع استفاده از مشروبات الکلی و فراوانی استفاده و مقدار نوشیدن، وجود داشت. دانشجویان در مورد تخدیر با الکل آگاهی کمی داشتند. بیش از یک سوم دانشجویان نسبت به عواقب بلند مدت نوشیدن مشروبات الکلی اظهار نگرانی کردند و تقریباً به همین میزان ابراز داشتند که دوست دانشجویی دارند که احتمالاً مشکل مشروب‌خواری دارد. بیش از ۳/۴ دانشجویان، در صورتیکه احساس می‌کردند مشکل مشروب‌خواری دارند، مرجعی خارج از دانشگاه را برای مشاوره ترجیح می‌دادند.

و - میزان بازگشت پرسشنامه‌ها تقریباً پایین بود (۵۷ درصد)، ولی محققان افرادی را که پاسخ نداده بودند، مورد بررسی قرار دادند و به این نتیجه رسیدند که آنها از نظر جنسیت، جایگاه طبقاتی و نوع سکونت، مشابه کسانی بودند که پاسخ داده بودند. یافته‌ها همچنین بر مبنای اطلاعات گزارش شده توسط خود فرد، گردآوری شده بود.



## فصل سیزدهم

### تحقیقات کیفی و تاریخی

#### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- تفاوت بین تحقیقات کمی و کیفی را تمیز دهد.
- ۲- شیوه‌هایی را که براساس آنها، تحقیق کیفی قابل اعتمادی به دست می‌آید، توصیف نماید.
- ۳- جنبه‌های حائز اهمیت تحقیق کیفی را توصیف کند.
- ۴- ماهیت مورد پژوهیها و نقاط ضعف و قوت آنها را توصیف کند.
- ۵- تمایز بین مورد پژوهیها، مطالعات طبیعی و آزمایشهای تک آزمودنی را درک کند.
- ۶- ویژگیهای تحقیق تاریخی را توصیف کند.
- ۷- بین منابع اولیه و ثانویه در تحقیق تاریخی، تمایز قایل شود.
- ۸- نقد داخلی و خارجی را تعریف و برای آنها مثالهایی ذکر کند.

در روشهای تحقیقی که در فصول ۹ تا ۱۲ توصیف کردیم، برای پاسخگویی به مسئله تحقیق از اعداد استفاده می‌شود. از آنجا که این روشهای تحقیق از اندازه‌های کمی، چون شمارشها، میانگینها، همبستگیها و آزمونهای t استفاده می‌کنند، می‌توان آنها را تحت عنوان تحقیق کمی طبقه‌بندی کرد. بالعکس، تحقیق کیفی برای پاسخگویی به مسئله تحقیق، از واژه‌ها استفاده می‌کند. تحقیق تاریخی اغلب ماهیتی کیفی دارد، لکن در صورت لزوم از ارزشهای کمی هم استفاده می‌کند.

### /// بررسی کیفی

عبارت بررسی کیفی، اصطلاحی عام برای رویکردهای مختلف تحقیق در زمینه تعلیم و تربیت و ارزشیابی تلقی می‌شود که با عناوین مختلفی چون قوم‌نگاری، بررسی طبیعی، مورد پژوهیها، پژوهش میدانی، بررسیهای میدانی و مشاهده توأم با مشارکت خواننده می‌شوند<sup>(۱)</sup>. اگرچه این رویکردها بر حسب سنن مختلف فلسفی و تحلیلی قابل تفکیک هستند<sup>(۲)</sup>، اما دارای ویژگیهای مشترکی هستند که آنها را از رویکرد کمی مورد استفاده در تحقیقات آموزشی و ارزشیابی که در فصول ۱، ۲ و ۴ رتوس مطالب آن آورده شد و روشهایی که در فصول ۹ تا ۱۲ توصیف شد، جدا می‌سازد.

### // تمایز بین بررسیهای کیفی و کمی

بررسی کیفی از این نظر با روش کمی در بررسی پدیده‌های اجتماعی و رفتاری متفاوت است که ادعای شیوه کمی را مبنی بر اینکه حداقل در اصول، هدف و روشهای علوم اجتماعی همان هدف و روشهای علوم طبیعی یا مادی هستند، رد می‌کند.<sup>(۳)</sup> محقق طرفدار بررسیهای

1- Y. S. Lincoln and E. G. Guba (1985), *Naturalistic Inquiry*, Beverly Hills: Sage Publications; J. P. Goetz and M. D. LeCompte (1984), *Ethnography and Qualitative Design in Educational Research*, New York: Academic Press; S. B. Merriam (1988), *Case Study Research in Education*, San Francisco: Jossey-Bass; R. G. Bogdan and S. K. Biklen (1982), *Qualitative Research for Education*, Boston: Allyn and Bacon; and J. Spradley (1980), *Participant Observation*, New York: Holt, Rinehart and Winston.

2- Jacob, E. (1988), Clarifying qualitative research: A focus on traditions. *Educational Researcher*, 17, 16-24; and P. Atkinson, S. Delamont, and M. Hammersley (1988), qualitative research traditions: A British response to Jacob *Review of Educational Research*, 58, 231-250.

۳- دیدگاهی که علوم اجتماعی را دارای همان اهداف و روشهای بررسی علوم طبیعی می‌داند، طبیعت‌گرایی یا تفسیر طبیعت‌گرایانه علوم اجتماعی می‌نامند. در این مورد اغلب اشتباهی پیش می‌آید؛ زیرا بوعی از بررسی کیفی نیز بررسی طبیعی خوانده می‌شود. در این مورد، واژه طبیعی معنایی ←

کیفی چنین بحث می‌کند که هر دو علوم طبیعی و اجتماعی در جهت نظریه‌های قابل آزمون و تأیید شدنی کوشش می‌کنند تا بتوانند پدیده‌ها را با نشان دادن چگونگی استنتاج آنها از فرضیات نظری، توضیح دهند (رجوع شود به بحث نظریه علمی در فصل ۱). به عبارت دیگر، هدف هر دو نوعی تبیین علمی است، که مشتمل بر اکتشاف و استیناف قوانین است؛ قوانینی که از طرفی بر رفتار جهان مادی حکومت می‌کند و از طرف دیگر حاکم بر رفتار انسانی است.

بررسی کیفی یا فرض روش شناختی متفاوتی شروع می‌شود، یعنی موضوع علوم اجتماعی و انسانی اساساً با موضوع علوم مادی یا طبیعی متفاوت است و در نتیجه به هدفی متفاوت برای بررسی و مجموعه‌ای از روشهای تحقیق، نیازمند است. پژوهشگران کیفی چنین بحث می‌کنند که همواره رفتار انسانی مقید به زمینه‌ای است که در آن اتفاق افتاده است (واقعیت جامعه)؛ برای مثال، فرهنگها، موضوعهای فرهنگی، نهادها و چیزهای مشابه، نمی‌تواند همانند واقعیت مادی به متغیرهایی محدود شود و آنچه در رشته‌های علوم اجتماعی مهم‌ترین است، درک و تصویر سازی از آن معنایی است که توسط شرکت کنندگان درگیر در موقعیت و با رویدادهای خاص اجتماعی، ساخته می‌شود. بررسی کیفی تلاش می‌کند رفتارهای انسانی و اجتماعی را از دیدگاه «خودبها»، درک کند، یعنی دیدگاهی که شرکت کنندگان در یک موقعیت خاص اجتماعی آن را تجربه کرده باشند. (برای مثال، یک فرهنگ، مدرسه، جامعه، گروه و نهاد). این نوع تحقیق، یک تحقیق به شدت شخصی است، تحقیقی که «درک ذهنی و پیشداوری شرکت‌کنندگان و محقق را در چهارچوب تحقیق» آزادانه می‌پذیرد و به آن اذعان دارد.<sup>(۱)</sup>

طرفداران رویکردهای کیفی چنین می‌گویند که در مقابل، بررسی کمی اصولاً با کشف «واقعیت‌های اجتماعی» جدای از ادراکات ذهنی و نیت و منفک از بافت اجتماعی و تاریخی خاص، سروکار دارد.

رویکردهای کمی در علوم انسانی بر مبنای مدل تبیین فرضی - قیاسی استوار است. تحقیق کمی با نظریه‌ای در مورد پدیده‌های مورد بررسی شروع می‌شود. از آن نظریه تعدادی فرضیه استنتاج می‌شوند که به نوبه خود با استفاده از روشهای از قبل تعیین شده، مانند طرحهای آزمایشی، علی - مقایسه‌ای و همبستگی، مورد آزمون قرار می‌گیرند. هدف غایی استفاده از مدل

→ تفاوت دارد و به این وابست بر می‌گردد که بررسیهای کیفی در ضمنی «طبیعی» (در مقابل نوعیت ساخته شده) انجام می‌شوند. این بحث در بخش «جوانب حائز اهمیت بررسی کیفی» در صفحات ۶۰۸-۶۰۱ ادامه خواهد داشت.

1- Goeltz and LeCompte, *Ethnography and Qualitative Design in Educational Research*, p. 95.

فرضی - قیاسی عبارت از بازنگری و تأیید نظریه‌ها و یا بیان احکام قانونمند در مورد پدیده‌های اجتماعی و رفتاری است که بر مبنای نتایج آزمون فرضیه استوار هستند. نظریه‌ها پالایش می‌شوند و بسط می‌یابند (و گاهی رد می‌شوند) تا علل و دلایل نتایج حاصله از آزمون معانی ضمنی یا موارد آنها را بیان کنند (قیاسها).

بررسی کیفی بر مدل متفاوتی از تبیین استوار است و بر هدف متفاوتی از بررسی دلالت دارد. به طور کلی، بررسی کیفی بر این باور است که رفتن به دنبال تعمیم نتایج (احکام قانونمند یا نظریه‌هایی که در طول زمان و مکان تغییر ناپذیر باقی می‌مانند) نادرست است. رفتار بشری همواره در قید یک بافت تاریخی، اجتماعی، زمانی و فرهنگی است؛ بنابراین، تبیین از نوع قانون - و موارد - آن، به سود تبیین از نوع موضوعات - و تفاسیر - آنها کنار نهاده می‌شود.<sup>(۱)</sup> بررسی کیفی به دنبال تفسیر کنشها، نهادها، رویدادها، رسوم انسان و نظایر آنهاست و در این راستا از آنچه مورد بررسی است، یک نوشتار یا تصویر می‌سازد. هدف غایی این نوع بررسی، به تصویر کشیدن الگویی پیچیده از چیزهای مورد بررسی، با عمق و جزئیات کافی است؛ به گونه‌ای که فرد، بدون آنکه آن را تجربه کرده باشد، درک کند. زمانی که بررسیهای کیفی به تفسیر یا تبیین معانی رویدادها، کنشها و غیره می‌پردازند، معمولاً از یکی از انواع این تفسیرها استفاده می‌کنند:

۱- ساختن الگو از طریق تحلیل و ترکیب دوباره اجزای متشکله ۲- تفسیر معانی اجتماعی رویدادها یا تحلیل رابطه بین رویدادها و عوامل خارجی.<sup>(۲)</sup>

رویکردهای کمی و کیفی همچنین دیدگاههای مختلفی نسبت به ارزشها در بررسی دارند. محققان کمی می‌پذیرند که ارزشهای محقق ممکن است در انتخاب عنوان و مسئله مورد بررسی نقش داشته باشد، اما تحقیق واقعی خود باید فارغ از ارزشها باشد، یعنی محقق باید شیوه‌ای را در پیش گیرد که به طور مشخص برای جداسازی و حذف تمام عناصر ذهنی، نظیر ارزشها، از شرایط تحقیق طراحی شده باشد، به گونه‌ای که آنچه باقی می‌ماند فقط «واقعتهای عینی» باشد. برای مثال، بررسی آزمایشی را در نظر آورید که دانش آموزان دو کلاس سوم را در بر می‌گیرد. در این بررسی یک کلاس سوم، گروه آزمایشی و کلاس دیگر، گروه گواه تلقی می‌شود. باز فرض کنید که مشاهده گرانی در هر یک از کلاسها قرار داده شده‌اند تا عکس العملهای متقابل بین

1- Geertz, C. (1980). Blurred genres: The refiguration of social thought. *American Scholar*, 49, 165-178.

2- McCutcheon, G. (1981). On the interpretation of classroom observations. *Educational Researcher*, 10, 5-10; and P. Diesing (1972). *Patterns of Discovery in the Social Sciences*. London: Routledge & Kegan Paul.



معلمان و دانش آموزان را ثبت کنند. در بررسی کیفی بهتر است مشاهده‌گر از اینکه شاهد رفتار گروه آزمایشی یا گروه گواه خواهد بود، بی‌خبر باشد؛ به این معنی که از ویژگیهای آزمودنیها (طبقه اجتماعی، هوشبهر، موفقیت تحصیلی قبلی و غیره) اطلاعی نداشته باشد، و همچنین از برگه‌ای کاملاً ساخت دار برای مشاهده استفاده کند، که به استنباط بسیار کم و تفسیر اندک درباره آنچه بین معلم و شاگردان اتفاق می‌افتد، نیاز داشته باشد. در تحقیق کمی، محقق بدین منظور از این شیوه‌ها استفاده می‌کند که مطمئن شود ارزشها و باورهای مشاهده‌گر بر مشاهده‌ای که انجام داده است، اثر نمی‌گذارد و خلوص آن را خدشه دار نمی‌سازد. محقق با تبعیت از این شیوه‌های مشاهده، عاری بودن تحقیق را از ارزشها، کاملاً بیمه می‌کند (یا به عبارت دیگر تضمین می‌کند). بالعکس، رویکرد کیفی استدلال می‌کند که تحقیق همواره در قید ارزش است - و هرگز نمی‌توان آن را فارغ از ارزشها دانست - و ذهن محقق باید در مورد نقشی که ارزشها در هر بررسی ایفا می‌کنند، روشن باشد. پژوهشگران کیفی می‌گویند که تحقیق در این موارد وابسته به ارزشهاست: از نظر انتخاب مسئله مورد بررسی، در انتخاب اینکه آیا یک رویکرد کمی یا رویکرد کیفی نسبت به مسئله تحقیق در پیش گرفته شود، از نظر شیوه‌هایی که برای بررسی مسئله انتخاب می‌شود، در انتخاب شیوه تفسیر نتایج و یا یافته‌ها، و یا از طریق ارزشهای ذاتی در زمینه‌ای که بررسی در آن صورت می‌گیرد.<sup>(۱)</sup> پژوهشگران کیفی بر این باورند که درک تجربه بشری بدون در نظر گرفتن کنش متقابل باورها و ارزشهای محققان و شرکت کنندگان در یک بررسی، امری امکان ناپذیر است. بعلاوه، پژوهشگران کیفی می‌گویند که بررسی انسان، به تعامل فراوان، مداوم و معنی دار بین محققان و پاسخگویان آنها (آزمودنیها) نیاز دارد و تحقیق نه تنها نباید اینگونه تماسها را به حداقل برساند، بلکه باید سعی کند آنها را به حداکثر ممکن برساند.<sup>(۲)</sup> از آنجا که تحقیق کیفی به صورت بارز بر نقش ارزشها در تحقیق تأکید می‌کند و ارتباط و تعامل بین محققان و پاسخگویان را ضروری می‌داند، معمولاً ادعا می‌شود که یافته‌های (نتایج) اینگونه بررسیها به سادگی، تنها نوعی عقیده است. به منظور مقابله با این ادعا، محققان طرفدار شیوه کیفی، فنون مختلفی را برای نشان دادن قابل اعتماد بودن یافته‌هایشان، به کار می‌گیرند. این فنون در بخشهای بعد مفصلتر مورد بحث قرار خواهد گرفت.

1- Lincoln and Guba, *Naturalistic Inquiry*, pp. 160-166.

## / جنبه‌های حائز اهمیت تحقیق کیفی

هرچند پژوهشگران کیفی به صور مختلف عمل می‌کنند، اما در مجموعه خاصی از شیوه‌ها برای انجام یک تحقیق، وجه اشتراک دارند. برخی از جوانب حائز اهمیت شیوه عمل، شامل موارد ذیل می‌شوند.<sup>(۱)</sup>

توجه به زمینه تحقیق کیفی فرض بر آن دارد که رفتار بشر وابسته به اوضاع و احوال است و تجربه بشری معنای خود را از اثرات اجتماعی، تاریخی، سیاسی و فرهنگی می‌گیرد و در نتیجه از آن جدایی ناپذیر است. در نتیجه، تحقیق همواره به اوضاع و احوال و یا موقعیت خاصی وابسته است. طرفداران تحقیق کیفی می‌گویند که رویکرد کمی در مطالعه تجربه بشری، سعی در جداسازی رفتار بشر از شرایط و اوضاع و احوال دارد.<sup>(۲)</sup>

وضعیت طبیعی شرایطی که در آن تجربه انسان مورد بررسی قرار می‌گیرد، باید به صورت طبیعی (مانند یک کلاس درس، تمامی یک مدرسه، یک سازمان)، نه از قبل طرح ریزی شده و مصنوعی (مانند وضعیتهای ساخته شده در یک تجربه آزمایشگاهی)، باشد. بنابراین، تحقیق کیفی در شرایط طبیعی صورت می‌گیرد. بعلاوه، در تحقیق کیفی از قبل در مورد آنچه که قرار است بررسی شود، محدودیتی قائل نمی‌شوند. برای مثال، رابطه بین متغیرهای مستقل و وابسته بخصوصی را شناسایی، تعریف، بررسی و یا آزمون نمی‌کنند، بلکه تجربه انسان را به صورت کلی، با در نظر گرفتن تمامی عوامل و تأثیرات در یک موقعیت خاص، بررسی می‌کنند. ابزار انسانی در بررسیهای کیفی، محقق خود وسیله گردآوری داده‌هاست. او در محیط با مردم صحبت می‌کند، فعالیت‌های آنها را مشاهده می‌کند، مدارک و یادداشتهای کتبی آنها را مطالعه می‌کند، و این اطلاعات را در اوراق و دفاتر خاص بررسی، ثبت می‌کند. تحقیق کیفی بر روشهای پژوهش میدانی، یعنی مصاحبه، مشاهده بدون ساخت، تحلیل مدارک به عنوان ابزار اصلی گردآوری داده‌ها، تکیه دارد و از استفاده از آزمونهای کتبی، لوازم مکانیکی و برگه کاملاً ساخت‌دار برای مشاهده، اجتناب می‌کند. پژوهشگر به جای ارقام و آماره‌ها، با داده‌ها به صورت کلمات سر و کار دارد، ولی بعضی مواقع احتمال دارد داده‌ها را به صورت ارقام هم گردآوری کند. کنترل

۱ - برای اطلاع بیشتر مراجعه شود به:

R. R. Sherman and R. B. Webb (eds) (1988). *Qualitative Research in Education: Focus and Methods*, London: The Falmer Press; and D. M. Fetterman (1988). *Qualitative Approaches to Evaluation in Education*, New York: Praeger.

2- Mischler, E. G. (1978). Meaning in context: Is there any other kind? *Harvard Educational Review*, 49, 2-10.

حجم زیادی از داده‌ها که از طریق مصاحبه‌ها، مشاهده‌ها و گردآوری مدارک به دست می‌آید، از مسائل قابل ملاحظه در بررسیهای کیفی است.<sup>(۱)</sup>

همچنین پژوهشگران کیفی معمولاً گزارش روزانه و یا دفتر یادداشت‌های روزانه خصوصی و یا یادآوریها را نگاه می‌دارند که در آنها افکار، احساسات، فرضیات، انگیزشها و منطق خود را برای تصمیم‌گیریهایشان، ثبت می‌کنند. این یکی از راههایی است که محقق کیفی از طریق آن، مقوله وابسته بودن تحقیق به ارزشها را مطرح می‌سازد.

**طرح نوظهور** در مطالعات کمی، محققان کلیه جوانب یک بررسی را دقیقاً قبل از آنکه عملاً داده‌هایی گردآوری شود، طراحی می‌کنند. متغیرها و اندازه‌های به کار برده شده برای متغیرها، آماره‌های لازم برای تحلیل داده‌ها و غیره را از قبل مشخص می‌کنند. این امر به این دلیل امکان‌پذیر است که محقق از قبل می‌داند به دنبال چیست. آنها فرضیه‌ها و سؤالات خاص را در ذهن دارند و می‌توانند تصور کنند که یک آزمون فرضیه و یا پاسخ به سؤال بررسی، به چه صورت خواهد بود. گذشته از مسائل و پدیده‌های خاص مورد بررسی، پژوهشگران اصرار دارند که مشخص سازی عناصر طرح بررسی بسیار حائز اهمیت است. بالعکس، پژوهشگران کیفی به ندرت تمام جوانب یک طرح را قبل از شروع بررسی مشخص می‌کنند یا هرگز این کار را نمی‌کنند، بلکه طرح با آشکار شدن بررسی، نمایان می‌شود. آنها شیوه‌ها و راههای انجام کار (طرح) را با موضوعی که در دست دارند، منطبق می‌کنند. این امر الزامی است، زیرا پژوهشگر کیفی هرگز از آنچه که در یک موقعیت خاص به دست می‌آورد، مطمئن نیست، چون آنچه که از یک موقعیت خاص به دست می‌آید، به ماهیت و نوع تعاملهای بین محققان و مردم و موقعیت بستگی دارد و این تعاملها کاملاً قابل پیش‌بینی نیستند و همچنین مشخصه‌های حائز اهمیت که به بررسی نیاز دارند، همیشه قبل از آنکه عملاً توسط محقق مشاهده شوند، قابل شناسایی نیستند. در نتیجه، ویژگیهای بررسی کیفی تنها می‌تواند به شکلی بسیار کلی مشخص شود تا نشان دهد که آن بررسی چگونه می‌تواند صورت پذیرد: بررسی کیفی با یک مسئله، سؤال یا موضوع تحقیق آغاز می‌شود. برای مثال، مددکاران اجتماعی چگونه با فشارهای ناشی از شغل خود کنار می‌آیند؟ عکس العمل معلمان مدارس روستایی در قبال تکنولوژیهای آموزش از راه دور

۱- برای مثال مراجعه شود به:

H. G. Levine (1985), Principles of data storage and retrieval for use in qualitative evaluations, *Educational Evaluation and Policy Analysis*, 7, 169-186.

چیست؟ درد مزمن چیست؟ ارتباط مدارس روستایی با جامعه‌شان چگونه است؟ سپس، محقق در صدد دستیابی به یک مکان خاص یا گروهی از افراد برمی‌آید تا بتواند موضوع تحقیق خود را در آنجا و با آن افراد بررسی کند و با اتخاذ نقشی به عنوان مشارکت‌کننده - مشاهده‌گر کامل، یا فقط مشاهده‌گر، یا ترکیبی از آن دو، برای کسب اجازه ورود به آن مکان اقدام می‌کند. اموری چون کسب اجازه ورود به محل، اتخاذ نقش، ایجاد و حفظ اعتماد در شرکت‌کنندگان بررسی، مسائلی تقریباً پیچیده هستند که درباره آنها مطالب بسیار زیادی نوشته شده است.<sup>(۱)</sup> محقق، در پی یک مرحله مقدماتی آشنایی با مکان، آگاهی از افراد شرکت‌کننده در بررسی و غیره، با ساختن فرضیه‌ها و سؤالات آزمایشی (مقدماتی)، کار خود را بر جنبه‌ها یا مقوله‌های برجسته تحقیق متمرکز می‌کند. این فرضیه‌ها و یا سؤاله‌های اولیه، بعداً با تمرکز دقیق‌تر تحقیق، پالایش (محدود) می‌شود. معمولاً، طی این مرحله، محقق به بازنگری تحقیقات قبلی و مطالب مربوط به موضوع می‌پردازد و از این طریق به درک خود از پدیده‌های مورد بررسی، عمق بیشتری می‌دهد. در نهایت فرضیه‌ها از طریق روشهای گوناگون، تأیید و آزمون می‌شوند. مرحله نهایی ترک کردن مکانی است که بررسی در آنجا انجام شده است.

ایجاد حس اعتماد<sup>(۲)</sup> پژوهشگران کیفی از روشهای مختلفی برای کنترل اعتبار داده‌های به دست آمده و تأیید بینشها و فرضیات رو به تکامل خود استفاده می‌کنند. از بین این روشها می‌توان از اقامت طولانی در محل و مشاهده مصرانه به منظور دستیابی به گستره و عمق کافی برای مشاهدات، نام برد. سه وجهی کردن - استفاده از منابع چندگانه داده‌ها، مشاهده‌گران چندگانه و یا روشهای چندگانه - شیوه‌ای دیگر است که به منظور افزایش احتمال معتبر بودن فرضیات و تفاسیر، به کار گرفته می‌شود. کسب اطلاعات در مقاطع مختلف از هم‌تایان محقق و نظر پرسی از همکاران (ارائه تفاسیر محقق به همکاران در یک موقعیت خاص برای اعتباریابی)، شیوه‌های حائز اهمیت دیگر هستند.

به منظور افزایش قابلیت اطمینان (تقریباً معادل واژه اعتبار) یک بررسی، محقق کیفی اغلب یک شرح انجام کار تهیه می‌کند و در آن مطالبی را می‌آورد که نحوه انجام بررسی را مستند

۱ - برای مثال مراجعه شود به:

W. B. Shaffir, R. A. Stebbins, and A. Turowetz (1980), *Fieldwork Experience*, New York: St. Martin's; and R. M. Emerson (ed.) (1988), *Contemporary Field Research*, Prospect Heights, IL: Waveland.

2- Lincoln and Guba, *Naturalistic Inquiry*, pp. 289-331.

می‌سازند. این مطالب شامل کارهای انجام شده، زمان انجام آنها و دلیل انجام آنهاست. این شرح انجام کار شامل داده‌های گردآوری شده از مصاحبه‌ها و مشاهده‌ها، یادداشتهای مربوط به تصمیم‌گیریهای محقق در مورد اینکه با چه کسی مصاحبه کند و چه چیز را به چه دلیل، مشاهده کند، سابقه مستندات در مورد اینکه فرضیه‌های مقدماتی چگونه بر مبنای داده‌های خام پروارنده و در مراحل بعدی پالایش یافتند و آزمون شدند، یافته‌های بررسی و غیره، می‌شود. یک محقق مستقل ثالث، از شرح انجام کار به عنوان راهنما استفاده می‌کند تا مشخص سازد که آیا شیوه‌های مورد استفاده قابل اطمینان هستند و آیا یافته‌ها قابل تأیید هستند یا خیر. به عبارت دیگر، آیا این یافته‌ها به طور منطقی از داده‌های تحقیق حاصل شده‌اند و بر داده‌های تحقیق استوار هستند یا خیر.<sup>(۱)</sup>

**تحلیل استقرایی در غالب بررسیهای کیفی گردآوری و تحلیل داده‌ها به صورت همزمان** صورت می‌گیرد. به عبارت دیگر، محقق صبر نمی‌کند تا تمام داده‌ها به دست آید و سپس شروع به تفسیر آنها کند. با شروع اولین مصاحبه یا مشاهده، محقق کیفی به معنای چیزی که شنیده یا دیده است، فکر و در مورد معنای آنها گمانه‌زنی (فرضیه‌های مقدماتی) می‌کند و در مصاحبه‌ها یا مشاهدات بعدی در پی تأیید یا رد آن گمانه‌زنیهاست. این فرایند تحلیل داده‌ها استقرایی است - یعنی از داده‌ها شروع می‌شود و به فرضیه و سپس به نظریه ختم می‌شود. همچنان که محقق داده‌ها را از طریق فرایندهای رمزگذاری و طبقه‌بندی تقلیل می‌دهد و بازسازی می‌کند، به بسط نظریه مبتنی بر داده‌ها (نظریه درباره پدیده‌های مورد مشاهده که مستقیماً به داده‌های مربوط به آن پدیده‌ها وابسته است) نیز نظر دارد.<sup>(۲)</sup>

**ارائه گزارش.** مسلماً گزارشهای بررسی کیفی، بر حسب ماهیت نشریه‌ای که در آن چاپ می‌شود، متفاوت خواهند بود (برای مثال، مقاله تهیه شده برای مجله *The American Educational Research Journal* با یک رساله یا بررسی که در حد یک کتاب است، متفاوت هستند)<sup>(۳)</sup>، ولی همانگونه که در این قطعه منتخب از بررسی میدانی آن

1- Schwandt, T. A., and E. S. Halpern (1988). *Linking Auditing and Metaevaluation*. Beverly Hills: Sage Publications.

2- Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge University Press

۳- برای مثال مراجعه شود به:

M. L. Smith and L. A. Shepard (1988), Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices, *American Educational Research Journal*, 25, 307-333; and J. J. Blase (1988), The everyday political perspective of teachers: Vulnerability and conservatism, *Qualitative Studies in Education*, 1, 125-142.

پشکین<sup>(۱)</sup> از یک مدرسه بنیادگرای مسیحی، مشاهده می‌شود، معمولاً از نظر شکل کاملاً صورت روایی دارد و شامل تفسیرهای مفصل از وضعیت و زمینه است:

راهروهای آکادمی بتانی باپتیست<sup>(۲)</sup>، موجب افتخار مدیر آن، مک گرو<sup>(۳)</sup>، است. کلاسهای باروچ و وسیع ابتدایی، ظاهری برجسته و جوش مانند کلاسهای ابتدایی همه جای دیگر دارد. بالمکس، کلاسهای دوره متوسطه که دیوارهای آن مزین به تابلوی اعلانات زنده‌ای است که عکسها و سخنان برگزیده با سنجاقها و گیره های مورد استفاده در ادارات به آن نصب شده‌اند، تا حدودی دلگیر هستند. سالن ورزش دارای یک صحن در یک طرف است و اندازه آن می‌تواند دو برابر شده و به سالن اجتماعات تبدیل شود. در طرف مقابل صحنه یک آشپزخانه است برای خرید ناهار، کودکان کنار دیوار سالن ورزش صف می‌کشند، سینی‌ها را بر می‌دارند و به اطاق ناهارخوری وارد می‌شوند. کتابخانه مدرسه و بخش اداری در صحن ابتدایی قرار دارند. دفتر زیبا و تزئین شده مدیر پشت اطاق منشی او قرار دارد که همچنین نقش همه کاره کل مدرسه (پرستار، مشاور گروه تفریحات مدرسه، همراهی کنندهٔ پیانو برای تک خوانهای آواز و مجری قوانین مدرسه) را بازی می‌کند. یک کتابخانه تزئینی محتری چندین کتاب از ابراهام لینکلن<sup>(۴)</sup>، که با تخته‌های نگاهدارندهٔ کتاب، ثابت شده‌اند، کنار دیوار دفتر قرار دارد و در بالای آن دو عکس قاب شده از لینکلن و سخنرانی او در گنیسبورگ نصب شده است. نرده‌ای به ارتفاع کمر انسان محوطهٔ مخصوص ارباب رجوعها را از دفتر جدا می‌کند. بر دیوار پشت قسمت ارباب رجوعها تصویری از تئودور روزولت<sup>(۵)</sup> آویخته شده و برگزیده‌ای از سخنان او در زیر تصویرش چاپ شده است. در اطراف این تصویر نشانها قرار دارند: جایزه ویلیام مولر<sup>(۶)</sup> برای دانش آموزان دبیرستانی؛ جایزه آن مولر<sup>(۷)</sup> برای دختران سال آخر دبیرستان که زندگیشان بیش از دیگران نمونه‌ای از اصول مورد نظر مؤسس دبیرستان است؛ جایزهٔ دستهٔ کر ملی<sup>(۸)</sup> و غیره<sup>(۹)</sup>.

گزارشهای بررسیهای کیفی، به جای ارائهٔ جداول آماره و نمودارهای نشان دهندهٔ داده‌های عددی، نشان‌دهندهٔ بیان عادی شرکت کنندگان در یک بررسی هستند که از اطلاعات مربوط به

1- Alan Peshkin

2- Bethany Baptist Academy

3- McGraw

4- Abraham Lincoln

5- Theodore Roosevelt

6- William Muller award

7- Ann Muller award

8- National School Choral award

9- Peshkin, A. (1988). *God's Choice: The Total World of a Fundamentalist Christian School* (p. 33).

Chicago: University of Chicago Press.

مصاحبه‌ها و مدارک گرفته شده است. برای مثال، در همان بررسی، پشکین گزارشی از کتابدار مدرسه بتانی را می‌آورد که در آن، چگونگی اعمال سانسور را توسط خود، شرح می‌دهد:

برخی از کتابهای علوم، پیش از حد در مورد تکامل مطالب دارند و یا در بعضی جاها زیاده از حد مفرضانه هستند، ولی مطالب خوب هم در آنها هست، در نتیجه من این کتابها را نزد آقای کروگر<sup>(۱)</sup> (دبیر علوم) می‌برم و می‌گذارم آنچه را که مورد استفاده اوست، بردارد. من به دنبال نظریه تکامل هستم. این یکی از آن چیزهاست. من به دنبال کلمات زشت هستم. من به دنبال آنها می‌گردم. من دو صفحه در مورد میمونهایی که به شکل انسان تکامل یافته‌اند، پیدا کردم. البته اینگونه مطالب را به هیچ وجه تأیید نمی‌کنیم، در نتیجه من تنها صفحات را به هم چسباندم و این، مشکلی را در خواندن دو طرف آن ایجاد نکرد. بنابراین، ابتدا بخشی در مورد تکامل بود. من دو طرف آن برانتر باز کردم و با حروف سیاه نوشتم تکامل، تا هر کسی این مطلب را می‌خواند، بداند در مورد تکامل است، اما تمام کتاب را نابود نکردم؛ چون خیلی از مطالب آن خوب بود. در صورتی که تصویر فردی برهنه را بینم، یک لباس شنا و یا لباس روی آن می‌کشم. البته این در مورد کتابهایی است که به هنر ارتباطی ندارد، ولی در مورد هنر، هنر هنر است؛ و اگر فرد بی‌لباسی را در آن دیدید، باید بدانید که به همان منظور کشیده شده است. ما یک کتاب داستان داشتیم که در آن تمام بچه‌ها لخت شنا می‌کردند. چیز مهمی نبود. من فقط یک لباس شنا برایشان کشیدم. ما جدیداً بیست کتاب جدید را در مورد ارزش صداقت و این چیزها بین معلمان تقسیم کردیم. من از هریک از معلمان کلاسهای پایین‌تر خواستم که چهار عدد از این کتابها را بگیرند و بخوانند تا ببینند آیا چیزی دارد که ما درباره‌اش نگرانی داشته باشیم. یکی از کتابها، به نوعی اهمیت نظم و ترتیب را کوچک می‌شمرد و می‌دانید ما چه کار کردیم؟ بر روی تصویر پسر اخمویی که تنبیه شده بود، ولی زیر بار نمی‌رفت، تصویر دیگری چسبانیدیم که پسر بچه‌ای را با چهره خندان نشان می‌داد.<sup>(۱)</sup>

## // مورد پژوهیها

مورد پژوهیها و آزمایشهای تک آزمودنی (رجوع شود به فصل ۹) هر دو به بررسی یک فرد و

یا تنها یک واحد مجزای اجتماعی، مانند خانواده، باشگاه و یا گروه نوجوان سر و کار دارند. در حالی که در آزمایشهای تک آزمودنی تأکید بر روی یک رفتار و یا تعداد معدودی رفتار است، مورد پژوهیها سعی بر آن دارند که کل رفتارهای آزمودنی و رابطه این رفتارها را با سابقه و محیط آزمودنی مورد بررسی قرار دهند. در آزمایش تک آزمودنی محقق تیمار خاصی را به منظور بررسی اثر آن بر آزمودنی ارائه می‌کند، در حالی که در یک مورد پژوهی محقق واکنش آزمودنی را در قبال رویدادها، به صورت طبیعی مشاهده می‌کند.

در یک مورد پژوهی محقق سعی می‌کند که یک فرد یا واحد را عمیقاً بررسی کند. محقق سعی دارد که کلیه متغیرهایی را که در سوابق و یا رشد آزمودنی حائز اهمیت است، کشف کند. تأکید بر روی این امر است که بفهمیم چرا فرد کاری را انجام می‌دهد و چگونه در پاسخ به محیطش تغییر می‌کند. چنین کاری نیاز به بررسی مشروح، برای یک دوره زمانی قابل ملاحظه دارد. محقق در مورد وضعیت کنونی آزمودنی، تجارب قبلی، محیط و چگونگی رابطه این عوامل بایکدیگر، اطلاعاتی را جمع آوری می‌کند.

غالب مورد پژوهیها از کوشش برای حل مشکلات سرچشمه می‌گیرند. مورد پژوهیهای معروف فروید از آنجا آغاز شد که او سعی کرد به آزمودنیهای خود در حل مشکلات شخصی آنان کمک کند. فروید در حین کوشش برای جستجوی عمیق در پویاییهای شخصیتی بیمارانش، به این نتیجه رسید که روابطی که او بین بیماران و محیط آنها مشاهده کرده است، ممکن است ویژگی افراد دیگری با مشکلات مشابه باشد. فروید گزارش جزئیات مصاحبه‌های خود را با بیماران و تفسیر خود را از افکار، رویاها و اعمال آنها با این فرض منتشر ساخت که می‌توان از حاصل مطالعات او به تعمیم‌های همه‌گستری دست یافت.

مهمترین مزیت یک مورد پژوهی، امکان عمیق بودن آن است، که در آن سعی می‌شود کلیت یک کودک و یا فرد در تمامیت محیط او درک شود. نه تنها اعمال کنونی یک فرد، بلکه گذشته، محیط، عواطف و افکارش می‌تواند مورد بررسی قرار گیرد. محقق سعی دارد تعیین کند که چرا فردی به صورت خاصی رفتار می‌کند و صرفاً نمی‌خواهد رفتار او را ثبت کند.

مورد پژوهیها معمولاً این امکان را برای محقق فراهم می‌آورند که به پیشنی از جنبه‌های اصلی رفتار بشری دست یابند. ویژگی جستجوی عمیق در این روش، امکان دارد به کشف روابطی منجر شود که قبلاً مورد نظر نبوده‌اند.



از سوی دیگر، مزایای مورد پژوهی نقاط ضعف آن نیز هست. این نوع تحقیق، اگرچه از عمق برخوردار است، ولی مسلماً دچار عدم وسعت است. پویایی یک فرد و یا یک واحد اجتماعی ممکن است رابطه کمی با پویایی دیگری داشته باشد. در عمل، غالب مورد پژوهیها از کوششهای مشاوره‌ای و یا درمانی ناشی می‌شوند و در نتیجه اطلاعاتی که به دست می‌دهند، به افراد استثنایی و نه معرف جامعه، مربوط می‌شود.

امکان بینش در یک مورد پژوهی نیز به وجود آورنده زمینه ذهنی گرابی و یا حتی تعصب می‌شود. مفهوم قبلی در ذهن محقق می‌تواند تعیین کننده این امر باشد که چه رفتارهایی مورد مشاهده قرار گیرد و چه رفتارهایی نادیده گرفته شود و همچنین با چه روشی، مشاهدات تفسیر شوند.

وجهه رویکرد مورد پژوهی خدشه‌دار شده است، زیرا پاره‌ای از محققان دوران گذشته مشاهداتشان را با سازه‌هایی توضیح داده‌اند که تأیید و یا تکذیب آن از طریق مطالعه تجربی غیر ممکن است.

از آنجا که مقدار تعمیم‌های معتبر حاصل از مورد پژوهیها بشدت محدود است، مورد استفاده اصلی آنها به عنوان ابزاری برای آزمون فرضیه نیست، بلکه تبیین فرضیه‌هایی است که بعداً به وسیله تحقیقات دقیق‌تر آزمون خواهند شد. برای مثال، بینشهایی که ژان پیاژه از مورد پژوهیهای معروفش بر روی بلوغ ذهنی به دست آورد، فرضیه‌های مفیدی را تبیین کرد که هنوز هم به وسیله روشهای دیگر مورد بررسی قرار می‌گیرند.

در مواردی که مورد پژوهیها ناشی از کوششهایی برای کسب اطلاعات در مورد افراد، به منظور کمک به آنها، باشد، جنبه تحقیقاتی بررسی در مقام دوم قرار می‌گیرد. به هر حال مورد پژوهیها معمولاً با هدف اولیه کسب آگاهی انجام می‌گیرد. مورد پژوهی کلاسیک ایثار<sup>(۱)</sup> در مورد پسر وحشی آویرون، کوششی برای کسب اطلاعات در مورد اثرات تمدن، از طریق مطالعه احوال پسرپچه‌ای بود که در قرن هیجدهم در فرانسه دور از تمدن، بزرگ شده بود. مورد پژوهیهای پیاژه بیشتر به منظور حصول معرفت در زمینه رشد ذهنی کودکان انجام گرفت، تا سودرسانی به آزمودنیهایی که در آن شرکت داشتند.

1- Iard, J. G. *The Wild Boy of Aveyron*. Trans. G. and M. Humphrey (1982) New York: Appleton.

## /// تحقیق تاریخی

تحقیق تاریخی کوششی است به منظور تبیین حقایق و دستیابی به نتایجی دربارهٔ رویدادهای گذشته. محقق تاریخ به طور نظام‌مند و عینی شواهدی را می‌یابد و ارزیابی و تفسیر می‌کند که ما از نتیجهٔ کار او روزگار پیشین را درک می‌کنیم. بر مبنای شواهد گردآوری شده، نتایجی به دست می‌آید که به آگاهی ما دربارهٔ این که رویدادهای گذشته چرا و چگونه اتفاق افتاده است، می‌افزاید و فرآیند تبدیل گذشته به حال را به ما می‌نمایاند. درک بیشتر عصر حاضر و مبنای منطقی‌تر برای انتخاب و تصمیم‌گیری، آن نتیجه‌ای است که امید می‌رود از تحقیقات تاریخی حاصل شود.

محقق تاریخ، نسبت به محققان دیگر رشته‌ها، از کمبودهای بیشتری رنج می‌برد. کنترل بر روی تیمار، اندازه‌گیری و نمونه‌گیری محدود است و هیچ امکانی برای تکرار وجود ندارد. همانند تحقیقات علمی - مقایسه‌ای و توصیفی، متغیرهای مستقل (تیمار)، تحت کنترل محقق نیستند، در نتیجه کلیهٔ احتیاطهای لازم برای تفسیر چنین بررسی‌هایی می‌تواند در مورد تحقیقات تاریخی نیز صادق باشد. در تحقیقات علمی - مقایسه‌ای و توصیفی، محقق معمولاً اندازه‌گیری را کنترل می‌کند؛ بدین معنا که او می‌تواند اندازه‌هایی را که به عنوان متغیر وابسته مورد استفاده قرار می‌گیرند، انتخاب کند. در حالیکه مورخ دربارهٔ آنچه از اسناد، آثار باستانی، پیشینه‌ها و دست‌ساختها برای مطالعات او باقی مانده است، هیچ‌گاه حق انتخاب ندارد، اما کنترل محدودی دارد بر اینکه چه سؤالهایی را از این منابع خواهد پرسید و چه اندازه‌هایی را در مورد آنها اعمال خواهد کرد. حتی در اینگونه موارد هم، او تنها می‌تواند آن چیزهایی را که شاهدان به یاد می‌آورند یا اسناد حاوی آنها هستند، اندازه‌گیری کند.

در تحقیقات توصیفی و آزمایشی، محقق می‌تواند وسیعی می‌کند نمونه‌گیری را کنترل کند، بدین معنا که او می‌تواند تصمیم بگیرد که دربارهٔ چه کسانی مطالعه کند. مورخان تنها می‌توانند کسانی را مورد مطالعه قرار دهند که اسناد و دست‌ساختهایی از آنها به جای مانده باشد. چنانچه روزنامه‌ها یک بخش از جامعه را نادیده بگیرند و منابع دیگری در مورد آن جامعه وجود نداشته باشد، مورخ قادر نخواهد بود که سهم این بخش را بر روی زندگی جامعه ارزشیابی کند. محدودیت دیگری که محققان تاریخی با آن دست به‌گریبان هستند، این است که آنها نمی‌توانند به دلیل اینکه پیشینه‌ای از گذشته باقی نمانده، فرض کنند که اتفاقی نیفتاده است. از طرف دیگر، نمی‌توانند فرض کنند که سکوتی تعمودی در مورد واقعه‌ای تاریخی وجود داشته است.

## // منابع اولیه و ثانویه

مورخ، ماده کار خود را با عناوین منابع اولیه و ثانویه طبقه بندی می کند. منابع اولیه عبارتند از: اسناد اصلی، آثار باستانی و بقایا و دست ساختها. اینها نتایج مستقیم رویدادها و یا وقایع ثبت شده توسط شاهدان عینی هستند. برای مثال می توان از گزارش جلسه هیئت امنای مدرسه، یک نوار ویدیویی ویرایش نشده مسابقه بسکتبال و مجموعه کارهای هنری جمع آوری شده از کلاس سوم دبستان، نام برد. در منابع اولیه تنها ذهن مشاهده گر بین رویداد اصلی و محقق مداخله می کند. به خاطر داشته باشید که در هر یک از این مثالها، ذهن مشاهده گر، بین رویداد و اثر قرار دارد. کسی از قبل تصمیم گرفته است که چه چیزی از جلسه هیئت امنای مدرسه ثبت شود و چه چیزی نشود، دوربین چه موقع فیلم بگیرد و چه موقع نگیرد و در یک مسابقه بسکتبال بر روی چه چیزی تأکید شود و چه آثار هنری حفظ شود.

در منابع ثانویه، ذهن یک غیر مشاهده گر هم بین رویداد و استفاده کننده از اثر، دخالت می کند. چنانچه یک گزارشگر روزنامه در جلسه هیئت امنای دبستان حضور داشته باشد، گزارشش از نوع منابع اولیه می شود، اما چنانچه در تهیه گزارش متکی به یادداشتهای جلسه و یا مصاحبه با شرکت کنندگان در جلسه باشد، گزارش او از نوع منابع ثانویه خواهد بود. بین نمونه های متداول منابع ثانویه می توان از کتابهای تاریخی، مقالات دایرةالمعارف و پیشینه تحقیقات نام برد. مورخان سعی دارند تا آنجا که ممکن باشد، منابع اولیه را به کار گیرند.

## // نقد بیرونی و درونی

دو ایده که سودمندی آنها در ارزشیابی منابع تاریخی ثابت شده است، مفاهیم نقد بیرونی (یا پایینی) و نقد درونی (یا بالایی) می باشند.

اصولاً نقد بیرونی زمانی مطرح می شود که شواهد مورد نظر معتبر باشند. با در نظر گرفتن ماهیت بررسی، ممکن است به تکنیکهای اعتباریابی امضاها، تجزیه شیمیایی رنگها و یا تعیین قدمت آثار هنری به وسیله کربن نیاز باشد. فرض کنید که مورخی نامه ای در دست دارد که مدارس ماساچوست را توصیف می کند و تصور می شود به وسیله هوراس مان<sup>(۱)</sup> نوشته شده باشد. با استفاده از نقد بیرونی، محقق این سؤال را مطرح می سازد که «آیا کاغذ با زمانش تطابق

دارد؟ آیا نوشته دست خط مان است؟ آیا دیدگاه و شیوه نگارش آن با دیگر نوشته‌های مان سازگاری دارد؟»

پس از آنکه اعتبار بخشی از شواهد ثابت شد، محقق تاریخ به مرحله نقد درونی می‌رسد. لازمه این امر، ارزشیابی شواهد است. برای مثال آیا یک سند، گزارش واقعی از یک رویداد را به دست می‌دهد یا نه؟ چنین سؤالی به بهترین وجه با اثبات بخشی از شواهد، توسط شواهد دیگری که بتوانند پرتوی بر آن رویداد بینکنند، و با اطلاعات اضافی در مورد یک رویداد و یا مردم و یا اوضاع آن فراهم آورند، پاسخ داده می‌شود. در مثال فوق، محقق سؤال می‌کند: «آیا توصیف مان از مدارس، غیر مغرضانه است؟ آیا با دیگر توصیفات موجود از مدارس آن دوران تطبیق می‌کند؟»

با در نظر گرفتن محدودیتهای متعدد تحقیق تاریخی، فرد بیش از هر چیز دیگر گرایش می‌یابد که سؤال کند: «اصلاً چرا این نوع تحقیق را انجام دهیم؟» پاسخ اساسی این است که سؤلهای بسیاری مطرح هستند که با روش دیگری پاسخ داده نمی‌شوند. یک محقق با چه روش دیگری می‌تواند اثرات تیراندازی در محوطه دانشگاه کنت یا سایر اغتشاشات دانشجویی را در بهار سال ۱۹۷۰ بر نظام تحصیلات عالی آمریکا بسنجد؟

یک مزیت تحقیق تاریخی در عدم دخالت آن است. این مزیت شاید در پاره‌ای موارد دلیل استفاده از این شیوه باشد. محقق در موقعیت مورد بررسی درگیری عینی ندارد. خطر وجود تعامل آزمایشگر - آزمودنی وجود ندارد. محقق نیاز ندارد که از مقامات مدرسه برای بررسی خود اجازه بگیرد. او به سادگی اسناد لازم را می‌یابد، داده‌های مناسب را گردآوری می‌کند و نتیجه‌گیری لازم را در مکانی دور از محل مورد بررسی انجام می‌دهد.

علاوه بر این، تحقیق تاریخی، در یک موقعیت بحرانی، می‌تواند چشم‌اندازهای جدیدی در اختیار انسان قرار دهد. ماهیت غیردرگیرانه تحقیق تاریخی می‌تواند آن را در یک موقعیت ملامت از احساس و عاطفه، یعنی در جاییکه امکان انجام انواع دیگر تحقیق وجود ندارد، قابل‌پذیرش سازد.

به علت محدودیتهای تحقیق تاریخی، باید در تعمیم دادن نتایج تحقیق تاریخی شدت محتاط باشیم. به دانشجویانی که قصد انجام بررسیهای تاریخی را دارند، توصیه می‌شود به کتابشناسیهای مناسب و منابع مربوط به روش شناسی تاریخی مراجعه کنند.<sup>(۱)</sup>

## /// خلاصه

تحقیقات کیفی، مشتمل بر مورد پژوهیها، نوعی متمایز از تحقیقات در زمینهٔ تعلیم و تربیت و علوم اجتماعی هستند که قادرند گزارشی روشن و مشروح از تجربهٔ انسانی را ارائه دهند. اینگونه بررسیها برای مطالعهٔ واقعیت‌های اجتماعی بر مبنای رویکردی اساساً متفاوت بنا شده‌اند و با رویکرد کمی متداول برای مطالعهٔ مسائل تعلیم و تربیت تمایز دارند. لازمهٔ بررسیهای کیفی مهارتهایی است که نمی‌توان آنها را با خواندن کتب درسی در زمینهٔ روشهای تحقیق به آسانی به دست آورد. برای آنکه محقق در این شیوهٔ بررسی مهارت یابد، باید بتواند در تجربهٔ پژوهش میدانی تماسهای لازم را با افراد مربوط برای دستیابی به محل بررسی برقرار کند، نقش محقق را ایفا کند، در شرکت‌کنندگان بررسی اعتماد را برانگیزد و اعتماد آنها را نگاهدارد، مصاحبه‌ها و مشاهده‌ها را انجام دهد و نتایج آنها را ثبت کند، داده‌ها را به دست آورد و تحلیل داده‌ها را انجام دهد.

در تحقیق تاریخی از مدارک غیر معاصر و آثار باقیمانده برای خلق و آزمون فرضیه‌ها، استفاده می‌شود. منابع اولیه را تا حد امکان به کار می‌برند. محقق تاریخی سعی دارد به وسیلهٔ نقد خارجی سندیت منابع را به اثبات برساند و صحت آنها را از طریق نقد درونی ثابت کند. به علت ضعف ذاتی این شیوهٔ بررسی، باید در تعمیم نتایج به دست آمده از تحقیق تاریخی بسیار احتیاط کرد.

## // مفاهیم اصلی

بررسی طبیعی	ابزار انسانی
بررسی کیفی	بررسی کمی
تحلیل استفرایی	تحقیق تاریخی
سه وجهی کردن	توجه به اوضاع و احوال
طرح نوظهور	شرح انجام کار
کسب اطلاع	قوم نگاری
منبع اولیه	مشاهدهٔ مشارکت‌کننده
نقد بیرونی	منبع ثانویه
وضعیت طبیعی	نقد درونی

← Berlinger (1978). *Historical Analysis: Contemporary Approaches to Clío's Craft*. Malabar, FL: R. E. Krieger; W. N. Bnckman (1982). *Educational Historiography: Tradition, Theory, and Technique*. Cherry Hill, NJ: Emeritus; R. R. Sherman (ed.) (1984). *Understanding History of Education* (2d ed.). Cambridge, MA: Shenkman.

### /// تمرینها

۱- مشخص کنید که کدام یک از جملات زیر خاص رویکرد کمی و کدام خاص رویکرد کیفی است:

الف - هدفها و روشهای علوم اجتماعی را همان هدفها و شیوه های علوم فیزیکی فرض می کند.

ب - ادعا می کند که رفتار انسان همواره وابسته به اوضاع و احوالی است که رفتار در آن روی می دهد.

ج - از توضیحات فرضی - قیاسی استفاده می کند.

د - به منظور اثبات قابل اعتماد بودن از شیوه سه وجهی کردن استفاده می کند.

۲- تفاوت بین مورد پژوهی و آزمایشهای تک آزمودنی چیست؟

۳- یک روان شناس در مورد رخداد های یک گروهی از افرادی که انتظار داشتند در یک شب معین سیاره زمین ناپدید شود، گزارش می دهد. گزارش او شامل مشاهدات علمی او از گروه بوده است. این بررسی را می توان در کدام طبقه بندی قرار داد:

الف - تحقیق کیفی

ب - تحقیق آزمایشی

ج - تحقیق تاریخی

د - تحقیق علی - مقایسه ای

۴- روشهای به کار برده شده توسط پژوهشگران کیفی را جهت اثبات قابل اعتماد بودن نتایج مشخص کنید.

۵- محقق نامهای دارد که تعلیم و تربیت را در اوگاندا در سال ۱۹۷۷ توصیف می کند. گفته می شود که این نامه توسط پرزیدنت ایدی امین نوشته شده است. چه سوالی در این مورد مطرح می شود؟

الف - نقد بیرونی

ب - نقد درونی

۶- در چه صورتی مدرک تاریخی، منبع ثانویه تلقی می شود؟

۷- در مقایسه با دیگر انواع تحقیق، نقاط قوت و ضعف تحقیق تاریخی کدامند؟

/// پاسخها

۱- الف - کمی

ب - کیفی

ج - کمی

د - کیفی

۲- آزمایشهای تک آزمودنی بر یک رفتار و یا تعداد بسیار محدودی از رفتارها تأکید دارند. محقق تیمار خاصی را با این هدف به کار می‌برد که اثر این تیمار را بر آزمودنی بررسی کند. در یک مورد پژوهی سعی بر آن است که کل رفتار آزمودنی، آنچنان که در یک وضعیت طبیعی بروز می‌کند، توصیف شود. محقق رفتار آزمودنی را در رابطه با اثرات محیط طبیعی، اجتماعی و روان شناختی، مشاهده می‌کند.

۳- الف

۴- بررسی طولانی در محل؛ مشاهدهٔ مصرانه؛ سه وجهی کردن، استفاده از مشاهدات مکرر، منابع داده‌های متعدد و روشهای چندگانه، ارائه گزارشهای ادواری و تهیهٔ شرح انجام کار.

۵- الف - آیا نامه توسط ایدی امین نوشته شده است؟

ب - آیا این نامه تعلیم و تربیت را در اوگاندا سال ۱۹۷۷، دقیقاً توصیف می‌کند؟

۶- در صورتی که ذهنیت یک فرد غیر مشاهده‌گر بین مدرک تاریخی و رویداد مربوط حائل

شود، آن مدرک یک منبع ثانویه محسوب می‌شود.

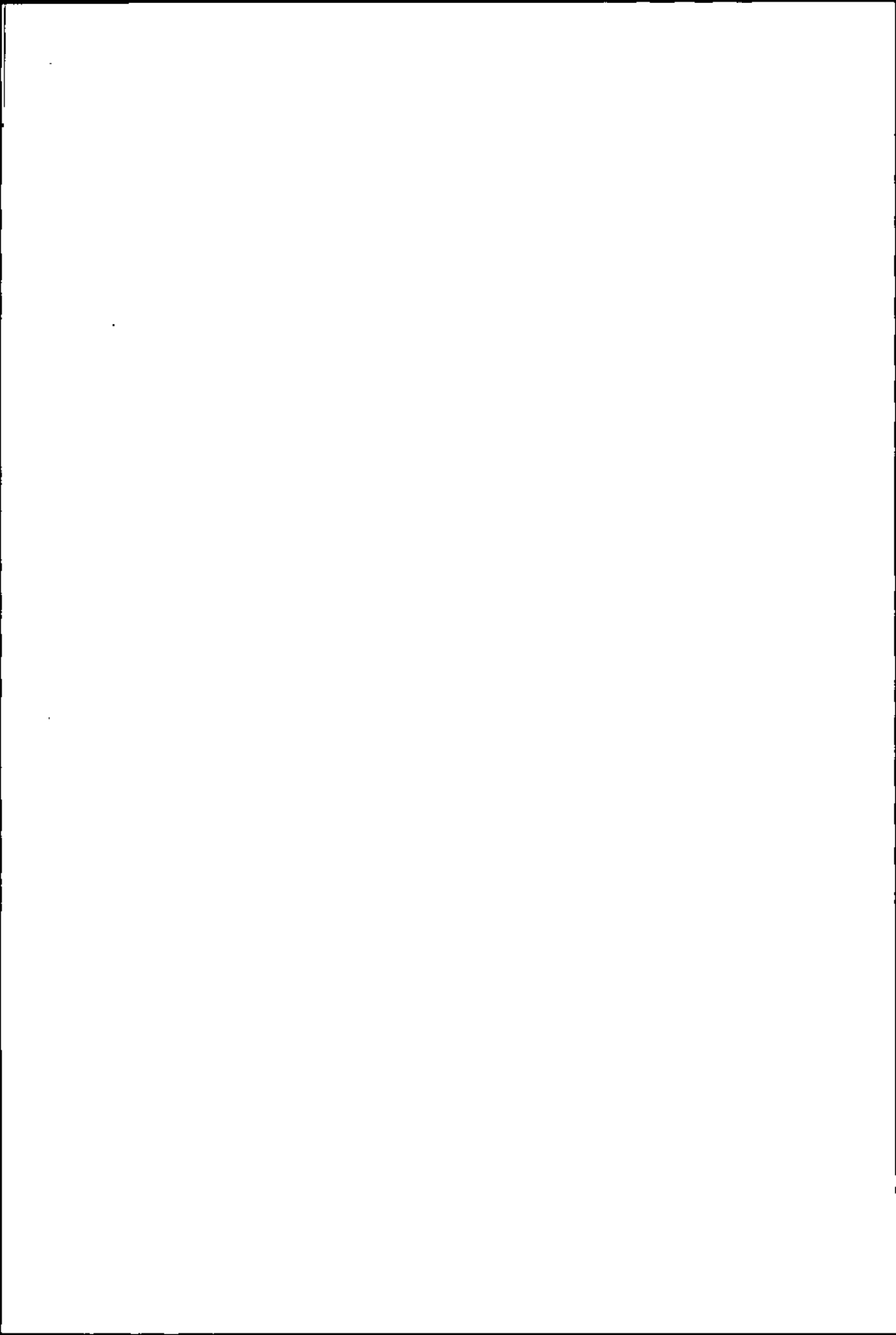
۷- یک مزیت تحقیق تاریخی عدم احتمال اثرات تعامل محقق و یا آزمایش است که تفسیر

یافته‌ها را مخدوش می‌سازد. دیدگاه تاریخی می‌تواند به مسائل و موقعیتهایی از گذشته بپردازد

که بررسی آنها نمی‌تواند به صورت آزمایشی انجام شود. نقطه ضعف اصلی آن، فقدان کنترل

آزمایشی است که تفسیر اطلاعات را مبهم و تعمیم آن را مشکل می‌سازد. همواره احتمال

گردآوری اطلاعات ناکافی و غیر دقیق که اثبات صحت و سقم آن امکان‌پذیر نیست، وجود دارد.





## بخش ششم

# تدوین گزارش تحقیق

## فصل چهاردهم

### رهنمودهایی برای نگارش طرح پیشنهادی تحقیق

#### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- اجزای یک طرح پیشنهادی تحقیق را توصیف و اهمیت آنها را ذکر کند.
- ۲- نقیصه های مشترک بین طرحهای پیشنهادی تحقیق را تشخیص دهد.
- ۳- طرحهای پیشنهادی تحقیق را نقد کند.
- ۴- آماره های مناسب برای پاسخگویی به سوالات تحقیق و داده های مربوط را انتخاب کند.
- ۵- ملاحظات اخلاقی و قانونی را در تحقیق توصیف کند.
- ۶- نقش یک شورای بازنگری سازمانی را توصیف کند.
- ۷- طرح پیشنهادی تحقیق بنویسد، به گونه ای که با رهنمودها همخوانی داشته باشد.

در غالب موارد، محقق باید طرح خود را که به صورت مکتوب سازمان یافته است، در دو مرحله ارائه کند: ۱- مرحله ابتدایی که لازمه آن تهیه یک طرح پیشنهادی تحقیق است و ۲- مرحله نهایی، که یک گزارش نهایی از نتایج تحقیق ارائه می‌شود.

### /// نگارش یک طرح تحقیق

نگارش طرح تحقیق، تعیین کننده ترین و پر هیجان ترین گام در فرآیند تحقیق است. در این مرحله، کل طرح به صورت عینی متبلور می‌شود. محقق در طرح پیشنهادی تحقیق روشن می‌سازد که در جستجوی چیست و چگونه آن را در خواهد یافت و چرا این جستجو ارزشمند است. الهامها و بینشهای محقق برای کشف آگاهیهای جدید در چهارچوب یک برنامه گام به گام گنجانده می‌شود.

این برنامه ممکن است یک طرح کلی غیر رسمی باشد که دانشجویی به عنوان تکلیف درسی به کلاس مقدماتی تحقیق ارائه می‌نماید، یا پایان نامه یا رساله‌ای رسمی باشد که به یک کمیته ارائه می‌شود، یا درخواست تأمین هزینه مالی باشد که به یک بنیاد یا سازمان دولتی تسلیم می‌شود. در زیر، رئوس مطالب جهت نگارش یک طرح پیشنهادی تحقیق، که شامل مراحل لازم جهت تدوین و پیشنهاد یک بررسی تحقیقاتی است، آورده شده است:

#### ۱- مقدمه

الف - بیان مسئله

ب - مرور تحقیقات پیشین

ج - سؤالات و یا فرضیه ها

۲- روش شناسی

الف - نمونه گیری

ب - ابزار گردآوری داده ها

ج - شیوه های کار

۳- تحلیل اطلاعات

الف - سازماندهی داده ها

ب - روشهای آماری

۴- اهمیت بررسی

الف - معانی ضمنی

ب - کاربردها

۵- بودجه مالی و زمانی

الف - بودجه مالی

ب - بودجه زمانی

اگرچه دنبال کردن این فهرست به صورت خیلی جدی لازم نیست، ولی خلاصه مفیدی را برای نگارش طرح پیشنهادی تحقیق فراهم می‌کند. در نظر گرفتن دقیق هر مرحله برای نگارش یک طرح پیشنهادی تحقیق لازم است.

// مقدمه

مقدمه طرح پیشنهادی یک بررسی در واقع بخش حیاتی آن است. نگارنده باید در ابتدا مسئله تحقیق را به صورت روشن و بدون ابهام بیان کند و سپس مسئله را به اطلاعات در دسترس در این زمینه ارتباط دهد و اهمیت و نیاز به اجرای تحقیق را ذکر کند. هرچند ممکن است طرح بررسی به صورت بسیار منظم تنظیم و روشهای آماری بسیار خوب انتخاب شده باشند، اگر مقدمه طرح با دقت و هوشیارانه نوشته نشده باشد، احتمالاً بقیه بخشهای طرح پیشنهادی نیز مورد توجه جدی قرار نخواهد گرفت. موارد زیادی وجود داشته است که طرحهای پیشنهادی صرفاً بر مبنای مقدمه ضعیف، بدون توجه به روش شناسی و طرح آماری پیشنهادی، رد شده‌اند. توصیه می‌شود این بخش، با دقت، احتیاط و با هدف افزایش علاقه خواننده به مسئله تهیه شود.

مقدمه یک طرح پیشنهادی تحقیق باید شامل (الف) بیان مسئله، (ب) مرور بر تحقیقات پیشین و (ج) سوالات و فرضیه‌ها باشد.

/ بیان مسئله

در ابتدای مقدمه، و ترجیحاً در آغاز پاراگراف اول، باید تعریفی روشن و دقیق از مسئله تحقیق ارائه و به دنبال آن توصیفی از پس زمینه مسئله آورده شود. این بخش از طرح پیشنهادی،

همچنین باید شرح مختصری از اهمیت بالقوه بررسی را شامل شود. باید در نظر داشت که اجتناب از تأکید بیش از حد بر اهمیت موضوع قبل از تبیین آن، ضروری است. دو اشتباه متداول که باید از آنها پرهیز کرد، عبارتند از: ۱- آغاز مقدمه با طراحی ماهرانه زمینه مسئله، قبل از آنکه خود مسئله به صورت روشن بیان شده باشد و ۲- ارائه یک توجیه برای انجام بررسی در حالی که تبیین مسئله در لابلای بحث مدفون شده باشد، و یا تنها به صورت مبهم در انتهای آن آمده باشد. اشتباه متداول دیگر آن است که فرض کنیم خواننده در مورد محتوای مسئله به اندازه نگارنده آگاهی دارد. مسئله باید به صورت مشخص و دقیق و با واژه‌هایی بیان شود که برای فردی فرهیخته که تقریباً در مورد زمینه مسئله بی اطلاع است، قابل درک باشد.

معمولاً زمینه مسئله مورد بررسی، در بخش مربوط به تحقیقات انجام شده پرورانده می‌شود، اگرچه در پاره‌ای مواقع در بیان مسئله، ذکر مطالعاتی که به صورت مستقیم به مسئله مورد بررسی منتهی شده‌اند نیز مناسب است. چنانچه مسئله از تجربه شخص نگارنده سرچشمه گرفته باشد، آن را می‌توان در این بخش به صورت خلاصه آورد.

در جایی مناسب از این بخش، هر واژه‌ای که ممکن است برای خواننده ناآشنا باشد و نیز واژه‌هایی که نگارنده در بررسی‌اش به معنای خاص به کار می‌برد، به همان نحو که در بررسی مورد استفاده قرار گرفته‌اند، باید تعریف شوند. محدودیت‌های ویژه حیطه بررسی و نشانه‌هایی حاکی از فرضیه آن باید در پایان این بخش بیابند.

### / مرور تحقیقات پیشین

محقق در بخش تحقیقات مرتبط با مسئله، آنچه را که تا زمان حال در مورد مسئله مورد نظر، دانسته شده است، ارائه می‌کند. این امر زمینه را برای طراحی مسئله‌ها و یا فرضیه‌های پیشنهادی آماده می‌سازد. تحقیقات مرتبط را در چهارچوب اهداف بررسی پیشنهاد شده، محدود کنید. لازم نیست که توصیف تحقیقات شامل کلیه جزئیات باشد، بلکه باید دربرگیرنده شایسته‌ترین بررسی‌های مرتبط و نمایانگر آگاهی از تجربیات امیدبخش زمان خود باشد.

نویسنده در این بخش، نه تنها راه پیشنهادی خود را برای آنکه از دانسته‌ها به نادانسته‌ها برسد، مطرح می‌کند، بلکه تسلط خود را بر حیطه بررسی و آگاهی خود را از پیشرفت‌های اخیر به نمایش می‌گذارد. این بدان معنا نیست که این بخش باید به صحنه فضل فروشی بدل شود. تنها تحقیقاتی که به وضوح با اهداف بررسی مرتبط هستند، آورده می‌شوند.

مرور تحقیقات پیشین باید بر مبنای عناوین، سازمان یابند. این سازماندهی به خواننده نشان می‌دهد که در مورد جوانب مختلف بررسی، چه می‌دانیم. بنابراین، تصویر کاملی از زمینه بررسی، مرحله به مرحله کنار هم گذارده می‌شود.

اشتباهی که باید از آن در بخش مرور تحقیقات پیشین پرهیز کرد، ارائه مجموعه خلاصه‌های تحقیقات، هریک در یک بند است. با انجام این کار، نگارنده، مخاطبانش را خسته می‌کند و فرصت پی‌ریزی مستحکم بررسی خود را از دست می‌دهد. بسیار مفیدتر خواهد بود اگر این بخش بر مبنای عنوان سازمان یابد و روشن شود که چگونه بررسیهای ارائه شده به مسئله یا مسائل تحقیق مربوطند.

نیازی نیست درباره کلیه بررسیهای مرتبط مفصلاً بحث شود. در مرور چند بررسی مشابه، نگارنده ممکن است مهمترین آنها را توصیف کند و سپس به سادگی اظهار دارد که نتایج آن توسط بررسیهای مشابه، که فهرست شده ولی به صورت مفصل توصیف نشده‌اند، تأیید شده است. محققان تازه کار پرشور معمولاً تصور می‌کنند که بررسی پیشنهاد شده آنها منحصر به فرد است و هیچ تحقیقی مربوط به آن، در دسترس نیست. به ندرت چنین پیش می‌آید. جستجوی کامل تقریباً همیشه منجر به یافتن چند گزارش تحقیقاتی می‌شود که حداقل به پاره‌ای از جوانب بررسی مورد نظر مربوط هستند. حتی اگر احتمالاً هیچ تحقیقی در زمینه مورد بررسی وجود نداشته باشد، معمولاً مطالبی با ماهیت نظریه‌ای و یا اندیشمندانه وجود دارند که باید به عنوان پیشینه بررسی آورده شوند.

مسلم است که محقق باید نظریه‌ها و نتایج تحقیقات مخالف با فرضیه‌اش را هم مانند آنهایی که موافق با فرضیه‌اش هستند، بیاورد.

در بخش مرور بر تحقیقات پیشین باید از طریق بحث در مورد یافته‌ها و کاربرد آنها، نتیجه‌گیری شود. در این بخش، نگارنده بینشی را که از مرور تحقیقات پیشین بدست آورده است، به دیگران عرضه می‌کند و به خلاهای موجود در مجموعه دانسته‌های روز، در زمینه مورد بررسی خود اشاره و بنابر این خواننده را مستقیماً به مسئله پیشنهادی تحقیق هدایت می‌کند.

### / سوالات و / یا فرضیه‌ها

حال باید مسئله، که تا اینجا به طور کلی تبیین شده است، مشخص شود. اگر طرح تحقیقاتی به صورت یک تحقیق پیمایشی باشد، مسئله باید به صورت سؤال مطرح شود، مثلاً «چه

درصدی از معلمان در ایالت آیووا در استخدام رسمی هستند؟»

چنانچه طرح تحقیق برای آزمون یک نظریه طراحی شده باشد، به هر حال مسئله به صورت فرضیه تبیین می‌شود. اگرچه پاسخ یک سؤال تحقیق پیمایشی ممکن است مقادیر عددی متفاوتی باشند، اما پاسخ یک آزمون فرضیه همیشه بلی یا خیر است.

تبیین فرضیه تحقیق معمولاً از طریق مفاهیم ضمنی تحقیقات قبلی مرتبط با آن و منطق قیاسی انجام می‌شود. برخی صاحب‌نظران بر این عقیده‌اند که فرضیه باید به صورت فرضیه صفر تبیین شود، زیرا در آزمون آماری با فرضیه صفر سر و کار داریم. به هر حال، پیشنهاد می‌کنیم که در بخش طرح پیشنهادی تحقیق، فرضیه به شکل روابط مورد انتظار بین متغیرها، تبیین شود. بدین نحو، مؤلف در مقایسه با ارائه فرضیه صفر در این مرحله، دلایل روشن تری از هدف بررسی را به خواننده ارائه می‌دهد. استفاده از فرضیه تحقیق در این مرحله همچنین به محقق این اجازه را می‌دهد که منطق قیاسی را که در زمینه مطالعه او وجود دارد، بنا کند. فرضیه صفر را می‌توان در بخش تحلیل داده‌ها آورد.

بسیار مناسب است اگر بتوان فرضیه‌ای را مختصراً به صورت کاربردی تبیین کرد. اگر این امر امکان پذیر نباشد، فرضیه‌ای که با واژه‌های کلی بیان شده است، باید با تعاریف و تصریحات ضروری برای تعریف فرضیه به شکل کاربردی، همراه باشد.

### // روش شناسی

در بخش روش شناسی، نگارنده نشان می‌دهد که بررسی خود را چگونه انجام خواهد داد تا به سؤالهای تحقیق و یا روابط مفروض، پاسخ داده شود. اگر این روابط در واقعیت وجود داشته باشند، مشاهده نیز خواهند شد. در فصول گذشته، طرحهای تحقیق برای انواع مختلف تحقیق آورده شد. محقق باید از بین این طرحهای تحقیقی، مثلاً آزمایشی و غیره، آن را که برای سؤال مورد نظر یا فرضیه مورد بررسی مناسب‌تر از همه تشخیص می‌دهد، انتخاب کند. برای مثال، در صورتی که محقق بخواهد دو روش آموزش درس شیمی را مقایسه کند، در واقع سؤالی آزمایشی را مطرح کرده است. این مسئله تحقیق حداقل به دو گروه آزمودنی نیاز دارد: آزمایشی و گواه. در صورتی که محقق مایل باشد اثر تعامل بین روشهای تدریس شیمی و متغیرهای دیگر، برای مثال، هوش را بررسی کند، هرچند سؤال بازهم آزمایشی است، لکن نیاز به طرحی پیچیده‌تر از طرح دو گروهی دارد. برای چنین بررسی‌ای نیاز است که محقق طرح عاملی را با

حداقل چهار گروه به کار برد.

در بخش روش شناسی طرح پیشنهادی، نگارنده کلیه مراحل بررسی مسئله مورد نظر را ارائه می‌کند. روشهای نمونه‌گیری پیشنهادی، روشهای گردآوری داده‌ها و وسائلی مورد استفاده، توصیف می‌شوند.

طریقه مناسب برای ارائه روش شناسی تحقیق این است که تمامی اطلاعات مربوط به طرح را تحت عناوین (۱) آزمودنیها (۲) ابزار و (۳) شیوه‌های انجام کار، طبقه‌بندی کنیم.

### / آزمودنیها

اولین قدم برای شناسایی آزمودنیهای یک بررسی، توصیف جامعه آماری مورد نظر است: آیا جامعه آماری مورد بررسی، دانشجویان سال اول دانشگاه هستند، یا کودکان شش ساله‌ای هستند که قادر به خواندن نیستند، یا مدیران دبستانهای ابتدایی هستند و غیره. سپس مؤلف / محقق، روش نمونه‌گیری از جامعه آماری را توصیف می‌کند. در صورتی که انتخاب تصادفی امکان‌پذیر نباشد، باید توضیح داده شود چرا روش بخصوصی برای انتخاب نمونه در نظر گرفته شده است و آیا نمونه مورد بررسی معرف جامعه بررسی هست، یا خیر. توصیفی دقیق از نمونه‌ها می‌تواند به خواننده طرح پیشنهادی در تعیین این امر کمک کند که آیا نتایج بررسی به اندازه مورد نظر قابل تعمیم است یا خیر.

### / ابزار اندازه‌گیری

هدف از یک پروژه تحقیقاتی، بررسی روابط بین سازه‌هاست. به هر حال، از آنجا که معمولاً اندازه‌گیری مستقیم سازه‌ها امکان‌پذیر نیست، باید شاخصهایی را انتخاب کنیم و یا بسازیم که آن سازه‌ها را تا آنجا که ممکن است نزدیک به واقعیت تخمین بزنند. چنانچه وسیله از قبل ساخته شده باشد، طرح پیشنهادی باید شواهدی حاکی از پایایی و اعتبار آن را برای هدف بررسی، ارائه کند. چنانچه محقق وسیله را خود بسازد، باید روش تهیه آن را به صورت خلاصه بیاورد. این خلاصه باید مراحل را دربرگیرد که برای دستیابی به اعتبار و پایایی اطلاعات حاصل از این ابزار، لازم است طی شود. اگر این امر آنچنان به جزئیات طرح بیفزاید که در تداوم طرح پیشنهادی اختلال ایجاد کند، ممکن است ترجیح داده شود که این جزئیات به جای متن اصلی در یک پیوست بیاید.

## / شیوه های انجام کار

در این بخش از طرح تحقیق، محقق توصیف می‌کند که بررسی خود را چگونه انجام می‌دهد تا در صورتی که روابطی وجود داشته باشند، مشاهده شوند. در واقع محقق می‌گوید: «اگر این فرضیه درست باشد، پس این نتایج مشاهده خواهند شد.» محقق با طراحی بررسی، آشناکارا به عنوان عملیاتی که امکان مشاهده روابط فرضی را می‌دهد، بنیان بررسی را پی می‌نهد.

توصیف دقیق شیوه انجام بررسی، لازمه اولیه طرح پیشنهادی تحقیق است. در تحقیقات پیمایشی، نگارش این بخش تقریباً ساده است، زیرا روش کار صرفاً شامل فرستادن پرسشنامه‌ها برای تکمیل کردن، و برگرداندن آنها و یا مصاحبه است. لکن، در این شیوه بررسی نیز کلیه مراحل، یعنی تهیه پرسشنامه و یا برنامه مصاحبه، آموزش پرسشگران، دادن آموزش به آنها در مورد چگونگی دستیابی به آزمودنیها و انجام مصاحبه، باید توضیح داده شوند.

در تحقیق آزمایشی، روشهای کار پیچیده‌تر است. در این بخش، نگارنده باید فهرست گروهها را بیاورد، نحوه دستکاری برنامه ریزی شده برای هر گروه را مرحله به مرحله مشخص کند و هر تیمار را به سوالات و فرضیه‌ها ربط دهد. این مراحل باید کاملاً به صورت کاربردی مشخص شوند. نگارنده می‌تواند با قرار دادن مبنای شیوه کار بر فرضیه‌ها، تفسیر مستقیم و غیر مبهم نتایج را تسهیل کند. احتمال وجود متغیرهای مخدوش‌کننده نتایج، یعنی متغیرهایی که می‌توانند بر تفاوت‌های موجود در نمره‌های ملاکی اثر گذارده، ولی در بین متغیرها به حساب نیامده باشند، در اینجا باید مورد نظر قرار گیرند. در بخش شیوه انجام کار، نگارنده یا محقق، باید چگونگی روش کنترل این متغیرها را مشخص کند. برای مثال، ممکن است محقق تفاوت‌های موجود بین دانش‌آموزان و معلمان را با تخصیص تصادفی دانش‌آموزان و معلمان به گروههای گواه و آزمایشی، کنترل کند. همچنین در تحقیقی نظیر تحقیق مذکور می‌توان وضعیت فیزیکی و تجهیزات را نیز برای هر دو گروه یکسان نگاهداشت. این بخش باید شامل کلیه مراحل باشد که برای تیمارهای آزمایش و گواه تعریف کاربردی می‌دهد.

ضرورت ندارد مدارکی چون مواد درسی و خواندنی که در آزمایش از آنها استفاده می‌شود، در متن اصلی طرح پیشنهادی تحقیق آورده شود، زیرا بحث چنین جزئیاتی معمولاً تداوم طرح پیشنهادی را گسسته می‌کند. پیشنهاد می‌شود نگارنده، اینگونه مدارک را در یک پیوست بیاورد، ولی آنها را به صورت خلاصه و در عین حال روشن در بخش شیوه انجام کار توصیف کند. به هر



حال، در این بخش توضیح هرگونه تفاوتی در نحوه ارائه این مطالب به گروههای مختلف شرکت کننده در آزمایش، امری ضروری است.

پس از آماده کردن پیش نویس انجام کار، به منظور حصول اطمینان از اینکه تمام مراحل ضروری برای پاسخگویی به تمام سؤالات و آزمون تمام فرضیه ها توصیف شده است، باید آن را بازخوانی کرد. کامل بودن این بخش را می توان با پاسخ به این سؤال کنترل کرد که «آیا خواننده می تواند با دنبال کردن مراحل توصیف شده، همین تحقیق را انجام دهد؟». چنانچه بتوان به این سؤال پاسخ مثبت داد، می توان اطمینان داشت که این بخش کامل است.

### // تحلیل داده ها

قسمت بعدی طرح پیشنهادی تحقیق، روشهای آماده سازی و ارائه داده ها و روشهای آماری مورد استفاده، در دو بخش (۱) سازماندهی داده ها و (۲) روشهای آماری، آورده می شود.

### / سازماندهی داده ها

ارائه نتایج یک تحقیق بر حسب نحوه تنظیم یافته های آن می تواند صورتهای مختلف داشته باشد. ترتیب نتایج تحقیق به طور سازمان یافته، امری است که قبلاً باید برنامه ریزی شده باشد. این امر به بهترین وجه با رجوع به سؤالات و فرضیه های بررسی، امکان پذیر است. برنامه ریزی قبلی در سازماندهی و ارائه داده ها، این امکان را برای محقق فراهم می کند که تصمیم بگیرد آیا اطلاعاتی که گردآوری می شوند به سؤالات تحقیق مربوط هستند یا خیر. کسانی که از این مرحله می گذرند، در مراحل بعدی می فهمند که زمان و هزینه قابل ملاحظه ای را برای گردآوری اطلاعات نامرتبط، صرف کرده اند.

جدول، نمودارها و نگاره ها، در سازماندهی و خلاصه کردن کل داده ها ابزار ضروری هستند. در حالی که تحقیق هنوز در مرحله برنامه ریزی است، محقق باید بتواند نحوه سازماندهی و ارائه داده ها را در قالب جدول، تصور کند. در این مرحله محقق باید اجزای اطلاعات قابل دسترس برای هر آزمودنی را فهرست کند، تصمیم بگیرد چگونه اطلاعات ارائه و خلاصه خواهند شد و نیز تعیین کند که چه روشهای آماری به کارگرفته خواهند شد.

## / روشهای آماری

طرح بررسی تبیین می‌کند که چه روشهای آماری باید به کار گرفته شود و عکس این عمل صادق نیست. به عبارت دیگر، محقق تعیین می‌کند که چه طرحی او را قادر می‌سازد تا روابط مفروض را مشاهده کند، سپس او روش آماری را که مناسب سؤا‌لهای مطرح شده و ماهیت داده هاست، انتخاب می‌کند. محقق ابتدا روش آماری خاصی را انتخاب نمی‌کند تا بعد از آن بررسی خود را مناسب با آن روش آماری طراحی کند. متداولترین روشهای آماری در فصول قبل توصیف شدند. این روشها برای استفاده بهتر در جدول ۱-۱۴ (آمار توصیفی) و جدول ۲-۱۴ (آمار استنباطی)، خلاصه شده‌اند. جدول ۱-۱۴ برای کمک به شناسایی شاخصهایی که می‌توانند برای توصیف داده‌های یک بررسی، به طور خلاصه، مورد استفاده قرار گیرند، طراحی شده است. نوع مقیاس اندازه‌گیری که متغیر وابسته را مشخص می‌کند، تا حدی آماره مناسب را نیز تضمین می‌نماید. بنابراین، سطرهای جدول با عناوین فاصله‌ای، رتبه‌ای و اسمی مشخص می‌شوند. فهرست موارد استفاده آماره‌های توصیفی در ستونهای ۱، ۲ و ۳ آمده است. متداولترین موارد استفاده از این آماره‌ها عبارتند از:

- ۱- فراهم آوردن شاخصی به منظور توصیف یک گروه و یا تفاوت بین گروهها (اندازه‌گیری گرایش مرکزی)
- ۲- فراهم آوردن شاخصی به منظور توصیف تغییر پذیری یک گروه و یا تفاوت در تغییرپذیری گروهها (اندازه‌های تغییر پذیری)
- ۳- یافتن فردی در یک گروه (شاخصهای مکان یابی)
- ۴- فراهم آوردن شاخصی به منظور توصیف رابطه بین متغیرها در یک جامعه (اندازه‌های همبستگی)
- ۵- توصیف چگونگی تقسیم یک مجموعه به دو زیرمجموعه
- ۶- توصیف چگونگی تعامل بین دو یا چند متغیر در رابطه بایک معیار (اندازه‌های تعامل)

### جدول ۱-۱۴ - آمار توصیفی

هدف آمار:		(۱) گرایش مرکزی		(۲) تشریح پذیری		(۳) مکان‌یابی	
نوع مقیاس شمر	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه
واحد	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه	یک‌گروه
فصل‌های	میانگین	انحراف معیار	انحراف معیار	انحراف معیار	انحراف معیار	انحراف معیار	انحراف معیار
رویه‌های	ماه	انحراف	انحراف	انحراف	انحراف	انحراف	انحراف
زبان	ماه‌ها	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی
اسمی	ماه	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی
اسمی	ماه	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی	چهارگی

ادامه جدول ۱-۱۴

	هدف آمار		(۳) مستقیماً	نوع مقایسه دسته
	(۱) مثال	(۲) آزمون		
پیشاز یکگروه	یکگروه	پیشاز یکگروه	پیشاز یکگروه	نوع مقایسه دسته
تفاوت‌ها در زمان مشاهده شده بین گروه‌ها	تفاوت بین میانگین‌های مشاهده شده و میانگین مورد انتظار عاشق (زمان) مشاهده شده)	تفاوت بین نضریب‌های پرسون برای همان متغیرها در دو گروه	ضریب پرسون	نوع مقایسه دسته
		تفاوت در ضریب‌های اسپیرمن برای همان متغیرها در دو گروه	$\rho$ اسپیرمن یا tau کدال* یا $\eta^2$ *	رئای

ادامه جدول ۱۳-۱

مدن آمار

(۱)	(۲)	(۳)	(۴)	نوع مقیاس ضمیمه
تامل	آزادری	بکت‌گروه	مبستگی	رابط
بیش از بکت‌گروه	بیش از بکت‌گروه	بیش از بکت‌گروه	بیش از بکت‌گروه	نسبت با درصد
تفاوتها بین تامل مشاهده شده در بین گروههای مورد انتظار	تفاوتها بین فراوانیهای مشاهده شده و مورد انتظار	تفاوتها در نسبتها با درصدها	تفاوتها بین دورتهای تقویم برای همان ضمیمه یا در دو گروه	تفاوت بین مسبستگیهای دورتهای نقطه‌ای

• این آماره در کتاب توصیف شده است، ولی آن را می‌توان در بسیاری از جزئیات درسی یافت.

جدول ۲-۱۴ - آمار استنباطی

هدف از روش آماری					
(۲)	(۱)	(۱)	(۱)	نوع مقیاس	دسته
مکانیابی	تیم پذیری	گرایش مرکزی	بیکسای	نوع مقیاس	دسته
بیشاز بیکسای	بیکسای	بیکسای	بیکسای	نوع مقیاس	دسته
بیکسای	بیکسای	بیکسای	بیکسای	نوع مقیاس	دسته
خطای شمار	خطای شمار	آزمون گویا	خطای شمار	نوع مقیاس	دسته
تفاوت سرت *	انتاوه گیری *	با آزمون گویا	تعمیل واریانس	نوع مقیاس	دسته
		همگنی واریانس	بیکسویه	نوع مقیاس	دسته
		آزمایه های F-McN		نوع مقیاس	دسته
			آزمون جانه، آزمون	نوع مقیاس	دسته
			ملاصها، تعیل واریانس	نوع مقیاس	دسته
			بیک سویه	نوع مقیاس	دسته
			کورسکال - رانس *	نوع مقیاس	دسته
			آزمون ریس	نوع مقیاس	دسته

## ادامه جدول ۱۳-۲

هدف از روش آماری		(۱)		(۲)		(۳)	
		نمایل		در نمونه		مبتنی	
نوع مقیاس متر	رابطه	یکگروه	یکگروه	یکگروه	یکگروه	یکگروه	یکگروه
		پیش از	پیش از	پیش از	پیش از	پیش از	پیش از
		یکگروه	یکگروه	یکگروه	یکگروه	یکگروه	یکگروه
فصلی	رابطه	آزمون F برای	آزمون F برای	آزمون T برای	آزمون T برای	آزمون T برای	آزمون T برای
		تحلیل واریانس چند عاملی	تحلیل واریانس چند عاملی	تبدیل Z	تبدیل Z	تبدیل Z	تبدیل Z
				و یا آزمون F	و یا آزمون F	و یا آزمون F	و یا آزمون F
				برای خطی بودن	برای خطی بودن	برای خطی بودن	برای خطی بودن
				آزمون P	آزمون P	آزمون P	آزمون P
				استیون و یا	استیون و یا	استیون و یا	استیون و یا
				Statistical یا w*	Statistical یا w*	Statistical یا w*	Statistical یا w*
اسمی	اسمی	نظریه اقلات A*	نظریه اقلات A*	خود یا	خود یا	خود یا	خود یا
		آزمون غیر در	آزمون غیر در	آزمون دقیق غیر*	آزمون دقیق غیر*	آزمون سنی های	آزمون سنی های
		برای نظریه	برای نظریه			دورتهای	دورتهای
		اطلاعات A*	اطلاعات A*			تفصیلی	تفصیلی

\* این روشهای آماری در این کتاب نیامده است، ولی آنها را در بیشتر از کتابهای دومی آماری می توان یافت.

خانه مورد نیاز را در جدول، می‌توان با مشخص کردن عناوین سطر و ستون، که مناسب با مطالعه مورد نظر باشد، پیدا کرد. هر خانه به دو بخش تقسیم شده است و بخش مورد استفاده با در نظر گرفتن این که بررسی مورد نظر با یک گروه یا بیش از یک گروه سر و کار دارد، تعیین می‌شود (به خاطر داشته باشید که محقق می‌تواند آماره‌ای را برای مقیاس پایین‌تر اندازه‌گیری انتخاب کند و نه برعکس. برای مثال، محقق می‌تواند برای توصیف داده‌های فاصله‌ای از میانه یا نما استفاده کند، ولی نمی‌تواند برای توصیف داده‌های رتبه‌ای و اسمی از میانگین استفاده کند). به منظور تصمیم‌گیری در مورد نوع مقیاسی که برای بیان داده‌ها مورد استفاده قرار می‌گیرد، محقق باید مزایای سه مقیاس را در سطح توصیفی در نظر گیرد. داده‌های فاصله‌ای معمولاً اطلاعات بیشتری را از داده‌های رتبه‌ای، و داده‌های رتبه‌ای، اطلاعات بیشتری را از داده‌های اسمی، فراهم می‌آورند. در استنتاج آماری، آزمونهای داده‌های فاصله‌ای از قدرت بیشتری نسبت به آزمونهای داده‌های رتبه‌ای برخوردار هستند؛ بدین معنی که احتمال رد یک فرضیه صفر وقتی از اندازه‌های فاصله‌ای استفاده شده باشد، بیشتر از زمانی است که از اندازه‌های رتبه‌ای استفاده شده است. به همین ترتیب، آزمونهای رتبه‌ای از قدرت بیشتری نسبت به اندازه‌های اسمی برخوردارند. بنابر این، چنانچه امکان انتخاب باشد، محقق داده‌های فاصله‌ای را به رتبه‌ای و رتبه‌ای را به اسمی ترجیح می‌دهد.

برای مثال اگر ما برای متغیر وابسته داده‌های فاصله‌ای داشته باشیم، بنابر این، به دنبال شاخصی برای توصیف تفاوت بین گروهها هستیم. جدول، تفاوت بین دو میانگین را به عنوان آماره مناسب می‌شناسد (اگر حق انتخاب با ما بود، می‌توانستیم از تفاوت بین دو میانه و نماها استفاده کنیم، ولی این تفاوتها از قدرت کمتری نسبت به تفاوت بین میانگینها برخوردار هستند). چنانچه بررسی، ماهیتاً استنتاجی باشد، محقق تا مرحله آزمون معنی داری آماری شاخص انتخاب شده، پیش می‌رود. آماره‌های مناسب برای این منظور در جدول ۲-۱۴ فهرست شده است. در مثال ما، آزمون  $t$  ویا تحلیل واریانس یک سویه مناسب خواهد بود.

به خاطر داشته باشید که یک آماره بر مبنای همخوانی برای پاسخگویی به سؤال مورد نظر بررسی، انتخاب می‌شود. با کاربرد یک آماره پیچیده، زمانی که یک آماره ساده به همان خوبی جوابگو باشد، ثمری حاصل نمی‌شود. آماره‌ها در خدمت بررسی هستند، نه برای تسلط بر آن. توصیف ویژه برنامه‌هایی که برای گردآوری داده‌ها در پیش گرفته خواهد شد، باید در این



بخش، ارائه شود. این برنامه‌ها شامل بودجه زمانی، روشهای جایگزینی آزمودنیهایی که در طول دوره آزمایشی از دست می‌روند، برنامه‌هایی برای خنثی کردن اثرات ترتیبی در صورت لزوم و دیگر جزئیات لازم، خواهد بود. ما همیشه به دانشجویانمان می‌گوییم: «تصور کنید که بورس تحقیقات عالی را برای انجام یک بررسی گرفته‌اید، ولی روز بعد با کامیون تصادف می‌کنید. آیا یک همکار می‌تواند طرح پیشنهادی شما را بردارد و عملاً بررسی را اجرا کند؟» چنانچه به این سؤال به طور صادقانه پاسخ مثبت داده شود، قسمت تحلیل داده‌های طرح پیشنهادی کامل است.

### // اهمیت بررسی

برخی محققان ترجیح می‌دهند اهمیت مسئله را در مقدمه طرح پیشنهادی بیاورند. از طرفی، چنانچه این عنوان را در بخش بعدی مطرح کنیم، امکان ربط آن، هم به زمینه و هم به طرح بررسی، فراهم می‌شود. این بخش به بهترین وجه در دو قسمت تهیه می‌شود: (۱) دلایل و (۲) کاربردها.

### / دلایل

از آنجا که هدف غایی تحقیق، افزایش آگاهی است، نویسنده طرح پیشنهادی باید با بحث در مورد نتایج بررسی خود بر نظریه‌ها و اطلاعات موجود در آن حیطه، نشان دهد که چگونه این هدف تحقق پیدا خواهد کرد. بعلاوه او باید نشان دهد که نتایج بررسی او، تا چه حد، در حل مسائل و پاسخگویی به سؤالات آن حیطه راهگشا خواهد بود. در نهایت، نویسنده باید نشان دهد که چگونه این تحقیق می‌تواند مبنایی برای تحقیقات آتی باشد. همچنین، نویسنده باید توضیح دهد که تجربه و مهارت او، همراه با امکانات موسسه مجری طرح، به چه نحو، موقعیت مطلوبی برای حل مسئله فراهم می‌آورد.

### / کاربردها

نویسنده باید بتواند خوانندگان را در مورد کاربرد بالقوه نتایج در کار تعلیم و تربیت، متقاعد سازد. این بحث باید نشان دهد که چگونه و تا چه حد دست اندرکاران تعلیم و تربیت می‌توانند از یافته‌های بررسی برای بهبود کارشان استفاده کنند. برای رسیدن به میزانی برای کاربرد بررسی

در امر تعلیم و تربیت به صورت عملی، محقق می‌تواند این سؤالها را مطرح سازد که «آیا نتایج بررسی من چیزی را در رشته تعلیم و تربیت تغییر می‌دهد؟ آیا نتایج بررسی من به معلمان، مشاوران، مدیران و برنامه ریزان آموزش و پرورش کمکی در جهت بهبود کارشان می‌کند؟» این جنبه کار می‌تواند بار زیادی از جهت جذب بودجه تحقیقاتی برای انجام بررسی داشته باشد. بسیاری از بنیادهای فرهنگی، طرحهای پیشنهادی تحقیق را بر مبنای اینکه آیا در عمل کاربرد دارند یا خیر، ارزیابی می‌کنند.

### // بودجه مالی و بودجه زمانی

تمام تحقیقات باید با در نظر گرفتن عملی بودن انجام آن برنامه ریزی شوند. بخش پایانی یک طرح باید شامل (۱) بودجه و (۲) برنامه زمانبندی باشد.

#### / بودجه مالی

محقق با بازنگری بخشهای قبلی طرح پیشنهادی، اکنون فهرستی از کارکنان، لوازم، فضا و زمانی را که برای پروژه لازم است، ارائه می‌دهد. غالب نظامهای آموزشی (دانشگاهها و مدارس)، سازمانی برای کمک به تهیه بودجه های تحقیقاتی دارند. چنین سازمانهایی می‌توانند از جهت مقررات محلی و سازمانی نقش مشورتی داشته باشند و در تبدیل نیازهای طرح بررسی به ارقام مالی کمک کنند.

#### / بودجه زمانی

محقق همچنین برنامه واقع بینانه‌ای را برای تکمیل تحقیق در زمان لازم، آماده می‌سازد. این اطلاع، به بازبین طرح پیشنهادی تحقیق کمک می‌کند و همچنین محقق را یاری می‌دهد تا ببیند برای تکمیل طرح چقدر زمان لازم دارد و همچنین باید امکاناتی را برای ارزیابی ادواری پیشرفت طرح، فراهم آورد.

### /// نقد طرحهای پیشنهادی

پس از تکمیل پیش نویس یک طرح پیشنهادی، نگارنده، یا محقق، باید دوباره آن را با یک

دید انتقادی بازنگری کند. همچنین مفید خواهد بود که از همکاران خواسته شود تا طرح پیشنهادی را بخوانند. اغلب یک همکار می‌تواند نقاط ضعف و از قلم افتاده‌ای را بیابد که برای خود نویسنده مشهود نیست.

اسمیت در گزارش خود از کمیته مشورتی تحقیقاتی، مربوط به برنامه های تحقیقاتی اشتراکی USOE، شش نقطه ضعف متداول را که کمیته در طرحهای پیشنهادی ارائه شده جهت تأمین بودجه یافته، مشخص کرده است.<sup>(1)</sup>

۱- مسئله، بی‌اهمیت است. مسائلی که در حاشیه مسائل اساسی مورد علاقه کارشناسان تعلیم و تربیت قرار دارند و یا احتمال کمی وجود دارد که چیزی به آگاهیهای آموزشی بیفزایند، به عنوان مسائلی تلقی می‌شوند که شایستگی پشتیبانی مالی را ندارند. اسمیت مثالی از یک طرح بررسی برنامه آموزش بزرگسالان برای هیئت امنای کتابخانه می‌آورد و چنین نقد می‌کند: «نه تنها دارای ارزش حاشیه‌ای در زمینه تعلیم و تربیت است و بیشتر بر مبنای عمل، نه تحقیق استوار است، بلکه خیلی بی‌پرده بگوییم، باید آن را مسئله کم‌اهمیتی تلقی کرد.»

۲- حدود مسئله مشخص نشده است. مثال کلاسیکی از گود و اسکیتس نمونه‌ای بسیار خوب برای این نقطه ضعف است.<sup>(2)</sup> نامه یک دانشجوی فوق لیسانس به عضو عالی رتبه آموزش و پرورش آلاسکا نشان می‌دهد که این دانشجو عنوان پایان نامه‌اش را «آموزش زبان انگلیسی، آن گونه که در واحدهای تحصیلی ملل انگلیسی زبان دنیا مشهود است»، انتخاب کرده است. در بند دوم از عضو عالی رتبه می‌پرسد: «آیا شما کتابهای جالبی درباره آلاسکا و تاریخچه‌اش، مسائل اقتصادی، تجارت، واردات، صادرات، روابط انسانی و مذهبش و... می‌شناسید؟»

محقق برای تهیه یک طرح پیشنهادی عملی، باید مطالعه خود را بر مرکز مسئله متمرکز کند. این بدان معنی نیست که یک بررسی هیچوقت نباید چندین متغیر را دربرگیرد. محقق باید به جنبه‌هایی از مسئله بپردازد که بتوان در یک بررسی به آنها پرداخت. توده‌ای از متغیرهای مربوط می‌توانند و معمولاً باید در بررسی گنجانده شوند، ولی محقق باید از کوششهای غیرقابل مهار و بسیار وسیع اجتناب کند.

1- Smith, G. R. (1963). A critique of proposals submitted to the Cooperative Research Program. In J.A. Culbertson and S.P. Hency (eds.). Educational Research: New Perspectives (ch.17).

2- Good, C.V. and D. E. Skates (1954) Methods of Research (PP.82-83) New York: Appleton.

۳- اهداف، فرضیه ها و سؤالاتها بیش از حد گسترده تعیین شده‌اند. نویسندگان طرحهای پیشنهادی اغلب برای تبیین هدفها، فرضیه ها و سؤالاتهای خود دچار وسوسه کلی گویی می‌شوند. در چنین مواردی فرد به هنگام خواندن بخش روش تحقیق چنین در می‌یابد که بررسی برنامه ریزی شده واقعی، قابلیت برآوردن هدفهای بزرگ تعیین شده را ندارد. در پاره‌ای از مواقع، هدفها و فرضیه ها با چنان واژه های کلی بیان می‌شوند که فرد برای آنکه کشف کند بررسی واقعاً درباره چیست، باید به بخش روشها مراجعه کند. آشکار است که هدفها و روشها باید منطبق باشند. کمیته مشورتی تحقیق، عدم انطباق این دو را در بسیاری از طرحهای تحقیقاتی پیشنهادی یافته است.

۴- روش فاقد جزئیات است. اسمیت چنین خاطر نشان می‌سازد که «محققی که بیشتر از آنکه اضافه کند، حذف می‌کند، نباید از کمیته انتظار داشته باشد که جزئیات روش را از بین خطوط بخواند. کمیته، فاقد تمایل و نیز بصیرت برای این عمل است. به خاطر داشته باشید که روشها باید به اندازه کافی کامل باشند تا اجازه تکرار را بدهند».

۵- از یک طرح ساده برای بررسی یک مسئله پیچیده استفاده شده است. طرح یک بررسی باید با مسئله همخوانی داشته باشد. یک مقایسه ساده میانگینها در بین دو گروه، وقتی فقط یک متغیر مورد نظر است، مناسب است. بررسیهای پیچیده تر نیاز به طرحهای پیچیده تر دارد. «متغیرهای مربوطه مورد نظر قرار نگرفته‌اند و یا تا حدودی حذف شده‌اند. شکست در تشخیص متغیرهای نامربوط دخیل در بررسی، اشتباه جدی در یک طرح پیشنهادی تحقیق است. محقق باید نشان دهد که از چنین متغیرهایی آگاه است و توضیح دهد که چگونه می‌خواهد آنها را در طرح خود در نظر گیرد».

### /// اهمیت تکمیل طرح پیشنهادی تحقیق قبل از گردآوری داده‌ها

یک طرح پیشنهادی روشن، که خوب و کامل تبیین شده باشد، نشان دهنده این واقعیت است که محقق آینده‌نگر عملاً برای انجام بررسی خود آماده است. چنین طرحی نشان می‌دهد که او می‌داند چه می‌کند، چرا آن را انجام می‌دهد و چگونه آن را اجرا می‌کند. چنانچه محقق آینده‌نگر نتواند یک طرح پیشنهادی کامل و منسجم ارائه کند، مسلماً هنوز برای شروع مرحله گردآوری داده های پروژه آمادگی ندارد. محققان بی تجربه، غالب اوقات تمایل دارند بگویند: «بگذار فعلاً داده ها را گردآوری کنم و بعد تصمیم بگیرم که با آنها چه کنم.»

ممکن است چنین به نظر آید که گردآوری داده‌ها همزمان با نوشتن طرح پیشنهادی، صرفه جویی در وقت است، ولی به ندرت چنین است. تعداد بیشماری ساعت کار و هزاران دلار با چنین روشی هدر رفته است. قبل از آنکه چهارچوب طرح پیشنهادی مشخص شده باشد، محقق نمی‌تواند دقیقاً مطمئن باشد که به چه داده‌هایی نیازمند است و یا بهترین راه برای پرداختن به این اطلاعات در پرتو هدف بررسی چیست. کسانی که باید در یک زمان مشخص کارشان را تمام کنند، باید تاریخی را برای تکمیل طرح پیشنهادی تعیین کنند که به اندازه کافی جلوتر از زمان مورد نظر برای اتمام کل پروژه باشد.

این بدان معنا نیست که بگویم گردآوری اطلاعات هیچوقت نباید قبل از طرح پیشنهادی صورت بگیرد. یک بررسی مقدماتی می‌تواند در برنامه ریزی پروژه، وارسی ابزار گردآوری داده‌ها، تعیین عملی بودن شیوه‌ها و غیره، مفید باشد. به هر حال، از این بررسی‌های مقدماتی، داده‌های اساسی برای بررسی حاصل نمی‌شود. به آنها باید به چشم یک عملیات آزمایشی و نه چیزی بیشتر نگاه کرد.

### /// ملاحظات اخلاقی و حقوقی

مهمترین مسئله در اجرای یک تحقیق، پیروی دقیق از معیارهای اخلاقی است. محققان در قبال آزمودنیها و شغلشان، تعهداتی دارند. معیارهای اخلاقی تحقیق که از قوانین انجمن روان شناسی آمریکا گرفته شده است، می‌تواند راهنمای مفیدی در تحقیقاتی باشد که در آن از انسان به عنوان آزمودنی استفاده می‌کنند:

۱- تنها در مواردی که یک مسئله علمی از اهمیت برخوردار است و بررسی آن از هیچ طریق دیگری امکان پذیر نیست، روان شناس مجاز است که آزمودنیهای تحقیق، چه کودکان و چه بزرگسالان، را تحت فشار فیزیکی و یا عاطفی قرار دهد.

۲- در صورتی که احتمال زیادی برای اثرات زیانبار بعدی وجود داشته باشد، تنها در صورتی تحقیق قابل اجراست که آزمودنیها و یا افراد مسئول آنها کاملاً از این احتمال آگاه باشند و به هر صورت با شرکت در تحقیق موافقت کنند.

۳- روان شناس شدیداً به احتمال اثرات زیان‌آور بعدی توجه دارد و از آنها پرهیز می‌کند و یا آنکه به محض آنکه طرح آزمایش اجازه دهد، آنها را قطع می‌کند.<sup>(۱)</sup>

## // تعهد نسبت به آزمودنیها

اگر محقق با انسانها به عنوان آزمودنی سر و کار دارد، باید برای امانت و انسانیت، احترام قائل باشد. سه زمینه عمده‌ای که باید مورد توجه باشند، عبارتند از:

(۱) حفاظت از افراد آزمودنی در مقابل صدمات، (۲) احترام به حق آنها در آگاهی از ماهیت و هدف بررسی و نیز حق آنها در موافقت یا عدم موافقت برای شرکت در بررسی (حق موافقت آگاهانه) و (۳) احترام به حریم خصوصی آزمودنیها.

## / حفاظت آزمودنیها در مقابل صدمات

آزمودنیها باید در مقابل صدمات فیزیکی، روانی و اجتماعی حفاظت شوند. در تحقیقاتی که بودجه آنها را دولت تأمین می‌کند و در آنها از انسان به عنوان آزمودنی استفاده می‌شود، سالهاست قوانین وجود داشته است، ولی در سال ۱۹۷۵ مجموعه قوانین جامعتری وضع شد که کلیه تحقیقاتی را که بودجه آنها توسط سازمانهای دولتی تأمین می‌شود، در بر می‌گرفت. در نتیجه وضع این قوانین، غالب مدارس عالی و دانشگاهها و بسیاری از مدرسه‌ها، هیئت بازننگری سازمانی (IRB)<sup>(۱)</sup> را تشکیل دادند و این هیئت، بازننگری اولیه در کلیه طرحهای تحقیقاتی را که در آنها از انسان به عنوان آزمودنی استفاده می‌شود، ضروری دانست. کلیه تحقیقات در زمینه تعلیم و تربیت در زمره اینگونه تحقیقات بودند. از اجرای این قوانین استقبال زیادی نشد و صدای اعتراض سازمانهای متعددی، مبنی بر اینکه امر تحقیق بیش از حد در فشار قوانین قرار گرفته است، بلند شد. کنگره فرمان تشکیل کمیسیون ملی حفاظت آزمودنیهای انسانی را در تحقیقات زیست پزشکی و رفتاری، صادر کرد و کمیسیون در گزارش نهایی خود، کاهش گستره مقرراتی را که در قانون جدید ۲۶ ژانویه ۱۹۸۱ وضع شده بود، پیشنهاد کرد.<sup>(۲)</sup>

قانون جدید، مقررات دولتی را منحصر به پژوهشهایی کرد که بودجه آنها توسط اداره بهداشت و خدمات انسانی، تأمین می‌شود و مقوله‌های تحقیقاتی متعددی را نیز از شمول مقررات معاف نمود. پنج مقوله اصلی تحقیق که از این مقررات معاف شدند، عبارت بودند از:

۱- تحقیقاتی که در محیطهای آموزشی مرسوم یا مورد قبول عام صورت می‌گیرد، شامل

1- Institutional Review Board (IRB)

2- Code of Federal Regulations 45 (1982; Part 46, pp. 100-104). Washington: U.S. Government Printing Office.

تجربیات آموزشی متعارف، نظیر: الف - تحقیق در زمینه راهبردهای آموزشی در مورد تعلیم و تربیت خاص و معمولی و ب - تحقیق در زمینه تأثیر و یا مقایسه روشهای تدریس، برنامه های درسی و شیوه های اداره کلاس.

۲- تحقیقات در زمینه استفاده از آزمونهای تعلیم و تربیت (آزمونهای شناختی - تشخیصی استعداد و پیشرفت تحصیلی) در صورتی که اطلاعات به دست آمده از این منابع به نحوی ثبت شوند که آزمودنیها، چه به صورت مستقیم و چه از طرق دیگر، قابل شناسایی نباشند.

۳- تحقیقات پیمایشی و یا روشهای مصاحبه‌ای، مگر آنکه شرایط زیر در آنها صادق باشد: الف - پاسخها به گونه‌ای ثبت شوند که آزمودنیها یا به صورت مستقیم و یا با علایم شناسایی مربوط به آزمودنیها، قابل شناسایی باشند، ب - در صورتی که پاسخهای آزمودنی به خارج از حوزه تحقیق تسری یابد و باعث شود آزمودنی در خطر بزهکاری قرار گیرد، مسئولیت مدنی او به خطر افتد و یا موقعیت مالی و یا امکان استخدام او، آسیب ببیند و ج - تحقیق در مورد جنبه‌های حساسیت آفرین رفتار آزمودنی، نظیر سوابق قانون شکنی، استعمال مواد مخدر، رفتارهای جنسی یا استفاده از الکل، باشد. تمام تحقیقاتی که از روش پیمایشی یا مصاحبه استفاده می‌کنند، بدون استثنا، در صورتی که آزمودنیها در انتخابات پیروز شده باشند و یا نامزد مقامات رسمی یا مشاغل عمومی باشند، مشمول معافیت می‌شوند.

۴- تحقیقاتی که از روش مشاهده رفتار عمومی استفاده می‌کنند (این طبقه شامل مشاهده توسط شرکت کنندگان نیز می‌شود)، در شرایط زیر، از این معافیت مستثنی هستند:

الف - مشاهدات به نحوی ثبت شوند که آزمودنیها به صورت مستقیم و یا به وسیله نشانه هایی که به آزمودنی مربوط می‌شود، قابل شناسایی باشند.

ب - در صورتی که پاسخهای آزمودنی به خارج از حوزه تحقیق تسری یابد، باعث شود موقعیت آزمودنی در معرض خطر بزهکاری قرار گیرد یا مسئولیت مدنی او به خطر افتد و یا موقعیت مالی و امکان استخدام او آسیب ببیند.

ج - تحقیق به جنبه های حساس رفتار آزمودنی نظیر سابقه قانون شکنی، استعمال مواد مخدر، رفتارهای جنسی و یا استفاده از الکل، بپردازد.

۵- تحقیقاتی که به گردآوری و بررسی اطلاعات، مدارک، یادداشتها، نمونه های آسیب شناسی و نمونه های قابل تشخیص موجود می‌پردازند، در صورتی که منابع در دسترس عموم

باشند و یا چنانچه اطلاعات ثبت شده توسط محقق به گونه‌ای باشد که آزمودنیها به صورت مستقیم و یا به وسیله نشانه‌هایی که به آزمودنیها مربوط می‌شود، قابل شناسایی نباشند.<sup>(۱)</sup> هرچند این مقررات عملاً تحقیقات در زمینه تعلیم و تربیت را از مقررات مستقیم دولتی مستثنی می‌سازد، لیکن IRB در جایگاه خود باقی می‌ماند و احتمال دارد مؤسسات آموزشی تصمیم بگیرند زیر لوای مقررات دیگری قرار گیرند که با ویژگی‌شان محدودتر یا گسترده‌شان وسیعتر از مقررات دولتی است. فشار غیر رسمی بر مؤسسه‌ها موجب شده است که آنها به تبعیت از مقررات وسیعتری وادار شوند. با در نظر گرفتن سردرگمی که در اثر وضعیت قانونی فعلی وجود دارد، محقق باید جهت تعیین قوانین با IRB مشورت کند و مقررات جاری را در قوانین دولتی که در کتاب مفاد قوانین دولتی (۴۵ CFR بخش ۴۶)<sup>(۲)</sup> آمده و در غالب کتابخانه‌ها موجود است، بررسی کند.

در مورد تحقیقاتی که بودجه‌شان توسط اداره بهداشت و خدمات انسانی تأمین می‌شود و در تحقیقاتی سوای طبقات مشمول معافیت که از انسان به عنوان آزمودنی استفاده می‌کنند، مقررات دولتی، تعیین این امر را که آیا شیوه‌های تحقیق، آزمودنیها را در معرض خطر قرار می‌دهد یا خیر، الزامی می‌کند. چنانچه محقق بتواند نشان دهد که خطرهای جسمانی، روانی و اجتماعی ناشی از پروژه پیشنهادی، بیشتر از آنچه که آزمودنیها در زندگی روزمره و یا در ضمن اجرای آزمایشها و آزمونهای جسمانی و روان‌شناختی روزمره با آن مواجهند، نیست، آزمودنیها دیگر در معرض خطر تلقی نمی‌شوند.<sup>(۳)</sup> چنانچه تشخیص داده شود که آزمودنیها «در معرض خطر» قرار می‌گیرند، محقق باید به این بحث پردازد که (۱) ارزش دانشی که احتمالاً از این تحقیق کسب می‌شود، فواید از صدمات بالقوه است، (۲) تمام آزمودنیها از مراحل کار کاملاً مطلع خواهند بود و موافقت داوطلبانه و کتبی آنها کسب خواهد شد و (۳) خدمات پشتیبانی پزشکی و دیگر خدمات، در دسترس شرکت کنندگان در تحقیق قرار خواهد گرفت. چنانچه این شرایط فراهم شود، معمولاً IRB کار پیشنهادی را تصویب می‌کند. معمولاً IRB خود رأساً جهت رد یا تأیید طرحهای تحقیقاتی هیچ اقدامی نمی‌کند.

کسب موافقت آگاهانه از افراد کم سال و یا کسانی که قادر به فهماندن مقاصد خود نیستند، مسئله‌ای غامض است. مواردی از این قبیل برای تصمیم‌گیری به دادگاههای ایالتی و دولتی



ارجاع شده و بسیاری از آنها معلق باقی مانده‌اند. به محققانی که با این مشکل روبرو هستند، توصیه می‌شود با رئیس IRB مؤسسه خود (در صورتی که IRB در مؤسسه آنها وجود داشته باشد) و یا مشاوران قانونی مؤسسه یا سازمان خود، مشورت کنند. دفتر حفاظت، وابسته به اداره بهداشت و خدمات انسانی، در مقابل خطرات ناشی از تحقیق، آماده ارائه مشاوره در اینگونه موارد است و احتمال دارد بتواند اطلاعات روز را در زمینه قوانین مربوط به گروههای مختلف، ارائه کند.

### / مسئله موافقت آگاهانه

هرچند امروزه مقررات دولتی تنها در موارد محدودی موافقت رسمی آزمودنی را طلب می‌کند، اما بسیاری از موازین اخلاقی حرفه‌ای که از سوی دولت وضع نشده‌اند، مشعر بر این هستند که کلیه آزمودنیها باید کاملاً از اهداف و شیوه عمل باخبر باشند و به آنها کاملاً حق انتخاب شرکت کردن یا نکردن در تحقیق داده شود. محقق باید مطلوبیت آگاهی دادن به آزمودنیها را در مورد تحقیق و گرفتن موافقت رسمی از آنها را در نظر داشته باشد.

در برخی موارد، ممکن است گفته شود که اطلاع قبلی از اهداف بررسی می‌تواند نتایج را جهت دار کند. در چنین مواردی بازهم باید ویژگی شرکت داوطلبانه را، حتی اگر اهداف تحقیق کاملاً و با دقتاً قبل از بررسی گفته نشود، مد نظر داشت. در مواردی که بحث نکردن راجع به اهداف بررسی قبل از انجام آن، حائز اهمیت است، پس از پایان اجرای بررسی، باید توضیح کاملی به آزمودنیها داده شود. تمامی مقررات حرفه‌ای در زمینه معیارهای اخلاقی، اجبار آزمودنیها را به شرکت در بررسی و توضیح دروغین در مورد هدف بررسی را، منع می‌کند.

### / حق حفظ حریم خصوصی

این توافق حرفه‌ای در سطح وسیع وجود دارد که کلیه آزمودنیها، برای حفظ مسائل خصوصیشان حق ذاتی دارند. در صورتی که نام و دیگر اطلاعات برای شناسایی فرد لازم نباشد، توصیه می‌شود چنین اطلاعاتی اصلاً گردآوری نشود. در صورتی که به دلیل امر پیگیری یا اهداف دیگر، گردآوری چنین اطلاعاتی ضروری باشد، مسئولیت حفظ محرمانه این اطلاعات و کنترل دستیابی به آنها، بر عهده محقق است. به طور کلی، توصیه می‌شود که تنها محقق و آن

دسته از اعضای گروه تحقیق که آگاهی از اطلاعات فردی، از جهت اهداف مربوط به کار، برایشان ضروری است، به آن دسترسی داشته باشند. در صورتی که لازم باشد دانشجویان و یا همکاران دیگر به اطلاعات دسترسی داشته باشند و یا ضروری باشد که آنها را در آرشیو قرار دهند، امحاً کلیه اطلاعات شناسایی فرد از این پرونده ها، بر عهده محقق مسئول طرح است.

در مورد تحقیقاتی که بودجه شان از طرف اداره بهداشت و خدمات انسانی تأمین می شود، محقق باید ذخیره سازی ایمن و نگاهداری کلیه اسامی و دیگر اطلاعات شناسایی را تضمین کند. در صورتی که پروژه تحقیقاتی به گردآوری اطلاعاتی می پردازد که می تواند فردی را مقصر قلمداد کند و یا به موقعیت اجتماعی او ضربه بزند، این ضرورت حتی اهمیت بیشتری می یابد. ناتوانی در حراست از این نوع اطلاعات می تواند به تحت تعقیب قرار گرفتن از جانب قوانین دولتی (عدم کفایت برای دریافت کمکهای مالی جهت تحقیق در آینده) یا اعاده حیثیت طبق قوانین مدنی به علت صدمات شخصیتی، منجر شود.

### // تعهدات حرفه‌ای

همچنین محقق در مقابل مصرف کنندگان تحقیق نیز مسئول است. همانند تحقیقات در رشته های دیگر، غالب بررسیهای تحقیقاتی در زمینه تعلیم و تربیت، در مجلات، تک نگاریها، کتابها و دیگر رسانه ها چاپ شده و به آنها ارجاع داده می شود و همچنین توسط دست اندرکاران این رشته مورد استفاده قرار می گیرد. محقق اخلاقاً متعهد است که بررسی را چنان طرح ریزی کند که نتایج حاصل از آن، اطلاعات گمراه کننده ای را ارائه ندهد. حتی فراتر از این، محقق متعهد است که نتایج را، هر چه باشد، دقیقاً و صادقانه گزارش کند. تحقیق نباید به گونه ای گزارش شود که موجب گمراهی شود. گزارشی که گمراه کننده باشد، در واقع سوء استفاده جدی از مسئولیت حرفه ای محقق تلقی می شود.

### /// خلاصه

یک طرح پیشنهادی تحقیق، برنامه ای قدم به قدم برای کشف معرفتهای جدید است. در مرحله نگارش طرح پیشنهادی تحقیق، الهامات و بینشهای محقق به شکل عینی متبلور می شود. در یک طرح پیشنهادی تحقیق مقوله های اطلاعاتی متعددی باید گنجانده شود.

بیان مسئله، همراه با تعاریف غیر مبهم واژه ها، باید در بخش اولیه طرح پیشنهادی تحقیق

آورده شود. به دنبال این بخش، مرور پژوهشهای پیشین می‌آید. یک مرور خوب از پژوهشهای پیشین، آنچه را که تا زمان حال در مورد مسئله شناخته شده است، نشان می‌دهد. این امر زمینه را برای تبیین فرضیه‌ها، با توجه به روابط بین متغیرهای مورد بررسی هموار می‌کند. بعلاوه، این بخش باید با این هدف نوشته شود که زمینه را برای تفسیر نتایج فراهم کند.

در بحث روش شناختی، که به دنبال بخش مقدمه می‌آید، روشهای انتخاب آزمودنی، روشهای گردآوری داده‌ها، شیوه‌های مشاهده و روشهای اندازه‌گیری، همگی با جزئیات لازم به گونه‌ای توصیف می‌شوند که خواننده بتواند با دنبال کردن مراحل پیشنهادی، دقیقاً همانند محقق اصلی تحقیق را انجام دهد.

بخش بعدی طرح پیشنهادی تحقیق به توصیف روشهای مورد استفاده در ارائه داده‌ها، نظیر جداول، نمودارها، نگاره‌ها و معرفی روشهای آماری که برای تحلیل داده‌ها مورد استفاده قرار خواهد گرفت، می‌پردازد.

بحث در مورد اهمیت بالقوه بررسی باید به دنبال این بخش آورده شود. در اینجا محقق باید نشان دهد که چگونه یافته‌های بررسی به دانش موجود در این زمینه می‌افزاید و نتایج آن برای نظریه و تحقیق در زمینه مورد نظر، چه مفهومی دارد. بحث در مورد کاربردهای یافته‌ها، در عمل می‌تواند برای خوانندگانی که مایلند اهمیت تحقیق پیشنهادشده را ارزشیابی کنند، کمک‌کننده باشد. بخش نهایی طرح پیشنهادی حاوی برنامه‌زمانی و هزینه‌برآورد شده برای بررسی است. این اطلاعات می‌تواند در امر ارزیابی کلی طرح پیشنهادی برای خوانندگان مورد استفاده قرار گیرد. مسئله‌ای بسیار حائز اهمیت در برنامه ریزی تحقیق، در نظر گرفتن معیارهای اخلاقی است. برای آزمودنیها باید حق موافقت آگاهانه را قائل بود. آنها باید از هرگونه آسیبی حفظ شوند و برای مسائل خصوصی آنها، باید احترام قائل شد.

### // مفاهیم اصلی

اشتباههای متداول در طرح تحقیق  
حق موافقت آگاهانه  
قانون مقررات دولتی  
ملاحظات اخلاقی در تحقیق

اجزای طرح پیشنهادی تحقیق  
حفاظت از آزمودنی در مقابل صدمات  
شورای بازنگري سازمانی  
معافیت‌های مقوله‌ای  
ملاحظات قانونی در تحقیق

## /// تمرینها

- ۱- اجزای اولیه یک طرح پیشنهادی تحقیق کدامند؟
- ۲- چرا نگارش با دقت و هوشمندانه مقدمه بسیار حائز اهمیت است؟
- ۳- بیان روشن مسئله باید در کدام قسمت از طرح پیشنهادی تحقیق صورت پذیرد؟
- ۴- فرضیه های زیر را دوباره نویسی کنید، برای تمام متغیرها، تعریف کاربردی ارائه کنید:  
الف - کودکانی که خواندن را با روش  $i/i/a$  یاد می گیرند، بهتر از آنهایی که با روش سنتی یاد می گیرند، عمل خواندن را انجام می دهند.
- ب - دانش آموزان دبیرستانی که در آزمون استعداد مکانیکی XYZ در چارک بالا قرار می گیرند، مکانیکهای بهتری خواهند شد.
- ج - نمرات آزمون فرعی ریاضی، مربوط به آزمون موفقیت SRA دانش آموزان تیزهوش کلاس هفتمی که با روش جدید تدریس ریاضی برای مدت یک سال آموزش دیده اند، بالاتر از نمرات دانش آموزان تیزهوش کلاس هفتمی است که با روش سنتی آموزش دیده اند.
- ۵- برخی متغیرهای مداخله کننده را که می توانند بر تفاوت میانگین نمرات پیشرفت تحصیلی کلاسهای مختلف در یک سطح اثر گذارند، نام ببرید. شما در شیوه های پیشنهادیتان چگونه این متغیرها را کنترل می کنید؟
- ۶- آماره مناسب برای اندازه گیری همبستگی، در صورتی که مقیاس متغیر وابسته ما اسمی باشد، چیست؟
- ۷- کدام اندازه از گرایش مرکزی برای داده های فاصله ای مناسب است؟
- ۸- برای چه نوع داده هایی، تحلیل واریانس مناسب است؟
- ۹- کار شورای بازنگری سازمانی چیست؟
- ۱۰- چگونه ممکن است اطلاع آزمودنیها از مشارکت داشتن، نتایج تحقیق را تحت تأثیر قرار دهد؟ در اینگونه موارد، چگونه باید با ضرورت کسب موافقت آگاهانه کنار آمد؟
- ۱۱- در پروژه های تحقیقاتی، چه تدابیری باید برای تضمین محرمانه باقی ماندن پاسخها و مسائل خصوصی آزمودنیها، اندیشید؟
- ۱۲- برخی اشتباهات متداول را که باید در تهیه یک طرح پیشنهادی تحقیق از آنها اجتناب کرد، نام ببرید.

### /// پاسخها

- ۱- مقدمه، روش شناسی، تحلیل داده ها، اهمیت بررسی، بودجه زمانی و هزینه طرح.
- ۲- در صورتی که مقدمه خوب تهیه نشده باشد، خواننده تمایلی به خواندن بقیه بخشهای طرح پیشنهادی نخواهد داشت.
- ۳- در اولین پاراگراف مقدمه.
- ۴- الف - کلاس سومی هایی که در کلاسهای اول و دوم خواندن را با شیوه  $i/u/a$  آموخته‌اند، نسبت به کلاس سومی هایی که خواندن را با شیوه خواندن پایه آموخته‌اند، در آزمون خواندن کالیفرنیا، نمرات بهتری می‌گیرند.
- ب - دانش آموزان کلاس یازدهم و دوازدهمی که قبل از کارآموزی در آزمون استعداد مکانیکی XYZ نمراتشان در بالای چارک اول قرار دارد، پس از یک سال دوره کارآموزی مکانیک، توسط سرپرستان خود، بسیار بالاتر از آنهایی که نمره‌شان زیر چارک اول بوده است، رتبه بندی می‌شوند.
- ج - نمرات آزمون فوعی ریاضی مربوط به آزمون موفقیت SRA کلاس هفتمی هایی که نمره هوشبهر آنها بر مبنای WISC بالاتر از ۱۱۵ است و برای مدت یک سال با شیوه جدید تدریس ریاضی آموزش دیده‌اند، بالاتر از دانش آموزان مشابهی خواهد بود که با شیوه سنتی آموزش دیده‌اند.
- ۵- تفاوت در متوسط سطح توانایی، محیطهای فیزیکی کلاس، معلمان و انواع و میزان مطالب، از جمله این عوامل هستند. از طریق انجام پیش‌آزمون برای سطح توانایی، انتخاب کلاسهایی که محیطشان یکسان است، آموزش معلمان تا سطح خاصی از کفایت، می‌توان این متغیرها را کنترل کرد. نوع متفاوتی از کنترل، استفاده از تعداد بسیار زیاد کلاس است که به صورت تصادفی به شرایط مختلف تخصیص داده می‌شوند.
- ۶- همبستگی دو رشته‌ای نقطه‌ای.
- ۷- میانگین.
- ۸- متغیر مستقل اسمی و متغیر وابسته فاصله‌ای.
- ۹- شورای بازنگری سازمانی در مورد اینکه آیا آزمودنیها در یک آزمایش پیشنهادی، بر طبق رهنمودهای DHEW در معرض خطر هستند یا خیر، تصمیم‌گیری می‌کند و سپس تعیین

می‌کند که آیا منافع انجام تحقیق احتمالاً بر خطر آن ارجحیت دارد یا خیر.

۱۰- اطلاع از اینکه از تیمار آزمایشی یا غیر معمولی استفاده می‌شود، می‌تواند بر وضعیت روانی و یا انتظارات آزمودنی اثر بگذارد، و این تأثیر گذاری می‌تواند اثر تیمار عملی را کاهش دهد و یا به طریقی بر آن اثر گذارد. در اینگونه شرایط باید به آزمودنیها گفته شود که پس از پایان بررسی، از اهداف آن آگاهی خواهند یافت.

۱۱- آزمودنیها نباید مجبور باشند که خودشان را معرفی کنند، مگر آنکه این امر ضروری باشد و در گزارشهای عمومی نیز نباید شناسایی شوند، مگر آنکه موافقت خود را اعلام کرده باشند.

۱۲- مسئله پیش پا افتاده است.

مسئله محدود نشده است.

اهداف، فرضیه ها و سوالات بسیار گسترده بیان شده‌اند.

جزئیات روشها قید نشده است.

طرحی ساده برای مسئله‌ای پیچیده پیشنهاد شده است.

متغیرهای مربوط مد نظر قرار نگرفته‌اند.

## فصل پانزدهم

### تحلیل، تفسیر و گزارش نتایج

#### هدفهای آموزشی

دانشجو بعد از مطالعه این فصل، قادر خواهد بود:

- ۱- نقش کامپیوتر را در تحقیق توصیف کند.
- ۲- اصولی را که محقق باید به هنگام تفسیر نتایج مورد انتظار در ذهن داشته باشد، ذکر کند.
- ۳- تفاوت بین معنی داری آماری و معنی داری عملی را تشخیص دهد.
- ۴- اصولی را که محقق باید در تفسیر غیر قابل انتظار، در نظر داشته باشد، ذکر کند.
- ۵- موارد مختلفی را که محقق باید در صورت قبول فرضیه صفر در تفسیر نتایج مدنظر داشته باشد، فهرست کند.
- ۷- عناوین مناسبی برای یک پایان نامه و یا گزارش تحقیق بنویسد.
- ۸- ماهیت محتوایی را که باید در هر بخش از یک رساله یا پایان نامه آورده شود، توصیف کند.
- ۹- مراحل کاری را که محقق به هنگام تهیه یک مقاله تحقیقی برای یک مجله تخصصی طی می‌کند، بیان کند.
- ۱۰- مراحل کاری را که محقق جهت تهیه یک مقاله تحقیقی برای خواندن در یک کنفرانس تخصصی طی می‌کند، بیان کند.

وقتی داده های تحقیق گردآوری شدند، محقق به پردازش داده ها، تحلیل نتایج و تفسیر دقیق یافته ها می پردازد و نهایتاً گزارش بررسی را می نویسد. بحثی مختصر از هر یک از این مراحل کار، در این فصل ارائه می شود.

### /// پردازش داده ها

اولین قدم در پردازش داده های گردآوری شده، رجوع به طرح پیشنهادی تحقیق به منظور واریسی برنامه های اصلی در مورد ارائه داده ها و انجام تحلیل آماری است. درباره این که چه سؤالاتی قرار است پاسخ داده شوند، متغیرهای بررسی چگونه به صورت کاربردی تعریف می شوند و چه طرحهای تحقیقی و شیوه های آماری مورد استفاده قرار خواهند گرفت، از قبل تصمیم گیری شده است. حال آنچه باقی مانده، سازماندهی داده های خام، انجام محاسبات ضروری و تفسیر نتایج است.

امروزه پروژه های تحقیقاتی در زمینه تعلیم و تربیت و دیگر علوم رفتاری، اغلب در زمره تحقیقات پیچیده محسوب می شوند و نیاز به بکارگیری کامپیوتر دارند. غالب محققان، به منظور دستیابی به دقت و صرفه جویی در وقت و انرژی، از امکانات ماشینهای محاسباتی الکترونیکی برای تحلیل داده هایشان، سود می جویند. در نتیجه بجاست که در اینجا مطلب مختصری در مورد کامپیوترها و نحوه استفاده از آنها در پردازش داده ها، بیاوریم.

### // استفاده از کامپیوترها برای پردازش داده ها

تکامل کامپیوتر، گستره تحقیق را تغییر داده و اجرای تحقیقاتی را میسر ساخته است که در غیر این صورت، کسی حتی تصور انجامش را نیز نمی توانست داشته باشد. قبل از عصر کامپیوتر، محققان از انجام بررسیهایی که در آنها چندین متغیر وجود داشت و تعداد نمونه هم زیاد بود، به علت زمان طولانی و نیروی کاری که برای جدول بندی و تحلیل داده ها ضرورت می یافت، اجتناب می کردند. معمولاً آزمونهای آماری پیشرفته و تحلیلهای پیچیده، انجام نمی شد. امروزه کامپیوتر می تواند مقادیر بسیار زیاد داده ها را پردازش کند و تحلیلهای آماری پیچیده را با سرعت و کفایت شگفت انگیزی انجام دهد. بنابر این، محققان می توانند بدون آنکه به تعداد متغیرها و پیچیدگی تحلیلهای توجه داشته باشند، تحلیل خود را صرفاً بر مبنای مناسب بودن آن برای پاسخگویی به سؤالات تحقیق، انتخاب کنند.



محققان ممکن است از کامپیوتر بزرگ<sup>(۱)</sup> استفاده کنند و یا از کامپیوترهای شخصی کوچکتر. کامپیوتر Main Frame بزرگ و گران قیمت است. در غالب دانشگاهها و سازمانها کامپیوترهای Main Frame اعمالی نظیر پرداخت حقوق، حسابداری و کارهای دیگر را علاوه بر ارائه خدمات در پاسخ به نیازهای محققان انجام می دهند. کامپیوتر Main Frame از مزیتهایی نظیر سرعت، ظرفیت و انعطاف پذیری برخوردار است. نقطه ضعف اساسی آن این است که محقق برای دسترسی به این کامپیوتر باید با افراد و ادارات مختلف رقابت کند و این امر می تواند به تأخیر و ناراحتی منجر شود. کامپیوترهای شخصی، کوچک و تقریباً ارزان قیمت هستند و بسیاری از محققان استفاده از آنها را آسانتر و مناسبتر می یابند. برنامه های کامپیوتری موجود برای کامپیوترهای بزرگ بیشتر است، لکن برنامه های کامپیوترهای شخصی نیز برای تحلیلهایی که اکثر محققان آموزشی به آنها نیاز دارند، در دسترس هستند و با قوی تر شدن کامپیوترهای شخصی و همه گیر شدن استفاده از آنها، این برنامه ها هم، بیشتر و بیشتر در دسترس قرار می گیرند. یک برنامه کامپیوتری در واقع دستورالعملهای قدم به قدم، به زبان کامپیوتر است که به وسیله آن به کامپیوتر اطلاع داده می شود که از چه داده هایی استفاده کند و چه عملیاتی را بر روی داده ها صورت دهد. کسانی که با زبانهای کامپیوتری و روشهای برنامه نویسی آشنایی دارند، می توانند برنامه ها را خودشان بنویسند.

با این همه، خوشبختانه تعداد زیادی برنامه های آماده موجود است. این برنامه ها به نام بسته نرم افزاری خوانده می شوند و توصیه می شود کسانی که بی تجربه یا کم تجربه هستند، از این برنامه ها استفاده کنند. برای کامپیوترهای بزرگ هم تعدادی بسته نرم افزار آماری وجود دارد. یکی از شناخته شده ترین این بسته نرم افزارها مجموعه زیست پزشکی (BMDP) است، که اساساً برای استفاده در زمینه های زیست شناسی و پزشکی تهیه شده است.<sup>(۲)</sup> تعداد زیادی برنامه های BMDP جهت تحلیل های آماری مختلف در دسترس است. مجموعه دیگری که بیش از مجموعه های دیگر مورد توجه محققان آموزشی است، بسته نرم افزاری برای علوم اجتماعی (SPSS-X) است.<sup>(۳)</sup> SPSS-X حاوی بسیاری از متداولترین شیوه های آماری است که توسط دست اندرکاران علوم اجتماعی، به کار برده می شوند. این مجموعه زبانی ساده دارد که یادگیری آن آسان است و نیازی به اطلاع از شیوه عمل کامپیوترها، ندارد. محقق در مورد چگونگی انجام

1- Main Frame.

2- *Biomedical Computer Programs* (1983). Berkeley: University of California Press.

3- *SPSS-X User's Guide*, 3d ed. (1988). Chicago: SPSS Inc.

تحلیل داده ها تصمیم گیری می کند و سپس با بررسی راهنمای SPSS-X برنامه مناسب را می یابد. مجموعه دیگر بسته نرم افزاری، سیستم تحلیل آماری (SAS) است.<sup>(۱)</sup> این مجموعه نسبت به SPSS-X از سطح پیچیدگی بیشتری برخوردار است. نسخه ای از SPSS-X نیز برای استفاده کامپیوترهای شخصی موجود است (MYSTAT)<sup>(۲)</sup>، که در واقع نسخه دیگری از نرم افزار SYSTAT می باشد که بسته نرم افزاری مفید دیگری است. کتاب گریبگ به نام «نحوه استفاده از کامپیوترهای کوچک و کامپیوترهای بزرگ برای تحلیل داده ها در علوم اجتماعی»<sup>(۳)</sup>، توصیفی مفید از کامپیوترهای کوچک و برنامه های آماری کامپیوترهای بزرگ و نحوه استفاده از آنها را ارائه می دهد.

### / توصیه ها

در کار با کامپیوتر رعایت احتیاط ضروری است. نتایج به دست آمده تا آنجا صحیح است که اطلاعات داده شده به کامپیوتر صحیح باشد. کامپیوتر مرتکب اشتباه محاسباتی نمی شود، اما در صورتی که داده های نادرست یا خطاهایی منطقی وارد کامپیوتر شود، کامپیوتر کورکورانه، ولی با کفایت، به جای نتیجه، مطالب بیهوده بیرون می دهد. دو نوع اشتباه می تواند در برنامه نویسی پیش بیاید. اشتباهات زبانی، نظیر هجی کردن نادرست دستورالعملها، که اینگونه اشتباهات را کامپیوتر تشخیص می دهد و در اینگونه موارد، دستورالعمل را قبول نمی کند، و اشتباهات در منطق برنامه نویسی نظیر دستورالعملهای غلط، که کامپیوتر اینگونه اشتباهات را قبول می کند و سپس نتایج بی معنای بسیار گران قیمتی را ارائه می دهد.

نهایتاً، مایلیم توصیه های زیر را در مورد استفاده از کامپیوتر ارائه کنیم:

- ۱- برای پردازش داده هایتان، از کامپیوتر به طور کامل استفاده کنید، ولی محصول کامپیوتر را منتقدانه ارزیابی کنید. هرچند کامپیوترها اشتباه نمی کنند، ولی هرگز اشتباهاتی را هم که در آماده سازی داده ها برای پردازش و یا برنامه نویسی رخ می دهد، تصحیح نمی کنند.
- ۲- از هر برنامه ای که در راهنمای برنامه ها پیدا می کنید، به صورت کورکورانه استفاده نکنید. شما باید از قبل بدانید کدام شیوه آماری برای یک مجموعه معین از داده ها، مناسبترین است.

1- *SAS User's Guide* (1983). Raleigh, NC: SAS Institute.

2- Wilkenson, L. (1986). *SYSTAT: The System for Statistics*. Evanston, IL: SYSTAT, Inc.

3- Greenberg, B. (1987). *Using Microcomputers and Mainframes for Data Analysis in the Social Sciences*. Columbus, OH: Merrill.

### /// اصول تفسیر

وقتی داده های تحقیق گردآوری شدند و تحلیل آماری انجام شد، محقق می تواند به چالش در تفسیر نتایج بپردازد. افزودن به آگاهیها، کانون اصلی کوششهای تحقیقاتی بوده است. وقتی به مرحله تفسیر رسیدیم، محقق نشان می دهد که در این پروژه چه چیزی فراگرفته شده است و این آموخته چگونه با پیکره کلی آگاهیها در آن حیطه همخوانی دارد.

### // نقش طرح پیشنهادی در آسان سازی تفسیر

بنیان مناسب برای تفسیر نتایج یک تحقیق باید به طور نظام مند در هر مرحله از پیشرفت طرح پیشنهادی پی ریخته شود؛ حتی قبل از آن که تحقیق عملاً شروع شود. محقق، در حالی که به خاطر دارد داده های تحقیق شامل چه چیزی و گویای چه مطلبی خواهد بود، خود را برای تفسیر داده ها و انطباق آنها با پیکره آگاهیهای حیطه مورد نظر، آماده می سازد.

از برنامه ای که به طور دقیق درباره آن فکر شده و به صورت یک طرح تحقیق پیشنهادی تمام و کمال بیان شده باشد، می توان نتایجی را انتظار داشت که به آسانی و به صورت معنی داری قابل تفسیر هستند. اگر محقق، بررسی خود را به گونه ای پی ریخته باشد که پیامدهای فرضیه های او در قالب مشاهدات پایا بیان شود، پس تفسیر و ارزش مشاهدات باید روشن باشند.

### // اهمیت وفادار ماندن به برنامه اصلی

وقتی طرح تحقیق مورد قبول واقع شد و پروژه به اجرا درآمد، بررسی باید دقیقاً همانطور که برنامه ریزی شده بود، اجرا شود. این قاعده هم دلیل اخلاقی و هم دلیل عملی دارد.

به منظور نشان دادن دلایل اخلاقی، فرض کنیم که آقای ویلیامز یک معلم زبان خارجه، زمان و کوشش زیادی را صرف تکمیل یک روش تدریس زبان فرانسه کرده است که تصور می کند به طور قابل ملاحظه ای بر روشهای موجود ارجحیت دارد. او برای آزمون میزان تأثیر گذاری این روش، یک گروه آزمایشی را که با این روش، و یک گروه گواه را که با روشی دیگر آموزش دیده بودند، انتخاب کرد. او هر هفته آزمونهای پیشرفت در درس فرانسه را انجام داد و نتایج را به عنوان متغیر وابسته جمع آوری کرد. فرض کنید که در چند هفته اول به این نتیجه رسید که میانگین نمرات آزمون دو گروه تقریباً یکسان است. او از آنجا که سرمایه گذاری بزرگی بر روی روش خود

انجام داده است، مشکل می‌توانست قبول کند که این روش بهتر از روش دیگر نیست، بنابر این تصمیم می‌گیرد در هر دو کلاس حضور پیدا کند تا متوجه اشتباهات شود. وی سپس کشف می‌کند که گروه آزمایشی، آگاهی و درک بیشتری از زندگی و فرهنگ فرانسوی را نشان می‌دهند. از آنجا که او در مورد یافتن تفاوت بین نمرات مصر است، تصمیم می‌گیرد که متغیر وابسته را عوض کند و نمرات آزمونهایی در مورد زندگی و فرهنگ فرانسوی را به عنوان متغیر وابسته انتخاب کند.

چنین تغییری غیر اخلاقی است. در مورد دو گروه تصادفی چنانچه زمانی طولانی مورد بررسی قرار گیرند، بالاخره می‌توان یک نوع ارجحیت در یک گروه نسبت به گروه دیگر یافت. چنانچه گروه آزمایشی در مورد آشنایی با فرهنگ و زندگی فرانسوی ارجح به نظر نمی‌رسیدند، ممکن بود در زمینه مهارت کلامی، مهارت در گوش کردن، رفتار و سلوک و یا متغیر دیگری ارجحیت داشته باشند. معلم زبان باید آزمایش را همان طور که برنامه‌ریزی شده است، دنبال کند و متغیر وابسته‌اش را تغییر ندهد. او باید در گزارش خود قید کند که گروه آزمایشی در زمینه درک زندگی و فرهنگ فرانسوی به نظر ارجح می‌آمد، ولی باید برای خواننده روشن شود که این رابطه‌ای نیست که بر روی آن فرضیه ساخته شده باشد و به راحتی می‌تواند تابعی از احتمال باشد. محقق از نظر اخلاقی نمی‌تواند متغیر مستقل و یا وابسته‌ای را که به نظر می‌رسد «عمل نمی‌کند»، حذف و یا متغیرهای جدید نویدبخشی را اضافه کند. چنین تغییراتی را باید به بررسیهای آینده واگذار کرد.

افزودن متغیرهای جدید، از نظر علمی نیز عاقلانه نیست. چنین روشی نتایج یک بررسی را مخدوش می‌کند و معنای نتایج آن را مبهم می‌سازد. محققان معمولاً برای افزودن متغیرهای جدید و جالب که بررسی آنها را ثمربخش‌تر می‌کند، وسوسه می‌شوند. به هر حال مبنای نظری برای تفسیر این متغیرها از پیش ساخته نشده و بهترین توصیه این است که آنها را برای بررسی بعدی باقی بگذاریم.

### // تفسیر نتایج قابل انتظار

قابل فهم است که محققان از اینکه نتایج بررسی با چهارچوب ساخته شده قبلی همخوانی دارد و تفسیر می‌تواند به همان صورتی که انتظار می‌رفته است، انجام گیرد، احساس رضایت

کنند. می‌توان گفت که بررسی خوب عمل کرده است و بین دلایل منطقی و نتایج مطابقت وجود دارد.

تنها چند کلمه به عنوان هشدار در چنین حالتی ضروری است:

۱- فراتر از اطلاعات، تفسیر نکنید. این نکته ممکن است به نظر حکمی از قبل مسلم بیاید، ولی محققان اغلب زمانی که نتایج به همان صورت که انتظار داشتند، حاصل می‌شود، آنچنان هیجان زده می‌شوند و نتیجه‌گیری‌هایی می‌کنند که مبنای معتبری ندارد. حتی در تحقیقات منتشر شده نیز به فراوانی می‌توان بیشتر تفسیر یافت تا داده‌های مستند.

۲- محدودیت‌های بررسی را فراموش نکنید، البته این محدودیت‌ها باید از قبل در بررسی تعریف شده باشد. محدودیت‌های ذاتی در وسایلی که از نظر پایایی کامل نیستند، محدودیت‌های ناشی از مشکلات نمونه‌گیری، مسائل اعتبار درونی و غیره.

۳- اخلاقیات حکم می‌کند که فرد مشکلات اعتبار درونی را که به نتایج مربوط می‌شود، گزارش کند. چنانچه به رغم تمام کوشش‌های محقق، متغیرهای غیر آزمایشی، برای گروه آزمایشی، بویژه مساعد و آنهایی که به گروه گواه مربوط می‌شوند، نامساعد باشند، چنین موقعیت‌هایی باید گزارش شوند و در تفسیر نتایج مورد نظر قرار گیرند (برای مثال، به رغم تخصیص تصادفی معلمان به گروه‌ها، ممکن است اکثر معلمان با تجربه در گروه آزمایشی و معلمان بی‌تجربه در گروه گواه، قرار گیرند).

۴- به خاطر داشته باشید که معنی داری آماری بدین معنی است که با درجه آزادی مناسب، نتایج، تابعی از احتمال و اتفاق نباشند. معنی داری آماری<sup>(۱)</sup> بدین معنی نیست که نتایج بر حسب معنای عموماً پذیرفته شده معنی داری آماری، یعنی مهم، با معنا و یا خطیر هستند. گمان نکنید که معنی داری آماری با اهمیت بودن یافته‌های شما را تضمین می‌کند. فرض کنید دو گروه هم‌ارز طی یک دوره دوساله در معرض دو روش یادگیری هجی کردن قرار گرفته‌اند. دانش آموزانی که از روش A استفاده کردند، در طول آزمایش، میانگین پیشرفتی معادل ۲/۱۵ سال رشد را در آزمون‌های استاندارد شده به دست آوردند، در حالیکه دانش آموزان روش B پیشرفتی معادل ۲/۲۰ را در طول همان دوره نشان دادند. چنانچه گروه‌ها بزرگ باشند و / یا اگر تفاوتها در داخل گروه‌ها کم باشند، تفاوت بین میانگین‌ها از نظر آماری معنی دار می‌شود، اما تفاوت نیمه یک ماه

طی یک دوره دوساله، از نظر عملی، نسبتاً بی‌معنا است. اگر روش B، از نظر مدت زمان و مواد مصرفی گرانتر باشد، عاقلانه نیست که معلم، به صرف این که این روش، از نظر آماری پیشرفت بیشتری نسبت به روش A حاصل کرده است، آن را انتخاب کنند. از سوی دیگر، اگر روش B ارزانتر باشد، معلم به سمت آن تمایل پیدا می‌کند، زیرا نتایج آن از نظر عملی بسیار شبیه روش A است.

اهمیت بالقوه معنی دار بودن نتایج باید قبل از شروع بررسی، در طرح پیشنهادی مطرح شود. مطالعه‌ای که اطلاعات با معنایی را برای افزودن به پیکره دانش موجود فراهم نیاورد، از اهمیت برخوردار نیست، فارغ از این که نتایج آن از نظر آماری چقدر معنی دار هستند.

### // تفسیر نتایج منفی

محققانی که به نتایجی مخالف آنچه فرض کرده‌اند، دست می‌یابند، معمولاً در مورد کمبودهای بررسی خود افساگری می‌کنند و تفسیر نتایجشان همانند یک اعتراف می‌شود: وسایل اندازه‌گیری برای سنجش متغیرهای مورد نظر ناکافی بودند، نمونه بیش از حد کوچک و غیر معرف بود و در نتیجه نمی‌توان یافته‌ها را به طور معتبر به جامعه مورد نظر تعمیم داد و غیره. این بازاندیشی، مشکلات اعتبار درونی را آشکار می‌سازد و علت این امر را که چرا نتایج بررسی مطابق انتظار نبوده است، توضیح می‌دهد.

البته هریک از موارد فوق می‌توانند صحیح باشند و کمبودهای بررسی باید بدون در نظر گرفتن نتایج، گزارش شوند. به هر حال تحقیق، سفری جسورانه در ناشناخته‌ها است. بنابراین هیچ «باید بشود» غایی وجود ندارد. محقق نتایج قابل انتظار بررسی‌اش را بر مبنای نظریه، قیاس و نتایج تحقیقات قبلی پیش بینی می‌کند. چنانچه آنها آن قدر جامع باشند که محلی برای تردید در نتایج بررسی باقی نماند، بررسی در وهله اول بی‌فایده به نظر خواهد رسید. وقتی بررسی خود را آغاز می‌کنیم، به طور ضمنی اذعان داریم که نتایج، پیامد یک امر حدسی است و امری قطعی نمی‌باشد. در مرحله تکمیل طرح پیشنهادی، قابل درک است که اعلام کنیم، با بهره‌گیری از بهترین وسایل و روشهای موجود، بی‌غرضانه به دنبال تعیین وضعیت واقعی مسائل هستیم. بنابراین، باید یافته‌هایمان را قبول و بدون در نظر گرفتن وضعیت داده‌ها، آنها را تفسیر کنیم. چنانچه نتایج با منطق نظری بررسی تضاد داشته باشند، بخش مباحثه گزارش

ما، باید شامل تجدید نظر در نظریه اصلی، در پرتو یافته‌های بررسی باشد. محققان معمولاً تمایلی نسبت به ارائه و تفسیر داده‌هایی که با تحقیقات قبلی و نظریه‌های شناخته شده، تضاد دارند، ابراز نمی‌کنند، هرچند احتمال درستی نتایج آنها و نادرستی نتایج قبلی وجود دارد. چنانچه محققان نسبت به گزارش یافته‌هایی که با گزارشهای قبلی تضاد دارند، تمایل نداشته باشند، روند پیشرفت علم تعلیم و تربیت کند خواهد شد. نتایج متضاد نشان می‌دهد که مسئله‌جا نیفتاده است و می‌تواند انگیزه‌ای برای تحقیق بیشتر باشد. تحقیقات اضافی و یا ساخت نظریه می‌تواند در نهایت باعث تطابق یافته‌هایی شود که ظاهراً متضاد هستند. نظریه امری تجربی است و نباید کسی را از «اظهار چیزی به گونه‌ای که هست»، باز دارد.

بازنگری مبنای نظری بررسی به بخش مباحثه تعلق دارد. محقق نباید به عقب برگردد و بخش مرور پژوهشهای پیشین و فرضیات گزارش را دوباره نویسی کند.

### // تفسیر نتایج در صورتی که فرضیه صفر رد نشود

از آنجا که فرضیه صفر به دلایل مختلف ممکن است تأیید شود، تفسیر چنین نتایجی می‌تواند بویژه مشکل باشد. تأیید فرضیه صفر ممکن است اتفاق بیفتد؛ زیرا (۱) فرضیه صفر در حقیقت صحیح است. ممکن است رابطه‌ای بین متغیرها وجود نداشته باشد یا تیمار آزمایشی مؤثرتر از تیمار کنترل نباشد. (۲) فرضیه صفر اشتباه است، ولی مشکلکهای اعتبار درونی چنان تحقیق را آلوده ساخته است که رابطه عملی بین متغیرها قابل مشاهده نیست. (۳) فرضیه صفر اشتباه است، ولی طرح تحقیق توانایی رد آن را ندارد.

هریک از این موارد می‌تواند اتفاق بیفتد، ولی محقق نمی‌داند کدام صادق است و بنابراین نباید ادعایی برای هیچ یک از آنها به عنوان تبیین یافته‌هایش داشته باشد.

ارائه یک فرضیه صفر رد نشده به عنوان شاهدهی برای فقدان رابطه بین متغیرها، اشتباه است. چنانچه فرضیه صفر رد نشود، تنها می‌توانیم بگوییم که هیچ شاهدهی مبنی بر درست بودن یا غلط بودن فرضیه در دست نداریم. یک تبلیغ تجارتي که به طور گسترده‌ای مورد استفاده قرار می‌گیرد، حاکی از آن است که نوع خاصی از خمیر دندان، در کاهش پوسیدگی دندان، بر انواع مشابه آن تفوق دارد. اگر واژه «تفوق نداشتن» را به معنای «فقدان تفاوت معنی دار» تفسیر کنیم، می‌توانیم آزمایشی را تصور کنیم که در آن تعداد کمی آزمودنی شرکت داشته‌اند و یا مسائل اعتبار

درونی متعددی در آن وجود داشته است. اگر پذیرش فرضیه صفر، نتیجه مطلوب یک آزمایش باشد، ترتیب دادن مقدمات برای حصول چنین نتیجه‌ای بسیار آسان است.

البته، چنانچه فردی جامعه آماری کوچکی را مورد بررسی قرار دهد و بتواند آن جامعه را سرشماری کند، یک فرضیه صفر رد نشده می‌تواند به عنوان فقدان رابطه بین متغیرها در آن جامعه خاص تفسیر شود. همچنین اگر با یک نمونه بسیار بزرگ سر و کار داشته باشیم، فرضیه صفری که رد نشود، اعتبار کسب می‌کند. برای مثال، گزارش کلمن، با بیش از ۶۰۰۰۰۰۰ آزمودنی، چنان مبنای بزرگی را فراهم می‌سازد که در این مورد تمایل داریم فقدان رابطه مشاهده شده بین متغیرها را به عنوان شهادی بر فقدان رابطه واقعی بپذیریم. هرچند در غالب بررسیها، فرضیه صفر رد نشده، باید به عنوان فقدان شواهد تفسیر شود و نه بیشتر.

این خطر وجود دارد که محققانی که بیش از حد شیفته فرضیه آزمایشی خود هستند، وسوسه شوند که فرضیه صفر رد نشده را به گونه‌ای تفسیر کنند که اصلاً وجود نداشته است. آنها مشکلات اعتبار درونی را ذکر و ادعا می‌کنند که اگر فقط این مشکلات غیرمنتظره، آزمایش را ضایع نمی‌کردند، نتایج مسلماً معنی دار می‌شدند. مسلم است که فرد باید کلیه مشکلات اعتبار درونی را که در طول بررسی بروز کرده است، گزارش کند، ولی نباید از آنها برای توضیح نتایج ناخوشایند استفاده کند. محقق ممکن است تحقیقات بیشتری را پیشنهاد کند و بخواهد این تحقیقات را به گونه‌ای برنامه ریزی کند که با مشکلات اعتبار درونی، که در بررسی خود با آن مواجه شده است، روبرو نشوند، ولی هنوز هم فرضیه صفر رد نشده را باید به عنوان شواهد گزارش کند و نه بیشتر.

توان یک آزمایش به توانایی آماری آن در رد فرضیه صفر، در صورتی که واقعاً اشتباه باشد، اطلاق می‌شود. این توان تابعی از حجم نمونه، ناهمگنی آزمودنیها با توجه به متغیر وابسته و اعتبار وسایل اندازه گیری مورد استفاده و ماهیت آماره مورد استفاده جهت آزمون فرضیه است. چنین عواملی را باید در برنامه ریزی یک آزمایش مد نظر داشت. بسیاری از کتابهای درسی آمار توضیح داده‌اند که چگونه باید آزمایشها را برنامه ریزی کرد تا رابطه های بامعنا، به صورت مشاهداتی که از نظر آماری معنی دار هستند، بروز کنند.

توان یک آزمایش باید در برنامه ریزی بررسی مورد نظر باشد. نباید آن را در انتهای بررسی جهت توصیف فقدان معنی داری آماری آورد. برای مثال، محقق نباید بگوید: «چنانچه حجم



نمونه بزرگتر بود، نتایج از نظر آماری معنی دار می‌شدند.»  
به استثنای موارد نادر، تنها تفسیر قابل قبول در مورد فرضیه رد نشده، این است که شواهدی برای نتیجه‌گیری مشاهده نشده است.

### // تفسیر رابطه‌های از پیش فرض نشده

قبلاً تأکید کردیم که محقق مجاز نیست به منظور طی مسیرهای نوید بخش تری که در طول انجام بررسی خود را آشکار خواهند کرد، فرضیه‌ای را نادیده بگیرد. این بدان معنی نیست که رابطه‌های فرض نشده‌ای که ممکن است در اجرای بررسی مشاهده شوند، باید نادیده گرفته شوند. برعکس، آنها باید با همان دقتی ثبت و تحلیل شوند که در مورد روابط فرض شده، اعمال می‌شود. در طول تاریخ علوم، ثابت شده است که اکتشافات غیر مترقبه اغلب پر ارزش و با اهمیت هستند.

با این همه، به چنین یافته‌هایی همیشه باید با تردید بیشتری نگاه کرد، تا یافته‌هایی که مستقیماً با فرضیه مرتبط هستند؛ زیرا احتمال بروز یک رابطه فرض نشده نادرست تقریباً زیاد است. چنین روابطی باید گزارش شوند، ولی باید به آنها به عنوان ضمیمه‌ای بر جریان اصلی بررسی نگاه کرد. آنها قبل از آنکه به عنوان مبنای نتیجه‌گیری به کار روند، باید خود به عنوان موضوع اصلی یک تحقیق در نظر گرفته شوند.

### /// گزارش تحقیق

چنانچه نتوان بین نتایج یک تحقیق با تحقیقات دیگر ارتباط برقرار کرد، آن کار تحقیقاتی از ارزش کمی برخوردار خواهد بود. بنابراین، آگاهی از شیوه‌های نگارش یک گزارش تحقیقاتی برای کلیه محققان اهمیت دارد. هدف از این بخش، ارائه یک رهنمود کلی برای سازماندهی و ارائه یک گزارش است. به منظور آشنایی با قواعد خاص سبک و شکل نگارش گزارشهای تحقیقاتی می‌توان به یک کتاب راهنما مراجعه کرد. فهرست چند کتاب راهنما در پایان این فصل آمده است.

در یک گزارش تحقیقاتی، محقق هم فرآیند و هم یافته‌های تحقیق خود را ارائه می‌کند. او همچنین در مورد کاربرد یافته‌ها و رابطه آن با دیگر دانسته‌های آن حیطه، بحث می‌کند.  
از آنجا که گزارش، توسط افراد حرفه‌ای بسیار پر مشغله خوانده خواهد شد، باید تا آنجا که

ممکن است موجز و به صورت منطقی سازمان یافته باشد. داستان‌سرای، روایات از تجربیات شخصی و استدلال‌های بحث انگیز، جایی در گزارش تحقیقاتی ندارند. این بدان معنی نیست که گزارش باید کسل کننده و فضل فروشانه باشد. چنانچه محقق با روحی پرشور با بررسی خود روبرو شده باشد، این روح احتمالاً در بین خطوط گزارش ظاهر خواهد شد.

از آنجا که هدف گزارش، ارائه یک تحقیق، و نه شخصیت نویسنده است، گزارش از نظر دستور زبان، باید به صورت سوم شخص نوشته شود. در این مورد، ضمائر اول شخص هیچگاه به کار برده نمی‌شوند. بنابر این، محقق نمی‌نویسد: «من آزمودنیها را به صورت تصادفی به دو گروه تیمار تخصیص دادم»، بلکه می‌نویسد: «آزمودنیها به صورت تصادفی به دو گروه تیمار تخصیص یافتند». به رغم شوق طبیعی محقق در مورد اهمیت کار خود، نگارنده نباید در مورد آن لاف بزند، بلکه باید ارزشیابی آن را به خوانندگان خود و آیندگان واگذار کند.

یک روش رسمی و یکسان برای ارائه گزارشهای تحقیقاتی، شکل گرفته است. اگرچه در نگاه اول، این روش ممکن است به نظر بازدارنده بیاید، ولی در عمل برای مقصود ما بسیار مفید است. بسیار اهمیت دارد که گزارشها را به گونه‌ای ترتیب بدهیم که خواننده دقیقاً بداند در کدام بخش گزارش، دنبال چه بگردد. در فصل ۳ توصیه کردیم که در گردآوری تحقیقات مربوط، ابتدا خلاصه یک مقاله را از اول تا آخر بخوانید. چنانچه مقاله‌ای بخش خلاصه نداشته باشد، شما ناگزیرید که وقتی اضافه صرف کنید و تمام آن را مطالعه کنید تا بفهمید که با موضوع بررسی شما مرتبط است یا نه.

وجود یک روش رسمی گزارش نویسی، محقق را از ابداع آن بی‌نیاز می‌کند. بعد از مطالعه این بخش، خواهید دید که روش رسمی گزارش نویسی، به طور منطقی، مراحل یک پروژه تحقیقاتی را که در فصول پیش ارائه شد، دنبال می‌کند.

یک گزارش تحقیقاتی ممکن است به صورت (۱) پایان نامه و یا رساله، (۲) مقاله برای درج در یک نشریه و (۳) مقاله برای سمینار، ارائه شود. برای هر یک از اینها رویکردی متفاوت مورد نیاز است.

### // پایان نامه یا رساله

غالب دانشگاهها دارای یک کتاب راهنمای نگارش هستند که مفصلاً شکل ترجیحی گزارشهای آن دانشگاه را توصیف می‌کنند. برای آن دسته از دانشجویانی که آزادی انتخاب دارند،

فهرستی از چند کتاب راهنمای نگارش در انتهای همین فصل آورده شده است. چنانچه یک کتاب راهنما انتخاب شود، باید تا انتهای گزارش از دستورالعملهای آن پیروی کرد. رئوس مطالب، فهرست مراحل و ترکیب عمومی که در غالب کتابهای راهنمای نگارش توصیف شده است، در ذیل می آید:

۱- صفحات اولیه

الف - صفحه عنوان

ب - صفحه پذیرش پایان نامه

ج - صفحه سپاسگزاری یا پیشگفتار

د - فهرست محتوا

ه - فهرست جداول

و - فهرست نمودارها

۲- متن

الف - مقدمه

۱- بیان مسئله و توجیه بررسی

۲- بیان هدفها

۳- تعریف واژه های استفاده شده

۴- مرور تحقیقات مرتبط با عنوان بررسی

ب - روشها و نتایج

۱- آزمودنیها

۲- شیوه های مورد استفاده

۳- ابزار گردآوری داده ها

۴- ارائه و تحلیل داده ها

ج - بحث در مورد نتایج

۱- تفسیر یافته ها

۲- معانی ضمنی

۳- کاربردها

د - نتیجه گیریها و خلاصه

۱- نتیجه گیریها

۲- خلاصه

۳- صفحات متمم

الف - کتابنامه

ب - پیوستها (ضمایم)

ج - شناسنامه کاری مؤلف یا مؤلفان پایان نامه (اگر ضروری باشد)

د - چکیده

### ۱ / صفحات اولیه

تنظیم صفحات اولیه بیشتر بر مبنای پیروی از قواعد راهنمای نگارش است. ولی، یک جنبه از این صفحات که نیاز به توضیح بیشتری دارد، خود عنوان است.

عنوان باید ماهیت خاص بررسی را تا آنجا که ممکن است به صورت خلاصه توصیف کند. فرض کنید: (الف) در یک بررسی از کودکان محروم استفاده شده است و آمادگی خواندن در بین آنهایی که در برنامه طرح آموزش پیش دبستانی شرکت کرده‌اند، با آن دسته که هیچگونه تجربه رسمی پیش دبستانی نداشته‌اند، مقایسه شده است و (ب) عنوان «مقایسه نمرات آزمون آمادگی خواندن کودکان محروم که در کلاسهای دوره شش هفته‌ای آموزش پیش دبستانی شرکت کرده‌اند، با کودکان هم‌تای بدون هیچگونه تجربه پیش دبستانی»، با وجود آن که بررسی مورد نظر را خوب توصیف می‌کند، ولی بیش از حد طولانی است. عباراتی چون «یک مقایسه»، «یک بررسی» یا «تحقیقی در» زائد هستند. بعلاوه، غالب خوانندگان آینده نگر می‌دانند که طرح آموزش پیش دبستانی یک تجربه قبل از دبستان است که برای کودکان محروم برنامه ریزی شده است. هرچند، از سوی دیگر تهیه عنوانی که بیش از حد خلاصه، یا در رساندن ماهیت بررسی مبهم باشد، اشتباه جدی‌تری است. با عناوین بیش از حد خلاصه و یا مبهم، فرد باید مقاله را بررسی کند تا بفهمد در مورد چیست. عناوینی چون: «آموزش پیش دبستانی و آمادگی» و یا «خواندن در بین محرومین»، این کاستی را نشان می‌دهد. عنوان باید متغیرهای عمده و جامعه‌های آماری مورد نظر را مشخص کند. ضروری نیست تعاریف عملی متغیرهای عمده و توصیف نمونه‌ها در عنوان آورده شود.

از آنجا که انتخاب عنوان صحیح، فهرست بندی صحیح را تضمین می‌کند، یک راهبرد مفید برای محقق این است که ابتدا تصمیم بگیرد که بررسی خود را تحت چه لغات کلیدی فهرست بندی کند، و از آنجا برای تهیه یک عنوان مختصر قدم بردارد. در مورد بررسی ذکر شده، لغات کلیدی مهم برای فهرست بندی عبارت خواهند بود از: آمادگی خواندن و پروژه آموزش پیش دبستانی. بنابراین یک عنوان مناسب می‌تواند «آمادگی خواندن در بین کودکان تحت پروژه آموزش پیش دبستانی و دیگر کودکان» باشد. چنین عنوانی تقریباً خلاصه است، ولی نشانه معتبری از موضوع بررسی را به خواننده صاحب نظر می‌دهد.

از عناوینی که بازتاب عواطف هستند، مانند «ما باید برنامه های آموزش پیش دبستانی را توسعه دهیم» و یا «نگذاریم محرومین، در خواندن ضعف داشته باشند»، باید به هر قیمتی اجتناب شود. خواننده گزارش انتظار ندارد که یافته‌های تحقیقاتی تحت چنین عناوینی آمده باشند، اما از مقالاتی که با استفاده از کار دیگران تهیه شده و ضمن آن یک نقطه نظر خاص ارائه می‌شود، چنین انتظاری می‌رود.

## ۲- متن

**الف - مقدمه** بخش مقدمه هرچه را که برای آماده سازی زمینه تحقیق انجام شده است، در بر می‌گیرد. این بخش با تغییراتی جزئی، شامل همان مطالبی است که برای طرح پیشنهادی تهیه شده است. تبیین مسئله و توجیه بررسی و همچنین تبیین هدفها و تعاریف واژه‌ها و مرور تحقیقات پیشین به همان صورت باقی می‌مانند، ولی شرح مختصر منابع داده‌ها و روشهای مورد استفاده، در اینجا به زمان ماضی نوشته می‌شود. در حالی که در طرح پیشنهادی به زمان آینده نوشته شده است. بازنگری تحقیقات مرتبط معمولاً در فصلی جداگانه از بخش مقدمه، ارائه می‌شود.

**ب - روشها و نتایج** چهار طبقه از اطلاعات در این بخش گزارش آورده می‌شوند:

۱- **آزمودنیها:** یک توصیف مشروح از نمونه باید ارائه شود. این امر امکان قضاوت درباره اعتبار بیرونی جامعه آماری بالقوه تحقیق را به خواننده می‌دهد. جامعه آماری که نمونه از آن گرفته شده است، باید تعریف و روش نمونه گیری مشخص شود. نوع اطلاعات ارائه شده در مورد توصیف نمونه از یک بررسی به بررسی دیگر تغییر می‌کند، ولی به طور کلی محقق

می‌تواند با در نظر گرفتن متغیرهای احتمالی مؤثر بر نمرات ملاکی در بررسی، در مورد نوع اطلاعات قابل ارائه تصمیم‌گیری کند.

۲- روش: تدوین روش بررسی باید به اندازه کافی کامل باشد تا اگر کسی بخواهد بررسی را تکرار کند، تمام اطلاعات لازم را در این قسمت بیابد. یکی از ویژگیهای رویکرد علمی، امکان تأیید یافته‌ها به وسیله تکرار همان شیوه‌ها و اطلاعات مشاهده‌ای بررسی است. در این بخش باید طرح بررسی، تعداد گروهها (در صورتی که بررسی آزمایشی باشد)، آزمودنیها و دیگر اطلاعات مربوط، آورده شود.

۳- ابزار اندازه‌گیری: یک گزارش تحقیقاتی باید کلیه ابزارهای اندازه‌گیری و روشهای مشاهده‌ای مورد استفاده را مشخص کند. در صورتی که از اندازه‌های از پیش آماده استفاده شود، این مشخصات می‌توانند مختصر باشند؛ زیرا منابع و مآخذ گزارش حاوی اطلاعات مربوط به این ابزارها خواهند بود. چنانچه ابزار خاصی برای بررسی تهیه شده باشد، توصیف مبسوطی از این ابزار باید فراهم شود. همچنین باید شواهدی از پایایی و اعتبار و مبحثی در مورد شیوه نمره‌گذاری نیز ارائه شود.

۴- ارائه و تحلیل داده‌ها: یک روش پیشنهادی، سازمان دادن بحث حول فرضیه‌هاست. بدین معنا که فرضیه اول دوباره بیان و یافته‌های مربوط به آن ارائه شود، سپس این روش برای هر یک از فرضیه‌ها به نوبت تکرار شود.

ممکن است با ارائه جداول و نمودارها بتوان داده‌ها را صریح‌تر و موجزتر از صورت نوشتاری آنها ارائه کرد. غالب کتابهای راهنمای نگارش، نمونه‌هایی از انواع متداول جداول و نمودارها را ارائه و نحوه تهیه آنها را آموزش می‌دهند. یک جدول خوش‌ساخت دیدی کلی از داده‌ها به خواننده می‌دهد.

جداولی را که محقق حین اجرای بررسی ساخته است، معمولاً نمی‌توان مستقیماً در گزارش آورد. برای مثال، در پایان یک بررسی، محقق ممکن است فهرستی الفبایی از آزمودنیهای بررسی را با نمرات آنها در اندازه‌های ملاکی داشته باشد. او به جای آنکه این فهرست را به همان صورت ارائه کند، می‌تواند جدولی از این اطلاعات به صورت خلاصه بسازد (داده‌های پایه‌ای خام را در صورتی که احساس شود به درک مطلب کمک می‌کند، می‌توان در پیوسته‌ها آورد). معمولاً اولین جدول گزارش اطلاعات توصیفی نظیر میانگینها، انحراف معیارها، همبستگیها یا درصدها را

خلاصه می‌کند. جداول بعدی نتایج بکارگیری آمار استنباطی و آزمونه‌های معنی داری را در مورد داده‌ها، خلاصه می‌کند. برای مثال، از جدول خلاصه‌ای برای نشان دادن نتایج یک تحلیل واریانس، استفاده می‌شود.

مطلوب است که جداول به طریقی ترتیب یابند که رابطه داده‌ها را با فرضیه‌های مورد بررسی نشان دهند. محققان تازه کار معمولاً تمایل دارند که داده‌ها را به هر دو صورت ارائه دهند. این روش صرفاً گزارش را طولانی‌تر و خسته‌کننده‌تر می‌سازد. رویکرد بهتر این است که داده‌ها در جداول و نمودارها ارائه شوند و توضیحات کافی برای اشاره به مهمترین و جالبترین یافته‌ها با آنها همراه شود. بخصوص این نکته اهمیت دارد که اطلاعات جداول با فرضیه‌ها ربط داده شوند.

مبنای آماری تحلیل نتایج باید به صورت روشن بیان شود. مناسب است که شیوه‌های به کار گرفته شده آماری با ارائه داده‌ها تکمیل شوند.

**ج - بحث در مورد نتایج** یافته‌ها بار دیگر در ارتباط با فرضیه‌ها (یا سؤالات) تفسیر می‌شوند، معانی ضمنی و کاربردهای بررسی مورد بحث قرار می‌گیرند.

۱- **تفسیر یافته‌ها:** تفسیر یافته‌ها احتمالاً مشکلترین و همچنین ارضاکنده ترین بخش تهیه گزارش است. تفسیر محقق از نتایج، این یافته‌ها را به نظریه و سایر تحقیقات و روشهای تحقیق حیطه مورد نظر مرتبط می‌سازد.

۲- **دلایل:** سهم نتایج در دانش موجود در حیطه کلی بررسی، مسئله‌ای است که باید در این بخش بحث شود. محقق توضیح می‌دهد که چگونه نتایج می‌توانند نظریه‌های مرتبط را تغییر دهند و بررسیهای دیگری را که منطقاً به دنبال این بررسی می‌آیند، پیشنهاد کند.

۳- **کاربردها:** توضیحی در مورد کاربرد یافته‌ها می‌تواند به خوانندگان گزارش کمک کند تا در مورد چگونگی کاربرد یافته‌ها در عمل بیندیشند.

به بخش معانی ضمنی و کاربردهای نتایج معمولاً به اندازه کافی پرداخته نمی‌شود؛ زیرا فرض می‌شود که این مسائل برای خواننده گزارش نیز به اندازه خود محقق روشن است. در واقع، حین اجرای بررسی، محققان به پیشنی دست می‌یابند که عمیق‌تر از پیشنی است که می‌توان برای خواننده فرض کرد. در نتیجه، انتظار می‌رود تفسیر آنها معنی دارتر از تفسیری باشد که خوانندگان برای خود انجام می‌دهند.

د- نتیجه گیری ها و خلاصه      بحثهای نتیجه گیری و خلاصه، روی هم آخرین سنگ بنای گزارش است.

۱- نتیجه گیری ها: بحث در مورد نتیجه گیریهایی که یافته های تحقیق بر آنها دلالت دارند، باید به نتیجه گیریهایی محدود شود که یافته های تحقیق به طور مستقیم آنها را تأیید می کنند. محققان اغلب وسوسه می شوند که نتیجه گیریهای زیادی انجام دهند. فرضیه ها چهارچوب مطمئنی را برای تبیین نتیجه گیریها فراهم می سازند، بدین معنی که محقق در این بخش باید مشخص کند که آیا یافته ها فرضیه هایش را تأیید می کند یا نه.

بسیار مهم است که بین نتایج و نتیجه گیریها تمیز قائل شویم. نتایج عبارتند از مشاهدات مستقیمی که خلاصه شده و با تحلیل آماری نظیر تحلیل آماری میانگین های دو گروه، تکمیل شده اند. نتیجه گیری، یک استنتاج بر مبنای نتایج است که بر حسب فرضیه های بررسی بیان می شود، مثل اینکه بگوییم تیمار یک گروه، مؤثرتر از تیمار گروه دیگر است. برای مثال، یک بررسی ممکن است به این نتیجه برسد که میانگین نمرات آزمون هجی کردن دانش آموزانی که با روش A آموزش دیده اند، به صورت معنی داری بالاتر از دانش آموزانی است که با روش B آموزش دیده اند. این نتیجه گیری که روش A مؤثرتر از روش B است، نتیجه مستقیم بررسی نیست، بلکه استنتاجی بر مبنای نتایج بررسی می باشد.

محقق ممکن است ایده هایش را در مورد معانی ضمنی یافته ها و توصیه هایش را برای کاربرد احتمالی یافته ها، به صورت مختصر بیاورد. همچنین ممکن است هر سؤال جدیدی را که حاصل بررسی اوست، برای تحقیق بعدی پیشنهاد کند.

۲- خلاصه: از آنجا که خلاصه گزارش بیش از هر بخش دیگر آن، خواننده دارد، انشای آن باید بخصوص روشن و دقیق باشد. خلاصه، معمولاً شامل بیان دوباره مسئله (ها)، نکات اصلی روشها و مهمترین یافته های بررسی است. پس از تکمیل پیش نویس این بخش، نویسنده باید به دقت آن را واری و تعیین کند که آیا توصیف مختصر، اما تقریباً کاملی از بررسی و یافته های آن ارائه می دهد یا خیر. همچنین محقق باید مراقب باشد که اطلاعاتی را در این بخش نگنجانند که در بخشهای مربوط وجود نداشته باشند. ایده خوبی است که از یک همکار خواسته شود تا بخش نتیجه گیری را بخواند و ببیند که آیا این بخش مقاصد نویسنده را به خوبی می رساند یا خیر.



### ۳- صفحات متمم

**الف - کتابنامه** باید کلیه منابع ذکر شده در متن و پانویسها را شامل شود. غالب دانشگاهها اصرار دارند که فقط این منابع فهرست شوند، ولی تنها تعداد معدودی می‌خواهند مراجع مربوط هم که مشخصاً ذکر نشده‌اند، در فهرست بیایند. کتاب راهنمای نگارش انتخابی شما، جزئیات کامل روش ارائه فهرست مراجع را ارائه می‌دهد. دنبال کردن این قواعد به صورت دقیق و کامل بسیار اهمیت دارد. در حقیقت، روش بهتر این است که قبل از آنکه در جستجوی پیشینه تحقیقاتی بررسی خود باشیم، این نکته را بدانیم. با فهرست برداری صحیح از مراجع در حین برخورد با آنها، می‌توان از صرف وقت اضافی برای بازیابی مراجع به منظور استفاده کامل آنها در کتابنامه، اجتناب کرد. بهتر است آنها را بر روی کارتهایی بیاوریم و یا وارد کامپیوتر شخصی کنیم تا بتوانیم آنها را به ترتیب الفبایی بایگانی کنیم.

**ب - پیوستها** حاوی مطالبی هستند که برای قرار گرفتن در متن اصلی گزارش، به اندازه کافی حائز اهمیت نیستند، ولی ارزش خواندن را برای برخی خوانندگان دارند. از بین این مطالب می‌توان از نسخه های کامل آزمونها و پرسشنامه های ساخته شده داخلی، همراه با دستورالعملها و راهنمای نمره گذاری اینگونه ابزار، اطلاعات تحلیل سوالات پرسشنامه های مورد استفاده، دستورالعملهای شفاهی برای آزمودنیها، جداولی که بیش از حد طولانی هستند و یا از نظر بررسی اهمیت زیادی ندارند، نام برد.

**ج - شناسنامه کاری مؤلف یا مؤلفان پایان نامه** بعضی مواقع از نویسندگان گزارشهای تحقیقاتی خواسته می‌شود شرح مختصری از آموزش، تجارب، عضویت‌های حرفه‌ای و پژوهشهای قبلی خود را در این بخش بیاورند.

**د - چکیده** غالب سازمانها بر ضرورت ارائه چکیده‌ای جداگانه برای رساله‌ها تأکید می‌کنند. این چکیده شامل بیان مسئله به صورت مشخص و توصیف موجز روشهای بررسی، یافته های اصلی، نتیجه گیریها و کاربردها است. چکیده نباید بیش از حد طولانی باشد (نوعاً ۶۰۰ کلمه یا کمتر). صفحات مربوط به چکیده، جداگانه شماره گذاری می‌شود و در اول یا آخر پایان نامه، قرار می‌گیرد.

// نگارش یک مقاله برای مجله

در تهیه یک مقاله تحقیقاتی جهت نشر در یک مجله، بهترین کار این است که در اولین قدم به

کتابخانه مراجعه کنیم تا ببینیم کدام مجله بیشترین مقاله را در زمینه مورد نظر منتشر کرده است. اطلاعات مربوط به نحوه ارائه دستنویس مقالات را معمولاً می‌توان در صفحه پشت جلد یافت. بسیاری از مجلات، کتاب راهنمای نگارشی را که باید مورد استفاده قرار گیرد، مشخص می‌کنند؛ برای مثال، کتاب راهنمای نگارش انجمن روان‌شناسی آمریکا یا کتاب نحوه نگارش NEA. چنانچه کتاب راهنمای نگارش مشخص نشده باشد، سبک ارجح، روش مأخذنویسی و غیره را از بررسی یک مقاله که در آخرین شماره مجله آمده است، می‌توان تعیین کرد.

یک مقاله تحقیقاتی از همان مراحل یک پایان‌نامه تبعیت می‌کند، ولی بسیار کوتاهتر است. یک پایان‌نامه یا رساله نشان‌دهنده صلاحیت یک دانشجو و توصیف کاملی از تحقیقات مرتبط و شیوه مورد استفاده و نیز حاوی جدول بندی کامل نتایج و تفسیرهای اندیشمندانه است. از سوی دیگر، ویژگی یک مقاله تهیه شده برای مجله، تنها بیان سهم نویسنده در فعالیتهای علمی است. برای صرفه جویی در فضای نوشتاری مجله و جلب دقت خواننده، مقاله باید مختصر باشد. بخش تحقیقات مرتبط، فقط شامل آن نتایج و مباحثی می‌شود که مبنای مسئله مورد نظر فرد باشد. تبیین کلی مسئله در یک بند آورده می‌شود و حتی می‌تواند حذف شود، در این مورد مقاله با فرضیه شروع می‌شود. بخش روشها نیز باید بسیار مختصر آورده شود، هرچند در صورت امکان کلیه اطلاعات لازم برای تکرار آزمایش باید ذکر شود. بخش نتایج، قابل توجه ترین بخش برای خواننده خواهد بود و در نتیجه در مقاله بخش بزرگتری را، نسبت به پایان‌نامه، به خود اختصاص خواهد داد. تنها مهمترین یافته‌ها باید مفصلاً مورد بحث قرار گیرند.

یک نامه مختصر باید همراه نسخه دستنویس باشد. ویراستار معمولاً یک کارت تشکر از دریافت نسخه دست‌نویس را برای نویسنده ارسال می‌کند و نسخه‌هایی از آن را برای اعضای هیئت تحریریه جهت بازنگری می‌فرستد. به این دلیل معمولاً زمان قابل ملاحظه‌ای لازم است تا به نویسنده اطلاع داده شود که آیا مقاله او پذیرفته شده است یا نه (به طور معمول ۶ هفته). پس از آنکه یک مقاله پذیرفته شد، معمولاً ماهها طول می‌کشد تا در مجله به چاپ برسد.

وقتی نسخه دستنویس از طرف یک مجله رد می‌شود، اعلام رد، بعضی مواقع با ذکر دلایل آن همراه است. رد شدن یک مقاله از سوی یک مجله ضرورتاً بدین معنا نیست که مقاله ارزش چاپ ندارد. عوامل مختلفی مانند رقابت برای محل چاپ مقاله، تغییر در سیاست هیئت تحریریه و تعصب بازنگران، می‌توانند در تصمیم برای چاپ مقاله اثر گذارند. مقاله‌ای که به

وسیله یک مجله رد شده است، می‌تواند اصلاح و به مجله دیگری تسلیم شود. از نظر اخلاقی صحیح نیست که یک مقاله را در یک زمان به بیش از یک مجله ارائه کرد، ولی بسیاری از مقالات قبل از آنکه جای مناسب خود را پیدا کنند، به چندین مجله ارسال می‌شوند.

### // میثاق حرفه‌ای

بسیاری از محققان، استماع مقالاتی را که در انجمنهای حرفه‌ای قرائت می‌شوند، راه خوبی برای روزآمد کردن اطلاعات خود در حیطه مورد نظر می‌دانند. دلیل این امر آن است که فاصله زمانی زیادی بین اتمام یک پروژه و ظهور آن در مجله وجود دارد. این فاصله زمانی معمولاً آنچنان طولانی است که نشریات حرفه‌ای را اغلب دارای ماهیت بایگانی توصیف می‌کنند. نحوه تهیه مقالاتی که در جلسات حرفه‌ای ارائه می‌شوند، تا اندازه زیادی شبیه نحوه تهیه مقالات مجلات است. این مقالات لزوماً گزارش تحقیقهای تکمیل شده نیستند، بلکه می‌توانند گزارشهای پیشرفت کار مربوط به پروژه‌های در حال اجرا باشند. مقالات قرائت شده کمتر از مقالات مجلات حالت رسمی دارند و معمولاً می‌توانند با دقت برای مخاطبانشان آماده شوند. عموماً از مخاطبان می‌توان انتظار داشت که با جزئیات تحقیقات مربوطه و روشهای اندازه‌گیری آشنایی داشته باشند.

مقاله معمولاً به طریق زیر سازمان می‌یابد:

۱- تبیین مستقیم فرضیه

۲- توصیف مختصر روشها

۳- یافته‌ها، نتیجه‌گیریها و کاربردها

زمان اختصاص داده شده برای خواندن یک مقاله معمولاً خیلی کوتاه و در بیشتر موارد کمتر از ۱۵ دقیقه است. بنابر این، مقاله باید بر مهمترین جنبه‌های بررسی تأکید داشته باشد. برای اطمینان می‌توان  $\frac{۲}{۳}$  تا ۳ دقیقه برای هر صفحه که با فاصله دوبرابر تایپ شده باشد، اختصاص داد.

چنانچه نمودارها و جداول به تفهیم مطلب کمک کنند، می‌توان نسخی از آنها را در دسترس مخاطبان قرار داد. برخی سخنرانان نسخی از متن مقاله‌شان را نیز بین مخاطبان توزیع می‌کنند. اما به هر حال، از آنجا که مخاطبان می‌توانند قبل از آن که سخنرانی به اتمام برسد، متن آن را

بخوانند، امکان دارد دچار بی توجهی شوند. بهتر است متن کامل بررسی در اختیار کسانی که بعد از سخنرانی، آن را تقاضا می‌کنند، قرار داده شود.

### /// راهنماهای نگارش

کتاب زیر به صورت وسیعی مورد استفاده قرار می‌گیرند. این راهنما به صورت مفصل، شکل و سبک کلی را برای تهیهٔ رساله‌ها و پایان‌نامه‌ها ارائه می‌کنند.

- American Psychological Association (1983), *Publication Manual* (3d ed.).

Washington: American Psychological Association.

- Campbell, W.G. and S.V. Ballou (1982). *Form and Style: Theses, Reports, Term Papers* (6th ed.). Boston: Houghton Mifflin.

Reports, Term Papers (6th ed.). Boston: Houghton Mifflin.

- *The Chicago Manual of Style* (1982, 13th ed., rev.). Chicago: University

of Chicago Press.

- Turabian, K. (1987). *A Manual for Writers of Term Papers, Theses and*

*Desertations* (5th ed.). Chicago: University of Chicago Press.

### /// خلاصه

با در نظر گرفتن این امر که در غالب دانشگاهها، تجهیزات محاسبهٔ الکترونیکی در دسترس است، بسیاری از محققان می‌توانند از این تجهیزات در پردازش داده‌هایشان بهره‌جویند. استفاده از کامپیوتر باعث صرفه‌جویی در وقت و انرژی می‌شود و در مسائل پیچیده، دقت عمل را تا حد بسیار زیاد تسهیل می‌کند.

به منظور استفاده از کامپیوتر برای پردازش داده‌ها، محقق باید داده‌ها و دستورالعملهای لازم را به شکلی برای کامپیوتر آماده کند تا کامپیوتر بتواند به وسیلهٔ یک برنامهٔ مناسب داده‌ها را پردازش کند.

نوشتن برنامه‌های کامپیوتری نیاز به آموزش خاص دارد، ولی محقق معمولاً می‌تواند جهت تحلیل داده‌ها از برنامه‌های موجود استفاده کند.

در صورتی که در طرح پیشنهادی تحقیق، محقق بنیان مناسبی را برای تحقیق فراهم ساخته

باشد، تفسیر نتایج یک بررسی، امری ساده خواهد بود. موارد احتیاطی زیر را باید به خاطر داشت: ۱- تفسیر، منحصرأ باید بر مبنای داده های حاصل از بررسی باشد، ۲- مشکلات اعتبار درونی و بیرونی و دیگر محدودیتهای بررسی، مورد نظر قرار گیرند و ۳- نتایج باید به صورت عبارات احتمالی و نه واقعیتها ارائه شود.

شایسته است که نتایج منفی به همان صورت نتایج مثبت، مورد نظر و تفسیر قرار گیرند. فرضیه صفری که رد نشده است، باید به عنوان شواهد غیر کافی تفسیر شود و نه چیز دیگر. نتایج فرض نشده، به عنوان مبنایی برای فرضیات آینده، قابل توجه هستند. روشهای رسمی برای تهیه پایان نامه، مقالات مجلات و گزارشها تهیه شده اند. تبحر در این روشها و بکارگیری آنها به محققان کمک می کند که با دیگر محققان و دست اندرکاران تعلیم و تربیت ارتباط پیدا کنند. کتابهای راهنمای متعددی جزئیات این روشها را به دست می دهند.

### // مفاهیم اصلی

تفسیر فرضیه صفر رد نشده	بسته نرم افزاری
تفسیر نتایج منفی	تفسیر نتایج غیر قابل انتظار
تهیه رساله یا پایان نامه	تفسیر نتایج مورد انتظار
تهیه مقالات جهت کنفرانسها	تهیه مقالات برای مجلات
کتاب راهنمای شیوه نگارش	کاربرد کامپیوتر
	معنی داری عملی در مقابل آماری

### /// تمرینها

- ۱- ظهور کامپیوتر چه تغییری را در انجام تحقیقات در زمینه تعلیم و تربیت ایجاد کرده است؟
- ۲- سه برنامه کامپیوتری و تجارتمی موجود را برای تحلیل آماری، نام ببرید.
- ۳- تفاوت بین نتایج و نتیجه گیری چیست؟
- ۴- تفاوت بین معنی داری آماری و اهمیت بررسی را توضیح دهید.
- ۵- چه عواملی می توانند از رد شدن فرضیه صفر ممانعت کنند؟

۶- بگوئید کدامیک از عناوین زیر قابل قبول و کدام غیر قابل قبول هستند؟ دلایل خود را ذکر کنید.

الف - معدل دروس و آموزش رانندگی

ب - اثرات تدریس خصوصی سه جلسه در هفته توسط دانش آموزان کلاس ششم بر نحوه قرائت دانش آموزان کلاس دوم که از نظر توانایی پایین تر از میانگین هستند.

ج - باید به کودکان، ریاضیات جدید درس داد!

د - رابطه بین ویژگیهای شخصیتی و نگرش نسبت به پیشرفت تحصیلی دانش آموزان توانا و ضعیف در درس خواندن.

۷- چه تفاوتی در شکل تحقیقات گزارش شده به شکل پایان نامه، مقاله مجله و مقاله‌ای که

در یک کنفرانس خوانده می‌شود، وجود دارد؟

۸- آیا محقق، نتایج تحقیق را که با فرضیه او همخوانی ندارد، باید مورد بحث قرار دهد؟

### /// پاسخها

۱- کامپیوتر امکان طراحی و اجرای بررسیهای تحقیقاتی را بدون در نظر گرفتن تعداد متغیرها یا پیچیدگی تحلیل، فراهم آورده است. داده های حاصل از بررسیهایی را که حجم نمونه بزرگ دارند و از روشهای چند متغیری استفاده می‌کنند یا نیاز به آمارهای پیچیده دارند، می‌توان با سرعت و دقت تحلیل کرد.

۲- بسته نرم افزاری برای علوم اجتماعی (SPSS-X)، مجموعه زیست پزشکی (BIMED) و سیستم تحلیل آماری (SAS).

۳- نتیجه یک مشاهده مستقیم است. نتیجه گیری استنتاج انجام شده بر مبنای نتایج است.

۴- معنی داری آماری بدین معنی است که نتایج تابعی از احتمال نیستند. اهمیت بررسی براساس اهمیت یافته ها با توجه به آزمون فرضیه یا معانی عملی، تعیین می‌شود.

۵- رد نشدن یک فرضیه صفر می‌تواند به این دلیل باشد که در عمل فرضیه صفر ماهیتاً صحیح است یا می‌تواند ناشی از مخدوش شدن نتایج، به دلیل مشکلات اعتبار درونی باشد که اثرات تیمار را مبهم می‌سازد و یا به دلیل فقدان توانایی آماری طرح مورد استفاده بررسی و یا عدم توانایی در اندازه گیری دقیق اثرات تیمار بر متغیر وابسته باشد.

۶- الف - غیر قابل قبول: بیان رابطه صورت نگرفته است.

ب - غیر قابل قبول: از کلمات زیادی استفاده شده است.

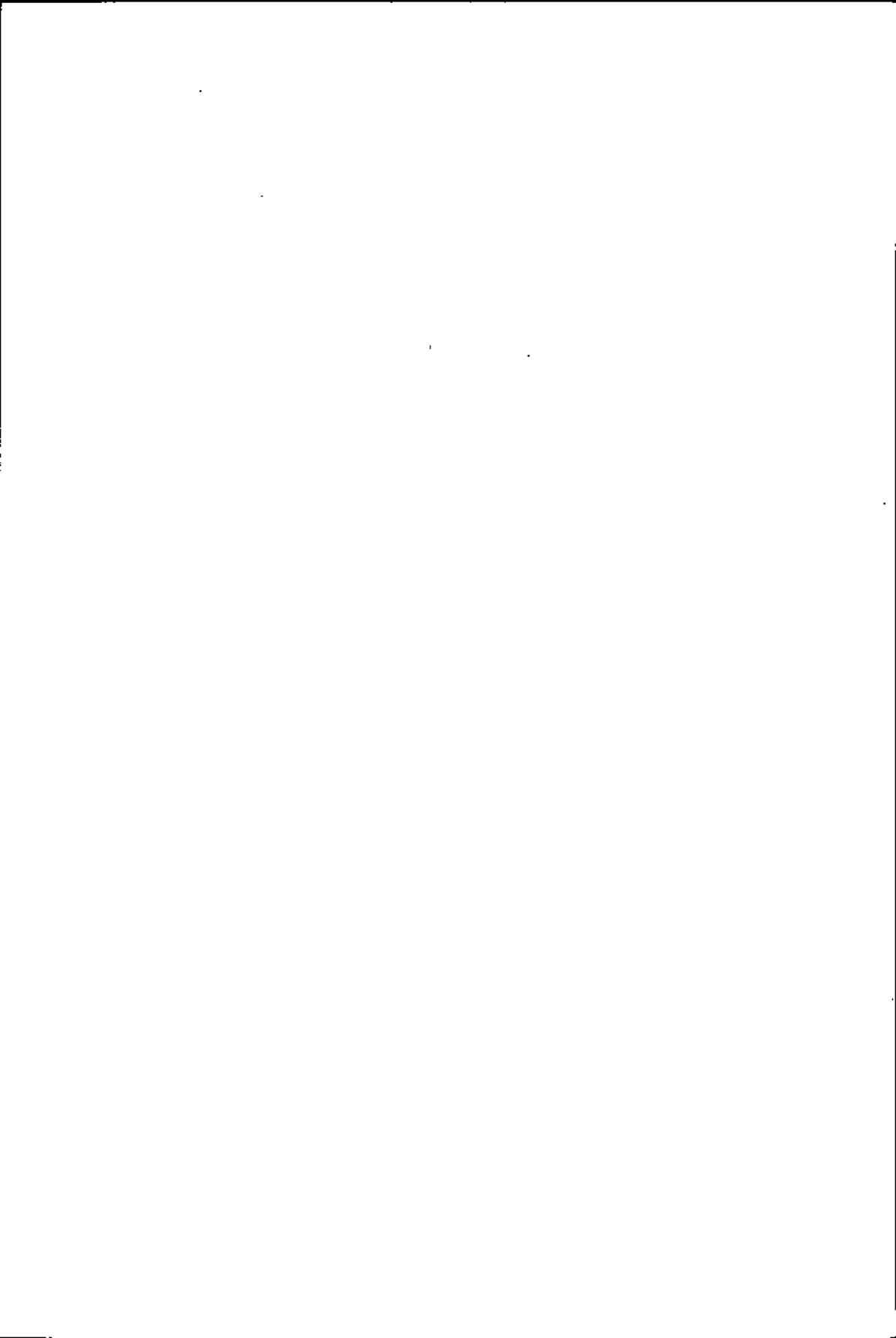
ج - غیر قابل قبول: عناوینی که متکی بر احساسات می‌باشند، برای مقالات تحقیقاتی مناسب نیستند.

د - قابل قبول: ضوابط لازم را برای عنوان دارا می‌باشد.

۷- شکل پایان نامه، رسمی‌ترین و مفصل‌ترین شکل مقاله است و از موارد مشخص شده در راهنمای شیوه نگارش خاصی پیروی می‌کند. مقاله تهیه شده برای مجله موجزتر و دارای یک تبیین مختصر مسئله، مرور بر پژوهشهای پیشین و روش شناسی است. بخش اعظم مقاله به نتایج اصلی و بحثی در مورد اهمیت آن اختصاص دارد. مقاله‌ای که برای خواندن در کنفرانس تهیه می‌شود، کمتر از همه جنبه رسمی دارد، خود را با مخاطباناش تطبیق می‌دهد، فرضیه‌ها را تبیین و روش را به اختصار توصیف می‌کند و بر مهمترین یافته‌ها تأکید دارد.

۸- بله، برخلاف انتظارات محقق، این نتایج به اندازه دیگر نتایج، معقول است و به همان

نحو نیز باید تفسیر شود.





ضمائم

جدول ۱- A سطوح زیر منحنی نرمال.

(2)			(2)		
(1)	سطح بین میانگین و z	(3) سطح فراتر از z	(1)	سطح بین میانگین و z	(3) سطح فراتر از z
z			z		
0.00	.0000	.5000	0.45	.1736	.3264
0.01	.0040	.4960	0.46	.1772	.3228
0.02	.0080	.4920	0.47	.1808	.3192
0.03	.0120	.4880	0.48	.1844	.3156
0.04	.0160	.4840	0.49	.1879	.3121
0.05	.0199	.4801	0.50	.1915	.3085
0.06	.0239	.4761	0.51	.1950	.3050
0.07	.0279	.4721	0.52	.1985	.3015
0.08	.0319	.4681	0.53	.2019	.2981
0.09	.0359	.4641	0.54	.2054	.2946
0.10	.0398	.4602	0.55	.2088	.2912
0.11	.0438	.4562	0.56	.2123	.2877
0.12	.0478	.4522	0.57	.2157	.2843
0.13	.0517	.4483	0.58	.2190	.2810
0.14	.0557	.4443	0.59	.2224	.2776
0.15	.0596	.4404	0.60	.2257	.2743
0.16	.0636	.4364	0.61	.2291	.2709
0.17	.0675	.4325	0.62	.2324	.2676
0.18	.0714	.4286	0.63	.2357	.2643
0.19	.0753	.4247	0.64	.2389	.2611
0.20	.0793	.4207	0.65	.2422	.2578
0.21	.0832	.4168	0.66	.2454	.2546
0.22	.0871	.4129	0.67	.2486	.2514
0.23	.0910	.4090	0.68	.2517	.2483
0.24	.0948	.4052	0.69	.2549	.2451
0.25	.0987	.4013	0.70	.2580	.2420
0.26	.1026	.3974	0.71	.2611	.2389
0.27	.1064	.3936	0.72	.2642	.2358
0.28	.1103	.3897	0.73	.2673	.2327
0.29	.1141	.3859	0.74	.2704	.2296
0.30	.1179	.3821	0.75	.2734	.2266
0.31	.1217	.3783	0.76	.2764	.2236
0.32	.1255	.3745	0.77	.2794	.2206
0.33	.1293	.3707	0.78	.2823	.2177
0.34	.1331	.3669	0.79	.2852	.2148
0.35	.1368	.3632	0.80	.2881	.2119
0.36	.1406	.3594	0.81	.2910	.2090
0.37	.1443	.3557	0.82	.2939	.2061
0.38	.1480	.3520	0.83	.2967	.2033
0.39	.1517	.3483	0.84	.2995	.2005
0.40	.1554	.3446	0.85	.3023	.1977
0.41	.1591	.3409	0.86	.3051	.1949
0.42	.1628	.3372	0.87	.3078	.1922
0.43	.1664	.3336	0.88	.3106	.1894
0.44	.1700	.3300	0.89	.3133	.1867

جدول ۱- A سطوح زیر منحنی نرمال (ادامه)

(2)			(2)		
(1)	سطح بین: میانگین و z	(3) سطح فراتر از z	(1)	سطح بین میانگین و z	(3) سطح فراتر از z
z			z		
0.90	.3159	.1841	1.35	.4115	.0885
0.91	.3186	.1814	1.36	.4131	.0869
0.92	.3212	.1788	1.37	.4147	.0853
0.93	.3238	.1762	1.38	.4162	.0838
0.94	.3264	.1736	1.39	.4177	.0823
0.95	.3289	.1711	1.40	.4192	.0808
0.96	.3315	.1685	1.41	.4207	.0793
0.97	.3340	.1660	1.42	.4222	.0778
0.98	.3365	.1635	1.43	.4236	.0764
0.99	.3389	.1611	1.44	.4251	.0749
1.00	.3413	.1587	1.45	.4265	.0735
1.01	.3438	.1562	1.46	.4279	.0721
1.02	.3461	.1539	1.47	.4292	.0708
1.03	.3485	.1515	1.48	.4306	.0694
1.04	.3508	.1492	1.49	.4319	.0681
1.05	.3531	.1469	1.50	.4332	.0668
1.06	.3554	.1446	1.51	.4345	.0655
1.07	.3577	.1423	1.52	.4357	.0643
1.08	.3599	.1401	1.53	.4370	.0630
1.09	.3621	.1379	1.54	.4382	.0618
1.10	.3643	.1357	1.55	.4394	.0606
1.11	.3665	.1335	1.56	.4406	.0594
1.12	.3686	.1314	1.57	.4418	.0582
1.13	.3708	.1292	1.58	.4429	.0571
1.14	.3729	.1271	1.59	.4441	.0559
1.15	.3749	.1251	1.60	.4452	.0548
1.16	.3770	.1230	1.61	.4463	.0537
1.17	.3790	.1210	1.62	.4474	.0526
1.18	.3810	.1190	1.63	.4484	.0516
1.19	.3830	.1170	1.64	.4495	.0505
1.20	.3849	.1151	1.65	.4505	.0495
1.21	.3869	.1131	1.66	.4515	.0485
1.22	.3888	.1112	1.67	.4525	.0475
1.23	.3907	.1093	1.68	.4535	.0465
1.24	.3925	.1075	1.69	.4545	.0455
1.25	.3944	.1056	1.70	.4554	.0446
1.26	.3962	.1038	1.71	.4564	.0436
1.27	.3980	.1020	1.72	.4573	.0427
1.28	.3997	.1003	1.73	.4582	.0418
1.29	.4015	.0985	1.74	.4591	.0409
1.30	.4032	.0968	1.75	.4599	.0401
1.31	.4049	.0951	1.76	.4608	.0392
1.32	.4066	.0934	1.77	.4616	.0384
1.33	.4082	.0918	1.78	.4625	.0375
1.34	.4099	.0901	1.79	.4633	.0367

جدول ۱- A سطوح زیر منحنی نرمال (ادامه)

(1) z	(2) Area سطح بین میانگین و z and z	(3) سطح فراتر از z	(1) z	(2) سطح بین میانگین و z and z	(3) سطح فراتر از z
1.80	.4641	.0359	2.25	.4878	.0122
1.81	.4649	.0351	2.26	.4881	.0119
1.82	.4656	.0344	2.27	.4884	.0116
1.83	.4664	.0336	2.28	.4887	.0113
1.84	.4671	.0329	2.29	.4890	.0110
1.85	.4678	.0322	2.30	.4893	.0107
1.86	.4686	.0314	2.31	.4896	.0104
1.87	.4693	.0307	2.32	.4898	.0102
1.88	.4699	.0301	2.33	.4901	.0099
1.89	.4706	.0294	2.34	.4904	.0096
1.90	.4713	.0287	2.35	.4906	.0094
1.91	.4719	.0281	2.36	.4909	.0091
1.92	.4726	.0274	2.37	.4911	.0089
1.93	.4732	.0268	2.38	.4913	.0087
1.94	.4738	.0262	2.39	.4916	.0084
1.95	.4744	.0256	2.40	.4918	.0082
1.96	.4750	.0250	2.41	.4920	.0080
1.97	.4756	.0244	2.42	.4922	.0078
1.98	.4761	.0239	2.43	.4925	.0075
1.99	.4767	.0233	2.44	.4927	.0073
2.00	.4772	.0228	2.45	.4929	.0071
2.01	.4778	.0222	2.46	.4931	.0069
2.02	.4783	.0217	2.47	.4932	.0068
2.03	.4788	.0212	2.48	.4934	.0066
2.04	.4793	.0207	2.49	.4936	.0064
2.05	.4798	.0202	2.50	.4938	.0062
2.06	.4803	.0197	2.51	.4940	.0060
2.07	.4808	.0192	2.52	.4941	.0059
2.08	.4812	.0188	2.53	.4943	.0057
2.09	.4817	.0183	2.54	.4945	.0055
2.10	.4821	.0179	2.55	.4946	.0054
2.11	.4826	.0174	2.56	.4948	.0052
2.12	.4830	.0170	2.57	.4949	.0051
2.13	.4834	.0166	2.58	.4951	.0049
2.14	.4838	.0162	2.59	.4952	.0048
2.15	.4842	.0158	2.60	.4953	.0047
2.16	.4846	.0154	2.61	.4955	.0045
2.17	.4850	.0150	2.62	.4956	.0044
2.18	.4854	.0146	2.63	.4957	.0043
2.19	.4857	.0143	2.64	.4959	.0041
2.20	.4861	.0139	2.65	.4960	.0040
2.21	.4864	.0136	2.66	.4961	.0039
2.22	.4868	.0132	2.67	.4962	.0038
2.23	.4871	.0129	2.68	.4963	.0037
2.24	.4875	.0125	2.69	.4964	.0036

جدول ۱-۸ سطوح زیر منحنی نرمال (ادامه)

(1)	(2)	(3)	(1)	(2)	(3)
$z$	سطح بین میانگین و $z$	سطح فراتر از $z$	$z$	سطح بین میانگین و $z$	سطح فراتر از $z$
2.70	.4965	.0035	3.05	.4989	.0011
2.71	.4966	.0034	3.06	.4989	.0011
2.72	.4967	.0033	3.07	.4989	.0011
2.73	.4968	.0032	3.08	.4990	.0010
2.74	.4969	.0031	3.09	.4990	.0010
2.75	.4970	.0030	3.10	.4990	.0010
2.76	.4971	.0029	3.11	.4991	.0009
2.77	.4972	.0028	3.12	.4991	.0009
2.78	.4973	.0027	3.13	.4991	.0009
2.79	.4974	.0026	3.14	.4992	.0008
2.80	.4974	.0026	3.15	.4992	.0008
2.81	.4975	.0025	3.16	.4992	.0008
2.82	.4976	.0024	3.17	.4992	.0008
2.83	.4977	.0023	3.18	.4993	.0007
2.84	.4977	.0023	3.19	.4993	.0007
2.85	.4978	.0022	3.20	.4993	.0007
2.86	.4979	.0021	3.21	.4993	.0007
2.87	.4979	.0021	3.22	.4994	.0006
2.88	.4980	.0020	3.23	.4994	.0006
2.89	.4981	.0019	3.24	.4994	.0006
2.90	.4981	.0019	3.30	.4995	.0005
2.91	.4982	.0018	3.40	.4997	.0003
2.92	.4982	.0018	3.50	.4998	.0002
2.93	.4983	.0017	3.60	.4998	.0002
2.94	.4984	.0016	3.70	.4999	.0001
2.95	.4984	.0016	3.90	.49995	.00005
2.96	.4985	.0015	4.00	.49997	.00003
2.97	.4985	.0015	4.50	.4999966	.0000034
2.98	.4986	.0014	5.00	.4999997	.0000003
2.99	.4986	.0014	5.50	.499999981	.000000019
3.00	.4987	.0013	6.00	.499999999	.000000001
3.01	.4987	.0013			
3.02	.4987	.0013			
3.03	.4988	.0012			
3.04	.4988	.0012			

جدول ۲- A جدول مقادیر t

df	سطح معنی داری برای یک آزمون جهت دار (یک دامنه)			سطح معنی داری برای یک آزمون فاقد جهت (دو دامنه)		
	.10	.05	.025	.10	.05	.02
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: Table A.2 is taken from Table III of Fisher and Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Longman Group UK Ltd., London (previously published by Oliver and Boyd, Ltd., Edinburgh), and by permission of the authors and publishers.

جدول A-۳ نقاط ۵ درصد (اعداد کم رنگ) و ۱ درصد (اعداد سیاه) برای توزیع F

مخرج کسر df = n <sub>۱</sub>	درجه های آزادی برای صورت کسر																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	101	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	254
2	4,052	4,999	5,403	5,625	5,744	5,829	5,928	5,981	6,022	6,056	6,083	6,106	6,122	6,149	6,208	6,234	6,258	6,286	6,302	6,323	6,334	6,352	6,361	6,366	
3	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,48	19,49	19,50	19,50	
4	98,49	99,00	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50	
5	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,68	8,64	8,62	8,60	8,58	8,57	8,50	8,54	8,54	8,54	
6	34,12	36,82	39,44	41,71	43,24	44,37	45,17	45,67	46,00	46,23	46,41	46,55	46,65	46,73	46,79	46,80	46,81	46,81	46,81	46,82	46,83	46,83	46,83	46,83	
7	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,60	5,65	5,65	5,65	
8	21,20	18,48	16,69	15,78	15,52	15,21	14,98	14,80	14,64	14,54	14,45	14,37	14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,44	
9	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	
10	16,26	13,27	12,06	11,39	10,97	10,47	10,45	10,27	10,15	10,05	9,96	9,89	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02	
11	5,99	5,14	4,78	4,53	4,39	4,28	4,21	4,15	4,10	4,00	4,03	4,00	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	
12	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,96	6,96	6,96	6,96	
13	6,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	
14	12,25	9,55	8,45	7,85	7,46	7,17	7,00	6,84	6,71	6,62	6,54	6,47	6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65	
15	5,32	4,48	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	
16	11,26	8,45	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,81	5,74	5,67	5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,94	4,91	4,88	4,86	
17	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,90	2,80	2,82	2,80	2,77	2,70	2,73	2,72	2,71	
18	10,54	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31	
19	4,90	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	
20	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91	
21	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40	
22	9,45	7,28	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60	
23	4,75	3,68	3,49	3,20	3,11	3,00	2,92	2,85	2,80	2,70	2,72	2,69	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,30	2,35	2,32	2,31	2,30	
24	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16	4,05	3,96	3,86	3,78	3,70	3,61	3,54	3,49	3,46	3,41	3,38	3,36	
25	4,07	3,00	3,41	3,18	3,02	2,92	2,84	2,77	2,73	2,67	2,63	2,60	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,20	2,24	2,22	2,21	
26	9,07	6,78	5,74	5,20	4,86	4,62	4,46	4,30	4,19	4,10	4,02	3,96	3,85	3,78	3,67	3,59	3,51	3,43	3,37	3,30	3,27	3,21	3,18	3,16	

جدول ۳-۸ نقاط ۵ درصد (اعداد کمزنگ) و ۱ درصد (اعداد سیاه) برای توزیع F (ادامه)

df = $n_1$	درجه‌های آزادی برای صورت کسر																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.00	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.11	2.13
15	8.84	6.51	5.56	5.03	4.59	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00	
16	4.54	3.69	3.29	3.00	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	
17	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87	
18	4.40	3.63	3.24	3.01	2.85	2.74	2.68	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	
19	8.51	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.89	2.86	2.80	2.77	2.75	
20	4.45	3.69	3.20	2.96	2.81	2.70	2.63	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	
21	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
22	4.41	3.65	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	
23	8.36	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
24	4.38	3.62	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.89	
25	8.30	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.28	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
26	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	
27	8.25	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.62	2.56	2.53	2.47	2.44	2.43	
28	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	
29	8.20	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.43	2.38	2.36	
30	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	
31	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.01	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
32	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.08	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	
33	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
34	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	
35	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
36	4.24	3.38	2.98	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	
37	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.33	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17	
38	4.22	3.37	2.98	2.74	2.58	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	
39	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	



جدول ۳-۱ نقاط ۵ درصد (اعداد کم رنگ) و ۱ درصد (اعداد سیاه) برای توزیع F (ادامه)

۳ درجه های آزادی برای صورت کسر

منبع کسر  
df =  $n_1$

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21	3.35	2.98	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	1.66
	7.68	5.49	4.80	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.11	2.10
28	4.20	3.34	2.95	2.71	2.60	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	1.65
	7.84	5.45	4.87	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.53	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.08	2.06
29	4.18	3.33	2.93	2.70	2.64	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	1.64
	7.68	5.42	4.84	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.05	2.03
30	4.17	3.32	2.92	2.69	2.63	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	1.62
	7.56	5.39	4.81	4.02	3.70	3.47	3.30	3.17	3.04	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01	2.01
32	4.15	3.30	2.90	2.67	2.61	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.88	1.82	1.77	1.74	1.69	1.67	1.64	1.61	1.59	1.59
	7.56	5.34	4.86	3.97	3.66	3.43	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.43	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96	1.96
34	4.13	3.28	2.85	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	1.57
	7.44	5.29	4.82	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91	1.91
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	1.55
	7.30	5.25	4.58	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87	1.87
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	1.53
	7.35	5.21	4.54	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.23	2.14	2.08	2.00	1.97	1.90	1.86	1.84	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	1.51
	7.31	5.18	4.51	3.83	3.51	3.29	3.12	2.99	2.85	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.84	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.48	1.48	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.89	1.83	1.78	1.75	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.46	1.46	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.74	1.71	1.71
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.73	1.73

جدول ۳-۵ نقاط ۵ درصد (اعداد کمرنگ) و ۱ درصد (اعداد سیاه) برای توزیع F (ادامه)

درجه‌های آزادی برای صورت کسر

df	df <sub>2</sub>																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
75	7.17	5.06	4.28	3.72	3.41	3.18	3.02	2.88	2.76	2.70	2.61	2.56	2.46	2.39	2.26	2.16	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
100	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.53	1.50	1.46	1.43	1.41
125	7.12	5.01	4.14	3.68	3.37	3.15	2.98	2.85	2.73	2.66	2.57	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
150	4.00	3.15	2.76	2.52	2.37	2.26	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.73	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
175	7.08	4.98	4.11	3.65	3.34	3.12	2.95	2.81	2.71	2.64	2.56	2.46	2.38	2.28	2.12	2.05	1.95	1.93	1.87	1.79	1.74	1.66	1.63	1.60
200	3.99	3.14	2.76	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
225	7.04	4.95	4.10	3.63	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.28	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56
250	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
275	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
300	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.86	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
325	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49
350	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
375	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
400	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
425	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
450	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
475	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
500	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
525	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97	1.88	1.79	1.69	1.63	1.53	1.46	1.39	1.33	1.28
550	3.86	3.02	2.63	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
575	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.30	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.41	1.32	1.24	1.19
600	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
625	6.66	4.62	3.80	3.34	3.04	2.83	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.36	1.28	1.19	1.11
650	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.06
675	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.34	1.25	1.15	1.09

Source: Reprinted by permission from *Statistical Methods*, 7th ed., by George W. Snustek and William G. Cochran. Copyright © 1980 by Iowa State University Press, Ames, Iowa 50010.

جدول ۲-۱. مقادیر  $\chi^2$

درجه‌های آزادی	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01
1	.000157	.000628	.00393	.0158	.0842	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635
2	.0201	.0404	.103	.211	.440	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	.115	.185	.352	.584	1.005	1.424	2.366	3.065	4.221	6.251	7.879	10.241	12.838
4	.297	.459	.711	1.064	1.649	2.195	3.357	4.278	5.989	7.779	9.488	11.668	13.277
5	.554	.732	1.145	1.610	2.343	3.000	4.351	6.084	7.289	9.236	11.070	13.388	15.086
6	.872	1.124	1.635	2.204	3.070	3.828	5.348	7.231	8.538	10.645	12.592	15.033	16.912
7	1.239	1.584	2.147	2.853	3.892	4.871	6.348	8.383	10.017	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.624	7.144	9.224	11.030	13.162	15.507	18.168	20.090
9	2.088	2.532	3.225	4.108	5.380	6.383	8.045	10.056	12.242	14.684	16.919	19.670	21.668
10	2.558	3.059	3.940	4.805	6.179	7.267	9.342	11.751	13.442	15.987	18.307	21.161	23.200
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.614	24.725
12	3.571	4.178	5.220	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.034	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.689
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.004	23.685	26.873	29.141
15	5.220	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.250	30.678
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.253	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.709	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.780	25.980	28.580	32.346	34.805
19	7.633	8.587	10.117	11.651	13.710	15.352	18.338	21.689	23.900	27.204	30.144	33.697	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.560
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.859	26.171	29.615	32.671	36.341	38.932
22	9.542	10.600	12.338	14.011	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.098	29.553	33.196	36.415	40.270	42.987
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.362	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.503	38.885	42.856	45.612
27	12.870	14.125	16.114	18.114	20.703	22.719	26.336	30.319	32.912	36.711	40.113	44.140	46.873
28	13.545	14.847	16.824	18.939	21.588	23.647	27.336	31.391	34.027	37.910	41.337	45.419	48.103
29	14.226	15.574	17.704	19.768	22.473	24.577	28.336	32.461	35.139	39.087	42.557	46.691	49.368
30	14.953	16.308	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.692

Source: Reprinted with permission of Macmillan Publishing Co. from *Statistical Methods for Research Workers*, by R. A. Fisher. Copyright © 1970 by the University of Adelaide, Australia.

جدول ۵- A مقادیر بحرانی ضریب همبستگی گشتاوری پیرسون

df = N - 2	سطح معنی داری برای یک آزمون فاقد جهت (دو دامنه)				
	.05	.025	.01	.005	.0005
	سطح معنی داری برای یک آزمون جهت دار (یک دامنه)				
	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Source: Table A.5 is taken from Table VII of Fisher and Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Longman Group UK Ltd., London (previously published by Oliver and Boyd, Ltd., Edinburgh), and by permission of the authors and publishers.

جدول ۶-۸ جدول اعداد تصادفی

ردیف	اعداد ستونی									
	00000	01234	00000	11111	11111	56789	22222	22222	22222	33333
00	23157	01837	54859	01837	25993	76249	70896	36744	95230	36744
01	05545	10537	55043	10537	43508	90611	83744	21343	10962	21343
02	14871	32404	60350	32404	75923	50051	00322	80834	11543	80834
03	38976	74951	74951	94051	75953	78905	90194	71696	32428	71696
04	97312	99755	61718	99755	30870	94251	25841	10513	54882	10513
05	11742	44339	69381	44339	30872	32797	33118	06850	22647	06850
06	43361	28859	28859	11016	45623	93009	00409	74030	43640	74030
07	93806	20478	20478	38268	04491	55751	18932	52571	58475	52571
08	49540	13181	13181	08429	84187	69538	29661	09527	77738	09527
09	36768	72633	72633	37948	21569	41959	68670	83840	45274	83840
10	07092	52392	52392	24027	12067	06558	45344	45328	67338	45328
11	43310	01081	01081	44863	80307	52555	10148	94617	89742	94617
12	61570	06360	06360	06173	63775	63148	95123	46903	35017	46903
13	31352	83799	83799	10779	18941	31579	76448	86919	62584	86919
14	57048	86526	86526	27795	93692	90529	56516	32204	35065	32204
15	09243	44200	44200	68721	07137	30729	75756	27660	09298	27660
16	97957	35018	35018	40894	88329	31520	82521	61587	22532	61587
17	93732	59570	59570	43781	98985	56671	66826	44509	95990	44509
18	72621	11225	11225	00922	68264	35666	59434	58107	71687	58107
19	61020	74418	74418	45371	20794	95917	37866	19378	99530	19378
20	97839	85474	85474	33055	91718	45473	54144	23000	99530	23000
21	89160	97192	97192	22232	90637	35055	45489	16301	88438	16301
22	25966	88220	88220	62871	92823	02823	52862	54883	84919	54883
23	81443	31719	31719	05049	54806	74690	07567	16543	65017	16543
24	11322	54931	54931	42362	34386	08024	97687	23215	46245	23215

جدول ۶-۸. جدول اعداد تصادفی (ادامه)

ردیف	اعداد ستونی									
	00000	00000	11111	11111	22222	22222	22222	33333	33333	33333
	01234	56789	01234	56789	01234	56789	01234	56789	01234	56789
00	64755	83885	84122	25920	17696	15655	95045	95947		
01	10302	52289	77430	34430	38112	49067	07348	23328		
02	71017	98495	51308	50374	66591	02887	53765	69149		
03	60012	55605	88410	34879	79655	90169	78900	03666		
04	37330	94656	49161	42802	48274	54755	44553	65090		
05	47869	87001	31591	12273	60626	12822	34691	61212		
06	38040	42737	64167	89578	39323	49324	88434	38706		
07	73508	30908	83054	80078	86669	30295	56460	45336		
08	32823	46474	84061	04324	20628	37319	32356	43069		
09	97501	99549	36630	35106	62069	92975	95320	57734		
10	74012	31955	59700	96982	66224	24015	96740	07589		
11	56754	26457	13351	05014	90966	33674	69096	33488		
12	49800	49908	54831	21948	08528	26372	92923	65026		
13	43584	89647	24878	56670	00221	50193	99591	62377		
14	16653	79664	60325	71301	35742	83636	73058	87229		
15	48502	69055	65322	58748	31446	40237	31252	96367		
16	98765	54692	36316	86230	48296	38352	23816	64091		
17	38923	61550	80357	81784	23444	12463	33992	28128		
18	77958	81604	25225	05587	51073	01070	60218	61961		
19	17928	28065	25586	08771	02641	85064	65796	48170		
20	94036	85978	02318	04499	41054	10531	87431	21596		
21	47460	60479	56230	48417	14372	85167	27558	00368		
22	47856	56088	51992	82439	40644	17170	13463	18286		
23	57616	34653	92298	62018	10375	76515	62986	90756		
24	08300	92704	66752	66610	57188	79107	51222	22013		

جدول A-۶ جدول اعداد تصادفی (ادامه)

ردیف	اعداد ستونی									
	00000	00000	11111	11111	56789	56789	22222	22222	33333	33333
00	89221	02362	65787	74733	51272	30213	92441	39651	01261	33333
01	04005	99818	63918	29032	94012	42363	01261	10050	53254	56789
02	88548	38066	50856	75045	40645	22841	53254	41125	49988	
03	41719	84401	59226	01314	54581	40398	49988	04579	85567	
04	28733	72489	00785	25843	24613	40797	85567	81471	68476	
05	65213	83927	77762	03086	80742	24395	68476	83792	69400	
06	65553	12678	90906	90466	43670	26217	69400	31205	45986	
07	05668	69080	73029	85746	58332	78231	45986	92998	47262	
08	39302	99718	49757	79519	27387	76373	47262	91612	18067	
09	64592	32254	45879	29431	38320	05981	18067	87137	82579	
10	07513	48792	47314	83660	68907	05390	82579	91582	56541	
11	86593	68501	56638	99800	82839	35148	56541	07232	32410	
12	83735	22599	97977	81248	36838	99560	32410	67014	56268	
13	08595	21826	54655	08204	87990	17033	56268	06384	35431	
14	41273	27149	44293	69458	16828	63962	15864	35431	47252	
15	00473	75908	56238	12242	72631	76314	47252	00347	07009	
16	86131	63789	81383	07868	89132	96182	07009	86132	22546	
17	33849	78359	08402	03580	03176	88663	22546	08018	20775	
18	61870	41657	07468	08612	98083	97349	20775	46091	91340	
19	43898	86923	25078	86129	78491	97653	91340	80786	40473	
20	29939	39123	04548	45985	60952	06041	40473	28726	67631	
21	38505	85555	14388	55077	18657	94887	67631	82268	53279	
22	91430	38431	67125	25511	72044	11562	53279	30753	02391	
23	03767	03767	13561	15597	06750	92552	02391	30753	01116	
24	38635	68976	25498	97528	96458	03805	01116	30651		

جدول ۶-۸ جدول اعداد تصادفی (ادامه)

ردیف	اعداد ستونی									
	00000 01234	11111 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789	33333 56789
00	02490	64122	54122	27944	39364	94239	72074	11679	54082	
01	11967	36469	60627	60627	83701	09253	30208	37482	37482	
02	48256	83465	49699	49699	24079	05403	35154	30613	03130	
03	27246	73080	21481	21481	23536	04881	89977	49484	93071	
04	32532	77285	72430	72430	70722	86529	18457	92657	10011	
05	66757	98955	92375	92375	93431	43204	55825	45443	60265	
06	11266	34545	76505	76505	97746	34668	26999	26742	97516	
07	17872	39142	45561	45561	80146	93137	48924	64257	50281	
08	62561	30365	03408	03408	14754	51798	08133	61010	97730	
09	62798	30779	35497	35497	70501	30105	08133	00997	01970	
10	75510	21771	04339	04339	33660	42757	62223	87565	48468	
11	87439	01691	63517	63517	26590	44437	07217	98706	39032	
12	97742	02621	10748	10748	78803	38337	65226	92149	59051	
13	98811	06001	21571	21571	78803	21828	83912	85188	61624	
14	51264	01852	64607	64607	92553	29004	26695	78583	62998	
15	40239	93376	10419	10419	68610	49120	02941	80035	99317	
16	26936	59186	51667	51667	27645	46329	44681	94190	66647	
17	88502	11716	98299	98299	40974	42394	62200	69094	81646	
18	63499	38693	25593	25593	61995	79867	80569	01023	38374	
19	93801	81206	03317	03317	78710	73828	31083	60509	44091	
20	29856	87120	22322	22322	57017	59334	30647	43061	26680	
21	97720	88265	56311	56311	50053	25365	81265	22414	02431	
22	85237	09105	74601	74601	13050	71017	15177	06957	92919	
23	75746	75268	31727	31727	46377	59938	15647	34177	92753	
24					95773	72364	87324	36879	06802	





جدول ٤-٨ جدول اعداد تصادفی (ادامه)

ردیف	اعداد ستونی									
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789	44444 01234	44444 56789
00	29935	06971	63175	52579	10478	89379	61428	21363		
01	15114	07126	51890	77787	75510	13103	42042	48111		
02	03970	43225	10589	87629	22039	94124	38127	65022		
03	79390	39188	40756	45269	65959	20640	14284	22060		
04	30035	06915	79196	54428	64819	52314	48721	81694		
05	29039	99861	28759	79802	68531	39198	38137	24373		
06	78196	08108	24107	49777	09599	43569	84820	94056		
07	15847	85493	91442	91351	80130	73752	21539	10086		
08	36614	62248	49194	97209	92587	92053	41021	84064		
09	40549	54884	91465	43862	35541	44466	88994	71180		
10	40878	08997	14286	09982	90308	78007	51587	10658		
11	10229	49282	41173	31468	69465	18756	08908	00660		
12	15918	76787	30624	25928	44124	25088	31137	71014		
13	13403	18796	74908	94404	04979	41462	18165	08335		
14	66523	94596	18796	90271	10009	08648	17040	04909		
15	91665	36409	06343	17870	25975	04602	21272	54020		
16	67415	87515	08207	73201	73201	57593	96917	00699		
17	76527	06996	23724	33448	63392	32304	60887	94017		
18	19815	47789	74348	17147	10954	34355	81194	54407		
19	25592	53587	76384	72575	84347	08918	05739	67222		
20	55902	45539	63646	31609	95099	82887	40060	04092		
21	02470	58376	79794	22482	42423	96162	47491	17204		
22	18630	53263	13319	35859	12350	12350	14032	87059		
23	89673	38230	16063	92007	59503	38402	76450	31333		
24	62986	67364	06595	17427	84623	14565	82860	57300		

Source: Kendall, M. G., and B. B. Smith (1938). Randomness and random sampling numbers. *Journal of the Royal Statistical Society*, 101, 164-166. Reproduced by permission of the Royal Statistical Society, London.

