آمار توصیفی و استنباطی

آمار توصیفی و استنباطی

فهرست مطالب

بخش اول: آمار توصیفی

بخش دوم: آمار استنباطی

بخش اول: آمار توصیفی

فصل اول : آمار توصیفی

کلمه «**statistics**» که به فارسی آن را «آمار» ترجمه کرده‌اند در اغلب زبان‌ها به دو معنی به کار می‌رود:

الف) به معنی ارقام و اعداد واقعی یا تقریبی درباره اموری از قبیل زاد و مرگ، طلاق و ازدواج، تصادفات رانندگی، میزان محصولات کشاورزی و صنعتی و ... .

ب) به معنی روش‌هایی برای جمع‌آوری، تنظیم و تجزیه و تحلیل اطلاعات عددی درباره موضوعی.

این دو مفهوم با هم ارتباط دارند، در این فصل بیشتر به قسمت «الف» که اغلب آمار توصیفی نامیده می‌شود خواهیم پرداخت. به‌طور کلی در بحث آمار توصیفی به سه قسمت عمدهٔ: 1ـ مفاهیم اولیه 2ـ مشخص‌کننده‌های مرکزی 3ـ مشخص‌کننده‌های پراکندگی 4ـ مشخص‌کننده‌های نسبی پراکندگی می‌پردازیم.

مفاهیم اولیه

علم آمار: به مجموعه‌ای از روش‌ها و مراحل مختلف که برای جمع‌آوری اطلاعات اولیه، دسته‌بندی داده‌ها و تجزیه و تحلیل آن‌ها و در نهایت تفسیر آن‌ها به‌کار می‌رود علم آمار می‌گوییم. علم آمار به دو بخش تقسیم می‌شود:

1ـ **آمار توصیفی**: قسمتی از روش‌های آماری است که شامل جمع‌آوری اطلاعات، دسته‌بندی آن‌ها و در انتها نمایش این داده‌ها می‌باشد.

2ـ **آمار استنباطی**: قسمتی از روش‌های آماری است که در آن اطلاعات به‌دست آمده از آمار توصیفی را مورد تجزیه و تحلیل قرار می‌دهند و نتایج حاصل از آن را به کل یا قسمتی از جامعه تعمیم می‌دهند.

جمعیت (جامعه آماری): مجموعه تمام عناصری که حداقل دارای یک ویژگی مشترک هستند و در یک زمان مشخص یا موقعیت مناسب، مورد توجه قرار می‌گیرند را جمعیت یا جامعه آماری می‌گوییم.

جامعه دانشجویان رشته کامپیوتر یا مجموعه دانشجویان رشته حسابداری در مقطع کاردانی یا وزن تمامی نوزادانی که از این لحظه به بعد در بیمارستان‌های ایران متولد می‌شوند یا تمامی نقاط درون یک دایره و ... همگی معرف یک جامعه آماری می‌باشند.

جامعه آماری به دو دسته تقسیم می‌شود:

1ـ **جامعه متناهی**: جامعه‌ای که تعداد عناصر آن محدود باشد جامعه متناهی نامیده می‌شود.

2ـ **جامعه نامتناهی**: جامعه‌ای که تعداد عناصر آن نامحدود باشد جامعه نامتناهی نامیده مـی‌شـود و خـود بـه دو دسـته: جامعه نامتناهی شمارش‌پذیر و جامعه نامتناهی شمارش‌ناپذیر تقسیم می‌شود.

تذکر: اگر جامعه‌ای متناهی باشد، تعداد عناصر آن را اندازه یا حجم جامعه می‌گوییم و با حرف N نشان می‌دهیم.

مجموعه دانشجویان رشته‌های حسابداری و کامپیوتر یک جامعه متناهی است، مجموعه وزن تمامی نـوزادانی کـه از ایـن لحظه به بعد در بیمارستان‌های ایران متولد می‌شوند یک جامعه نامتناهی شمارش‌پذیر و مجموعه تمامی نقاط درون یک دایره یک جامعه نامتناهی شمارش‌ناپذیر می‌باشد.

نمونه: بخشی از جامعه آماری را نمونه می‌گوییم یـا بـه بیـان دیگـر هـر زیرمجموعـه‌ای از جامعـه آمـاری را یـک نمونـه می‌گوییم. نمونه به دو دسته تقسیم می‌شود:

1ـ **نمونه ساده**: نمونه‌ای است که شخص آمارگر در انتخاب آن میل و سلقیه خود را اعمال کرده است.

2ـ **نمونه تصادفی**: نمونه‌ای است که میل و سلیقه شخص آمارگر در انتخاب آن دخالت ندارد.

تذکر:تمامی عناصریک جامعه آماری برای انتخـاب شـدن بـه عنـوان عضـوی از نمونـه تصـادفی دارای شـانس یکسـانی می‌باشند.

تذکر:در علم آمار نمونه‌های ساده فاقد ارزش می‌باشند و در این کتاب منظور از نمونه همان نمونه‌های تصادفی است. در صورت متناهی بودن نمونه، تعداد عناصر نمونه را اندازه یا حجم نمونه گویند و با حرف n نمایش می‌دهند.

مثال: برای خرید سیب به میوه‌فروشی مراجعه می‌کنیم صاحب مغازه تعدادی از ایـن سـیب‌هـا را در اختیـار مـا قـرار می‌دهد، با بررسی آن‌ها متوجه می‌شویم که نوع آن مرغوب بوده و در نتیجـه تصـمیم مـی‌گیـریم تـا یـک جعبـه از ایـن سیب‌ها را خریداری کنیم. چنانچه صاحب مغازه به منظور فریب دادن مشتری تمام سیب‌های روی جعبه را با سیب‌هـای مرغوب پوشانده باشد و بقیه سیب‌ها نامرغوب باشند در این‌صورت استنباط ما مبنـی بـر مرغـوب بـودن تمـام سـیب‌هـا نادرست است. در این مثال انتخاب اول یک نمونه تصادفی را در اختیار ما قرار می‌دهد ولی انتخاب دوم به دلیـل اعمـال میل و سلیقه صاحب مغازه یک نمونه ساده می‌باشد.

صفت: به کمیت یا کیفیتی که متعلق به عناصر جامعه آماری است صفت گفته می‌شود. در حقیقت صفت ویژگـی عناصـر یک جامعه آماری را بیان می‌کند و به دو دسته تقسیم می‌شود:

1ـ صفت ثابت (صفت مشخصه): صفتی است که بین همه عناصر یک جامعه آماری مشترک است.

2ـ صفت متغیر (صفت آماری): صفتی است که از فردی به فرد دیگر در بین عناصر جامعه آماری می‌تواند تغییر کند و به اختصار به آن متغیر می‌گوییم و با حرف **t** نمایش می‌دهیم.

مثال: اگر جامعه آماری را جامعه دانشجویان مقطع کاردانی در نظر بگیریم می‌توان «دانشجو بـودن» را صـفت مشخصـه آن دانست و صفاتی از قبیل: قد، وزن، رنگ مو یا هوش دانشجویان مقطع کاردانی را صفت متغیر یا متغیر به حساب آورد.

تذکر: متغیرها به دو دسته کمی و کیفی تقسیم می‌شوند. متغیرهای کمی آن‌هایی هسـتند کـه برایشـان ابـزار سـنجش وجود دارد مانند: قد، وزن، سن، مساحت، حجم، درجه حرارت و ... . متغیرهای کیفی آن‌هایی هستند کـه برایشـان ابـزار سنجش وجود ندارد مانند: رنگ چشم، رنگ مو، مرغوبیت، مهارت، هوش و ... .

تذکر: صفات متغیر کمی خود به دو دسته پیوسته و گسسته تقسیم می‌شوند. اگر مقادیر متغیر کمی قابل شمارش باشـد به آن متغیر کمی گسسته می‌گوییم و اگر مقادیر متغیر کمی غیرقابل شمارش باشد به آن متغیر کمی پیوسته می‌گوییم. به عنوان مثال به صفاتی از قبیل تعداد دانشجویان یک کلاس، تعداد اتومبیل‌های شهر تهران، متغیر کمی گسسـته و بـه صفاتی از قبیل طول، وزن، قد، مساحت، زمان، متغیر کمی پیوسته می‌گوییم.

مقیاس‌سازی: اختصاص دادن عدد به متغیرها را مقیاس‌سازی گویند، در حقیقت مـی‌خـواهیم عـدد حقیقـی **x** را تحـت قاعده خاص **F** به متغیر **t** نسبت دهیم یعنی $x = F(t)$ به عنوان مثال می‌توان فرض کرد که دمـای اتـاق متغیـر مـورد نظر باشد، در این صورت عدد **x** توسط دماسنج **F** به ویژگی دما اختصاص می‌یابد. برحسب این‌که قاعده **F** چگونه باشـد می‌توان چهار مقیاس گوناگون به‌دست آورد.

1ـ **مقایس اسمی:** از $x = F(t)$ یک مقیاس اسمی به‌دست می‌آید، هرگاه عدد **x** تنها برای تشخیص اسم یـا نـام اسـتفاده شود. این مقیاس به کد نیز شهرت دارد. البته قابل ذکر است که از این مقیاس برای مقایسه (بزرگتـر، کـوچکتر یـا چنـد برابر) یا چهار عمل اصلی حساب نمی‌توان استفاده کرد. به عنوان مثال بـه اتومبیـل‌هـای موجـود در کشـور یـک شـماره اختصاص یافته است، اگر شماره خاصی را در نظر بگیرید این شماره فقط بیانگر این است که اتومبیل مـورد نظـر از چـه نوعی، چه رنگی و ... می‌باشد.

2ـ **مقیاس ترتیبی:** از $x = F(t)$ یک مقیاس ترتیبی به‌دست می‌آید، اگر شـدت و ضـعف متغیـر **t** در **x** مـنعکس شـود، می‌توان **x** را برای مقایسه به کار برد ولی نمی‌توان روی آن چهار عمل اصلی حساب را انجام داد.

به عنوان مثال زمانی که مرغوبیت کالایی را با درجه 1، 2 و 3 طبقه‌بندی می‌کنیم مشخص می‌شود که در این‌جا 1 به معنی بهتر از 2 و 2 بهتر از 3 است. اما 1 به معنی دو برابر 2 یا نصف 3 نمی‌باشد.

3ـ مقایس فاصله‌ای: از $x = F(t)$ یک مقیاس فاصله‌ای به‌دست می‌آید اگر این تابع به‌صورت خطی $x = at + b$ با فرض $a > 0$ باشد. که در آن عرض از مبدأ مخالف صفر است $(b \neq 0)$. این مقیاس دارای سه ویژگی زیر است:

الف) صفر به معنی هیچ نیست.

ب) نسبت حفظ نمی‌شود.

ج) نسبت فاصله‌ها حفظ می‌شود.

به عنوان مثال می‌توان فرض کرد که **t** دمای سه جسم A_1 و A_2 و A_3 برحسب سانتی‌گراد به ترتیب برابر 0، 10 و 30 است، میزان گرمای این سه جسم با درجه فارنهایت طبق رابطه $x = \frac{9}{5}c + 32$ به ترتیب برابر است با: 32، 50 و 86. با مقایسه این دو مقیاس فاصله‌ای داریم:

اولاً: با مقایس اول میزان گرمای A_1 صفر و با مقیاس دوم 32 است بنابراین در اندازه‌گیری با مقیاس فاصله‌ای صفر به معنی «هیچ» نمی‌باشد و صرفاً جنبه قراردادی دارد.

ثانیاً: با مقایس اول میزان گرمای A_3 سه برابر میزان گرمای A_2 است در حالی‌که در مقیاس دوم چنین نیست.

ثالثاً: در مقیاس اول و هم در مقیاس دوم اختلاف میزان گرمای A_2 با A_3 دو برابر اختلاف میزان گرمای A_1 با A_2 است.

4ـ مقیاس نسبتی: از $x = F(t)$ یک مقیاس نسبتی به‌دست می‌آید اگر این تابع به‌صورت $x = at \quad (a > 0)$ باشد. در این مقیاس صفر به معنی «هیچ» است، نسبت حفظ می‌شود و نسبت فاصله‌ها نیز محفوظ می‌باشد. به‌عنوان مثال فرض کنید وزن سه جسم B_1، B_2 و B_3 به ترتیب برحسب کیلوگرم، 15، 20 و 30 می‌باشد، وزن همین سه جسم برحسب تن، (کیلوگرم * 1000) برابر است با: 15000، 20000 و 30000. ملاحظه می‌شود که در هر دو مقیاس نسبت وزن A_3 به A_1 برابر 2 می‌باشد.

تذکر: مقیاس‌های اسمی و ترتیبی عمدتاً برای متغیرهای کیفی و مقیاس‌های فاصله‌ای و نسبتی برای متغیرهای کمی مورد استفاده قرار می‌گیرند.

پارامتر: هر ویژگی عددی یک جامعه آماری پارامتر نامیده می‌شود مانند: میانگین جامعه (m) واریانس جامعه $\left(s^2\right)$ میانه (Md).

داده‌های آماری: فرض کنید می‌خواهیم صفت **t** از یک جمعیت را که معمولاً یک متغیر است مطالعه کنیم، اگر این متغیر را در مورد یکایک افراد جامعه آماری یا نمونه‌ای از آن با مقیاس مناسب اندازه‌گیری کنیم، یک مجموعه از اعداد **x** به‌دست می‌آیند که آن را داده می‌نامند. یا به بیان دیگر اعداد حاصل از اندازه‌گیری صفت مورد نظر در یک نمونه آماری را داده‌های آماری می‌نامند. از آن‌جا که توابع به دو دسته گسسته و پیوسته تقسیم می‌شوند، داده‌ها را نیز به دو دسته تقسیم می‌کنیم: داده‌های گسسته و داده‌های پیوسته. روش ساده برای تشخیص سریع داده‌های گسسته و پیوسته چنین است که اگر بین دو دادهٔ نزدیک به هم و قابل تصور هیچ عددی را نتوان تصور و تجسم کرد، آنگاه داده‌ها را گسسته گوییم، مانند تعداد فرزندان خانواده، و اما اگر بین دو مقدار نزدیک به هم و قابل تصور همواره بتوان عدد دیگری را تصور کرد آنگاه داده‌ها را پیوسته گوییم مانند: طول قامت انسان‌ها.

گرد کردن داده‌های پیوسته: داده‌های پیوسته یک عدد حقیقی است که به‌صورت اعشاری با تعدادی ارقام اعشار بیان می‌شود. منظور از گردن کردن عدد حقیقی x، تا **r** رقم اعشار، به‌دست آوردن یک عدد اعشاری x' با حداکثر **r** رقم اعشار است به طوری‌که داشته باشیم:

$$\left| x - x' \right| \leq 0/5 \times 10^{-r}$$

قاعده کلی برای گرد کردن یک عدد حقیقی تا r رقم اعشار: برای گرد کردن یک عدد حقیقی تا **r** رقم اعشار ابتدا نصف تقریب را به عدد اضافه کرده و سپس **r** رقم اعشار را نگه می‌داریم و بقیه ارقام را صفر می‌کنیم.

مثال: عدد 6512/90567 را با تقریب‌های 100، 1، 0/1، 0/01 و 0/001 گرد کنید.

الف) $\frac{100}{2} = 50 \Rightarrow 6212/90567 + 50 = 6262/90567 \rightarrow 6200$

ب) $\frac{1}{2} = 0/50 \Rightarrow 6212/90567 + 0/5 = 6213/40567 \rightarrow 6213$

ج) $\frac{0/1}{2} = 0/05 \rightarrow 6212/90567 + 0/05 = 6212/95567 \rightarrow 6212/9$

د) $\frac{0/01}{2} = 0/005 \rightarrow 6212/90567 + 0/005 = 6212/91067 \rightarrow 6212/91$

م) $\frac{0/001}{2} = 0/0005 \rightarrow 6212/90567 + 0/00005 = 6212/90617 \rightarrow 6212/906$

مثال: عدد 83/234113 را تا پنج رقم اعشار گرد کنید.

منظور از گرد کردن تا 5 رقم اعشار همان تقریب 10^{-5} یا 0/00001 می‌باشد پس:

$$\frac{0/00001}{2}=0/000005 \quad \Rightarrow \quad 83/234113+0/000005=83/234118 \quad \rightarrow \quad 83/23411$$

دسته‌بندی داده‌ها و جدول توزیع فراوانی

در آمار توصیفی بعد از جمع‌آوری داده‌ها، دسته‌بندی داده‌ها در یک جدول انجام می‌گیرد. برای انجام این کار ابتدا نیاز است تا تعاریف زیر را مطرح کنیم:

فراوانی: تعداد دفعات تکرار یک شیء و یا عدد را فراوانی آن شیء یا عدد می‌گوییم و با f_i نشان می‌دهیم.

جدول توزیع فراوانی: متداولترین جدول آماری به جدول فراوانی معروف است که این جدول شامل دو ستون که یکی نشان‌دهنده گروه‌ها و دیگری نشان‌دهنده تعداد دفعات تکرار گروه (فراوانی) می‌باشد.

تذکر: در بررسی جدول توزیع فراوانی دو حالت کلی زیر را در نظر می‌گیریم:

حالت اول: تعداد داده‌های آماری کم باشد.

حالت دوم: تعداد داده‌های آماری زیاد باشد.

حالت اول:

اگر تعداد داده‌های آماری کم باشد برای دسته‌بندی آن‌ها یک جدول توزیع فراوانی رسم می‌کنیم که در ستون اول مقادیری از متغیر (x_i) را می‌نویسیم و در ستون دوم فراوانی مربوط به هر ستون (f_i) را می‌نویسیم.

مثال: تعداد قرص‌های استامینافون که 20 خانواده در عرض یک ماه مصرف می‌کنند به‌صورت زیر داده شده است جدول فراوانی آن را تشکیل دهید.

$$0,2,3,11,8,6,5,2,1,1,4,3,3,5,5,2,2,4,0$$

تعداد قرص‌ها (x_i)	0	1	2	3	4	5	6	8	11
فراوانی $(f_i(x_i))$	2	2	4	3	2	3	1	1	1

حالت دوم:

اگر تعداد داده‌های آماری زیاد باشد بایستی داده‌ها را دسته‌بندی کنیم و برای طبقات حدود قائل شویم که برای این منظور چند اصطلاح آماری مورد نیاز را تعریف می‌کنیم.

1ـ **دامنه تغییرات**: اختلاف بین بزرگترین و کوچکترین داده را دامنه تغییرات می‌گوییم و با **R** نشان می‌دهیم:

2ـ **تعداد دسته‌ها**: در یک جدول توزیع فراوانی تعداد دسته‌ها را با حرف **K** نشان می‌دهیم و انتخاب آن بستگی به دامنه تغییرات دارد. هرچه دامنه تغییرات بزرگتر باشد تعداد دسته‌ها یا طبقات نیز بیشتر می‌شود.

3ـ **طول دسته**: هر دسته از دو عدد تشکیل می‌شود، عدد کوچکتر را کران پایین و عدد بزرگتر را کران بالای آن دسته یا طبقه می‌نامیم. اختلاف بین کران پایین و کران بالای هر طبقه را طول دسته (فاصله طبقه) می‌گوییم و با حرف **C** نشان می‌دهیم.

نکته: طول دسته را می‌توان از رابطه مقابل نیز به‌دست آورد:

$$C = \left[\frac{R}{K}\right]$$

4ـ **مرکز دسته (نماینده طبقه)**: میانگین کران پایین و کران بالای یک دسته را مرکز دسته می‌گوییم که از رابطه زیر به‌دست می‌آید:

$$x_i = \frac{\text{کران بالا + کران پایین}}{2} = \frac{L_i + U_i}{2}$$

نکته: فراوانی مرکز دسته‌ها برابر تعداد اعضایی است که در آن دسته قرار گرفته‌اند.

نکته: تفاضل دو مرکز دسته متوالی برابر است با طول دسته.

مثال: داده‌های زیر تعداد تلفن‌های زده شده روزانه به اداره مخابرات در 30 روز یک ماه می‌باشد. جدول فراوانی این 30 داده را تشکیل دهید.

164 ،166 ،173 ،171 ،177 ،172 ،178 ،164 ،159 ،169 ،170 ،184 ،172 ،172 ،177 ،173

154 ،162 ،170 ،161 ،166 ،180 ،177 ،170 ،169 ،169 ،156 ،180 ،171 ،169

ابتدا داده‌ها را به ترتیب صعودی مرتب می‌کنیم:

154 ،156 ،159 ،161 ،162 ،164 ،164 ،166 ،166 ،169 ،169 ،169 ،169 ،170 ،170،170

171،171 ،172 ،172 ،172 ،173 ،173 ،177 ،177 ،177 ،178 ،180 ،180 ،184

سپس عبارات مورد نیاز زیر را محاسبه می‌کنیم:

دامنه تغییرات : $R = \max - \min = 184 - 154 = 30$

تعداد دسته‌ها : $K=5$ (دلخواه)

طول دسته : $C=\frac{R}{K}=\frac{30}{5}=6$

حال جدول توزیع فراوانی را به‌صورت زیر تشکیل می‌دهیم:

حدود طبقات	154_160	160_166	166_172	172_178	178_184
فراوانی	3	4	11	8	4

تست: **داده‌های آماری با ماکسیمم ۸۵ و می‌نیمم ۲۳ را در ۷ طبقه دسته‌بندی کرده‌ایم. حدود طبقه چهارم کدام است؟**

1) 49 ـ 54   2) 49 ـ 58   3) 50 ـ 58   4) 50 ـ 59

$\max = 85 \quad , \quad \min = 23 \quad , \quad k = 7$

$R = \max - \min = 85 - 23 = 62$

$C = \left\lceil \frac{R}{K} \right\rceil = \left\lceil \frac{62}{7} \right\rceil = \lceil 8/8 \rceil \Rightarrow C = 9$

بنابراین داریم:

(23_32) : دسته اول   (32_41) : دسته دوم

(41_50) : دسته سوم   (50_59) : دسته چهارم

پس گزینه چهارم صحیح است.

تست: **در جدول توزیع فراوانی داده‌ها، نماینده طبقات اول و دوم و آخر به ترتیب ۴۴، ۴۹ و ۸۴ است تعداد طبقات کدام است؟**

1) 7   2) 8   3) 9   4) 10

اگر تعداد طبقات $\mathbf{K}$ باشد می‌توان نشان داد:

$x_1 = 44 , x_2 = 49 , \mathbf{K} , x_k = 84$

طول دسته اول $C = x_2 - x_1 = 49 - 44 = 5$

حد پایین دسته اول $L_1 = x_1 - \frac{C}{2} = 44 - 2/5 = 41/5$

حد بالای دسته آخر $U_K = x_K + \frac{C}{2} = 84 + 2/5 = 86/5$

$R = 86/5 - 41/5 = 45$

$C = \left\lceil \frac{R}{K} \right\rceil \Rightarrow K = \left\lceil \frac{R}{C} \right\rceil = \left\lceil \frac{45}{5} \right\rceil \Rightarrow K = 9$

تست: بزرگترین داده در یک جدول فراوانی ۵۰، طول دسته‌ها ۶ و تعداد دسته‌ها ۵ بـوده، ارزش یـا انـدازه مشـترک داده‌های دسته اول با چه عددی بیان می‌شود؟

1) 23   2) 20   3) 22   4) 25

$x_{\max}=50 \quad , \quad C=6 \quad , \quad K=5$

$$C=\frac{R}{K} \quad \Rightarrow \quad R=C\cdot K=6\times5 \quad \Rightarrow \quad R=30$$

$$R=x_{\max}-x_{\min} \quad \Rightarrow \quad 30=50-x_{\min} \quad \Rightarrow \quad x_{\min}=20$$

دسته اول : $(20-26)$

مرکز دسته اول : $x_1=\frac{26+20}{2}=23$

تست: یک سری از داده‌های آماری را به K دسته تقسیم کرده‌ایم طول هر دسته ۲۴ شده است، اگر به تعـداد دسته‌ها یکی افزوده شود طول هر دسته ۱۸ می‌شود، K کدام است؟

1) 1   2) 2   3) 3   4) 4

$$C_1=24 \quad \Rightarrow \quad C_1=\frac{R}{K} \quad \Rightarrow \quad 24=\frac{R}{K} \quad \Rightarrow \quad R=24K$$

$$C_2=18 \quad \Rightarrow \quad C_2=\frac{R}{K+1} \quad \Rightarrow \quad 18=\frac{R}{K+1} \quad \Rightarrow \quad R=18(K+1)$$

با استفاده از دو رابطه فوق می‌توان نشان داد که:

$$24K=18(K+1) \quad \Rightarrow \quad 24K=18K+18 \quad \Rightarrow \quad 6K=18 \quad \Rightarrow \quad K=3$$

انواع فراوانی

1ـ **فراوانی مطلق**: تعداد عناصر هر دسته یا تعداد دفعات تکرار هر داده را فراوانی مطلق آن دسته یا آن داده می‌نامیم و آن را با F_i نمایش می‌دهیم.

نکته: مجموع فراوانی‌های مطلق با تعداد کل داده‌ها (حجم جامعه) برابر می‌باشد:

$$\sum_{i=1}^{K} f_i=f_1+f_2+\dots+F_K=N$$

2ـ فراوانی نسبی: نسبت فراوانی مطلق هر دسته یا هر داده به تعداد کل فراوانی‌ها (حجم جامعـه) را فراوانـی نسـبی آن داده می‌نامیم و آن را با f_{p_i} نمایش می‌دهیم:

$$f_{p_i} = \frac{f_i}{\sum_{i=1}^{K} f_i} = \frac{f_i}{N}$$

نکته: مجموع فراوانی‌های نسبی برابر یک است.

$$\sum_{i=1}^{K} f_{p_i} = f_{p_1} + f_{p_2} + \ldots + f_{p_K} = 1$$ (K تعداد طبقات یا دسته‌ها می‌باشد)

نکته: به کمک فراوانی نسبی می‌توان تراکم داده‌ها را در هر دسته مشخص نمود.

3ـ درصد فراوانی نسبی: درصد فراوانی نسبی هر طبقه عبارت است از حاصل‌ضرب فراوانی نسبی آن طبقه در عدد 100 که آن را با p_i نمایش می‌دهیم.

$$P_i = f_{p_i} \times 100$$

$$\sum_{i=1}^{K} P_i = P_1 + P_2 + \ldots + P_K = 100$$ (K تعداد طبقات یا تعداد دسته‌ها می‌باشد)

4ـ فراوانی تجمعی: مجموع فراوانی مطلق هر طبقه با فراوانی‌های مطلق طبقات قبلی را فراوانی تجمعی آن طبقه می‌نامیم و آن را با F_{ci} نمایش می‌دهیم.

نکته: فراوانی تجمعی طبقه آخر با حجم کل جامعه برابر است.

نکته: فراوانی تجمعی هر طبقه نسبت به فراوانی تجمعی قبل از خودش صعودی است.

5ـ فراوانی تجمعی نسبی: نسبت فراوانی تجمعی هر دسته به تعداد کل داده‌ها **(N)** را فراوانی تجمعی نسبی می‌گوییم و آن را با F_{cpi} نمایش می‌دهیم.

$$F_{cpi} = \frac{F_{ci}}{N}$$

6ـ درصد فراوانی تجمعی نسبی: عبارت است از حاصل‌ضرب فراوانی تجمعی نسبی هر طبقه در عدد 100 که آن را با P_{ci} نمایش می‌دهیم.

$$P_{ci} = F_{cpi} \times 100$$

مثال: وزن ۳۰ بسته گز اصفهان تا نزدیکترین کیلو به‌صورت زیر داده شده است. یک جدول فراوانی کامل با ۸ طبقه به طول‌های مساوی پیدا کنید.

158، 146، 157، 149، 125، 144، 132، 150، 164، 138

119، 163، 176، 138، 126، 168، 144، 152، 148، 136

135، 161، 135، 140، 153، 135، 147، 142، 173، 146

$R = \max - \min = 176 - 119 = 57 \qquad K = 8$

$C = \left\lceil \frac{R}{K} \right\rceil = \left\lceil \frac{57}{8} \right\rceil = 8$

اکنون نماینده هر طبقه (مرکز دسته) را مشخص می‌کنیم:

$x_1 = L_1 + \frac{C}{2} = 119 + \frac{8}{2} = 123 \qquad x_2 = L_2 + \frac{C}{2} = (119+8) + \frac{8}{2} = 131$

$x_3 = 131 + 8 = 139 \qquad x_4 = 139 + 8 = 147$

$x_5 = 147 + 8 = 155 \qquad x_6 = 155 + 8 = 163$

$x_7 = 163 + 8 = 171 \qquad x_8 = 171 + 8 = 179$

سپس جدول فراوانی را به‌صورت زیر تشکیل می‌دهیم.

حدود طبقات	x_i	f_i	f_{pi}	P_i	F_{ci}	F_{cpi}	P_{ci}
119_127	123	3	0/1	10	3	0/1	10
127_135	131	1	0/033	3/3	4	0/133	13/3
135_143	139	8	0/266	26/6	12	0/4	40
143_151	147	8	0/266	26/6	20	0/666	66/6
151_159	155	4	0/133	13/3	24	0/8	80
159_167	163	3	0/1	10	27	0/9	90
167_175	171	2	0/066	6/6	29	0/966	96/6
175_183	179	1	0/033	3/3	30	1	100
$\sum$	_	30	1	100	_	_	_

از جدول فوق می‌توان نتیجه گرفت که حدود 10 درصد بسته‌ها وزنی بین 159_167 کیلوگرم دارند حدود 6/6 درصدوزنی بین 167_175 کیلوگرم دارند (با استفاده از P_i) و می‌توان نشان داد که حدود 90 درصد بسته‌ها وزنی کمتر از 167 کیلوگرم و 80 درصد آن‌ها وزنی کمتر از 159 کیلوگرم دارند (با استفاده از P_{ci}).

مثال: تعداد 120 داده به‌صورت جدول توزیع فراوانی زیر مشخص شده‌اند، جدول فراوانی آن را تکمیل کنید.

حدود طبقات	1_3	4_6	7_9	10_12
فراوانی مطلق	30	15	45	30

حدود طبقات	x_i	f_i	f_{pi}	P_i	F_{ci}	F_{cpi}	P_{ci}
0/5_3/5	2	30	0/25	25	30	0/25	25
3/5_6/5	5	15	0/125	12/5	45	0/375	37/5
6/5_9/5	8	45	0/375	37/5	90	0/75	75
9/5_12/5	11	30	0/25	25	120	1	100
$\sum$	_	120	1	100	_	_	_

تست: اگر فراوانی نسبی طبقه‌ای ۰/۱۵ و فراوانی مطلق همین طبقه ۱۲ باشد فراوانی تجمعی طبقه آخر کدام است؟

1) 60   2) 70   3) 80   4) 90

از آن‌جا که فراوانی تجمعی طبقه آخر همان تعداد کل داده‌هاست پس می‌توان نوشت:

$$f_i = 12 \quad , \quad f_{pi} = 0/15$$

$$f_{pi} = \frac{f_i}{N} \Rightarrow N = \frac{f_i}{f_{pi}} = \frac{12}{0/15} \Rightarrow N = 80$$

تست: اگر فراوانی تجمعی طبقه پنجم ۳۲ و فراوانی مطلق آن ۸ باشد، فراوانی تجمعی طبقه چهارم کدام است؟

1) 24   2) 26   3) 36   4) 40

فراوانی تجمعی طبقه چهارم + فراوانی مطلق طبقه پنجم = فراوانی تجمعی طبقه پنجم

$$f_{c_5} = f_5 + f_{c_4} \Rightarrow 32 = 8 + f_{c_4} \Rightarrow f_{c_4} = 24$$

تست: در جدول زیر کران بالای طبقه سوم و فراوانی نسبی این طبقه کدام است؟

1) 49 و 0/2   2) 49/5 و 0/2   3) 39 و 7   4) 50 و 7

متغیر	23_31	32_40	41_49	50_58	59_67
فراوانی	3	6	7	11	8

از آن‌جا که کران پایین هر طبقه با کران بالای طبقه بعدی برابر نیست پس باید میانگین گرفت:

$$\text{کران بالای دسته سوم} = \frac{49+50}{2} = 49/5$$

و برای به‌دست آوردن فراوانی نسبی دسته سوم داریم:

$$f_{p_3} = \frac{f_3}{N} \Rightarrow f_{p_3} = \frac{7}{35} = 0/2$$

تست: در یک دسته‌بندی داده‌ها، فراوانی کل برابر ۱۲۰ و فراوانی‌های نسبی دسته اول و دوم به ترتیب ۰/۴ و ۰/۲ می‌باشد فراوانی تجمعی دسته دوم کدام است؟

24 (1   36 (2   48 (3   72 (4

$$N = \sum f_i = 120 \quad , \quad f_{p_1} = 0/4 \quad , \quad f_{p_2} = 0/2$$

$$f_{2p} = \frac{F_2}{N} \Rightarrow f_2 = f_{p_2} \times N \Rightarrow f_2 = 0/2 \times 120 = 24$$

$$f_{1p} = \frac{F_1}{N} \Rightarrow f_1 = f_{p_1} \times N \Rightarrow f_1 = 0/4 \times 120 = 48$$

$$f_{c_2} = f_1 + f_2 = 24 + 48 = 72$$

تست: یک سری از داده‌های آماری را به ۳ دسته تقسیم کرده‌ایم اگر فراوانی هر دسته به اندازه ۰/۲ از دسته قبل بیشتر باشد با شرط این‌که فراوانی دسته اول $\left(\frac{r}{3}-1\right)$ باشد، r کدام است؟

2/8 (1   3/4 (2   3 (3   3/2 (4

می‌دانیم که $\sum f_{pi} = 1$ پس خواهیم داشت:

$$\left(\frac{r}{3}-1\right) + \left(\frac{r}{3}-1+0/2\right) + \left(\frac{r}{3}-1+0/4\right) = 1$$

$$r - 3 + 0/6 = 1 \Rightarrow r = 3/4$$

تست: در ۱۲۰ داده آماری کوچکترین و بزرگترین مقادیر به ترتیب ۳۵ و ۵۷ می‌باشند، این داده‌ها در ۹ طبقه دسته‌بندی شده‌اند. اگر ۳۲ درصد داده‌ها کمتر از ۴۵ و همچنین ۴۷ درصد داده‌ها کمتر از ۴۷/۵ باشند، فراوانی مطلق دسته وسط کدام است؟

1) 12 2) 15 3) 16 4) 18

$\max = 57 \quad , \quad \min = 35 \quad , \quad N = 120 \quad , \quad K = 9$

$R = 57 - 35 = 22 \quad , \quad C = \frac{R}{K} = \frac{22}{9} = 2/5$

(35 و 37/5): دسته اول و ...... و (45 و 47/5): دسته وسط و ...... و (55 و 57/5) دسته آخر

با استفاده از فرضیات مسئله می‌توان فهمید که:

درصد فراوانی نسبی طبقه وسط $= 47 - 32 = 15$

فراوانی مطلق طبقه وسط $= \frac{15 \times 120}{100} \Rightarrow f = 18$

تست: ۴۰ داده آماری در یک جدول توزیع فراوانی با ۸ دسته طبقه‌بندی شده است، اگر مجموع فراوانی‌های نسبی تا طبقه هفتم برابر ۰/۷۵ باشد آنگاه فراوانی طبقه هشتم کدام است؟

1) 12 2) 11 3) 10 4) 9

$$\sum_{i=1}^{8} f_{pi} = 1 \Rightarrow f_{p_1} + \ldots + f_{p_7} + f_{p_8} = 1 \Rightarrow 0/75 + f_{p_8} = 1 \Rightarrow f_{p_8} = 0/25$$

$$f_{p_8} = \frac{f_8}{N} \Rightarrow f_8 = 0/25 \times 40 = 10$$

نمودارهای فراوانی و تحلیل داده‌ها

گاهی برای ارائه یک تصویر روشن از ماهیت داده‌های مورد اندازه‌گیری، این داده‌ها و مقادیر فراوانی آن‌ها را به صورت نمودار ارائه می‌کنیم. به بیان دیگر، داده‌های موجود در جدول توزیع فراوانی را به‌وسیله ترسیم نمودار به بیننده منتقل می‌کنیم. برخی از نمودارها عبارتند از: 1ـ نمودار میله‌ای 2ـ نمودار سوزنی 3ـ نمودار هیستوگرام (بافت نگار) 4ـ نمودار چند ضلعی یا چند بر فراوانی 5ـ نمودار فراوانی تجمعی 6ـ نمودار دایره‌ای 7ـ نمودار ساقه و برگ.

1ـ **نمودار میله‌ای**: برای رسم این نمودار روی محور **x**ها متغیرها (مراکز دسته) و روی محور **y**ها فراوانی مطلق (فراوانی نسبی) همان دسته را قرار می‌دهیم. و سپس روی هر مقدار x_i میله‌ای به ارتفاع فراوانی مطلق (نسبی) رسم می‌کنیم.

تذکر: این نمودار بیشتر برای متغیرهای «کیفی» و «کمی گسسته» مناسب است.

2ـ **نمودار سوزنی**: برای رسم این نمودار مانند نمودار میله‌ای عمل می‌کنیم فقط با این تفاوت که از میله استفاده نمی‌کنیم بلکه نقاط (x_i , f_i) را مشخص می‌کنیم.

3ـ **نمودار هیستوگرام**: برای رسم این نمودار روی محور **x**ها حدود طبقات (حدود دسته‌ها) و روی محور **y**ها فراوانی (مطلق یا نسبی) دسته‌ها را مشخص می‌کنیم. در حقیقت این نمودارها شامل مستطیل‌هایی می‌باشند که در ارتباط با آن‌ها می‌توان به نکات زیر اشاره کرد:

تذکر: نمودار هیستوگرام برای متغیرهای کمی پیوسته مناسب است.

تذکر: به‌طور کلی نمودار هیستوگرام نمایشی از داده‌های دسته‌بندی شده است که در آن سطح مستطیل‌ها متناسب با فراوانی دسته‌ها می‌باشد. در صورتی‌که طول دسته‌ها (قاعده مستطیل‌ها) در نمودار مستطیلی با یکدیگر برابر باشند ارتفاع مستطیل‌ها (فراوانی دسته‌ها) با یکدیگر مقایسه می‌شوند و اگر طول دسته‌ها متفاوت باشند، مساحت مستطیل‌ها با یکدیگر مقایسه می‌شوند، هرچه مساحت یک مستطیل بیشتر باشد تعداد نفراتی که در آن دسته قرار دارند بیشتر است.

4ـ **نمودار چند ضلعی**: در این نمودار نقاط (x_i , f_i) را مشخص می‌کنیم، با به هم وصل این نقاط به‌وسیله خط‌های شکسته یک چند ضلعی به‌وجود می‌آید که به آن نمودار چندضلعی یا چندبرفراوانی می‌گوییم.

تذکر: در نمودار چند ضلعی دو دسته با فراوانی‌های صفر به ابتدا و انتهای دسته اضافه می‌کنند (فراوانی مرکز دسته قبل از اولین دسته و مرکز دسته بعد از آخرین دسته را صفر می‌گذاریم) و منظور از آن این است که مساحت زیر چند ضلعی فراوانی برابر مساحت نمودار هیستوگرام است.

تذکر: نمودار چندضلعی برای داده‌های کمی پیوسته مناسب است.

تذکر: نمودار چندضلعی فراوانی را می‌توان با در دست داشتن فراوانی نسبی نیز رسم کرد که آن را چند ضلعی فراوانی نسبی می‌گوییم، در این‌صورت اطلاعات منسجم‌تری در اختیار ما قرار می‌گیرد چون می‌توان فراوانی را با کل جامعه مقایسه کرد.

5ـ نمودار فراوانی تجمعی: برای رسم این نمودار روی محور **x**ها کران‌های هر دسته (حدود دسته‌ها) قرار می‌گیرد و روی محور **y**ها فراوانی تجمعی (مطلق یا نسبی) قرار می‌گیرد با وصل کردن این نقاط نمودار را رسم می‌کنیم.

تذکر: مقادیر فراوانی‌های تجمعی را به کران بالای هر دسته اختصاص می‌دهیم.

تذکر: نمودار فراوانی تجمعی هیچ‌گاه نزولی نمی‌شود.

تذکر: شیب منحنی در هر دسته، با فراوانی مربوط به همان دسته ارتباط مستقیم دارد.

6ـ نمودار دایره‌ای: برای رسم این نمودار، دایره‌ای رسم می‌کنیم و مساحت دایره را به قطاع‌هایی تقسیم می‌کنیم که سطح هر قطاع متناسب با مقدار یا فراوانی متغیر مورد نظر است. در این نمودار باید زاویه‌های متناسب با فراوانی دسته‌ها را پیدا کرد، برای این کار اگر فراوانی دسته‌ای برابر F_i باشد زاویه نظیر آن از رابطه زیر به‌دست می‌آید:

$$a_i = \frac{f_i}{N} \times 360$$

تذکر: نمودار دایره‌ای بر مبنای فراوانی نسبی$\left(\frac{f_i}{N}\right)$ رسم می‌شود.

تذکر: نمودار دایره‌ای برای نمایش متغیرهای کیفی مناسب است.

7ـ نمودار ساقه و برگ: داده‌های آماری معمولاً به‌صورت اعداد می‌باشند و از این اعداد می‌توان نموداری تشکیل داد که نمودار ساقه و برگ نامیده می‌شود، در واقع نمودار ساقه و برگ نموداری است که اعداد، تشکیل‌دهنده آن می‌باشند. این نمودار برای داده‌هایی که تفاوت کوچکترین داده و بزرگترین داده آن از نظر تعداد ارقام کم باشد مناسب است.

مثال: نمودار ساقه و برگ داده‌های زیر را تشکیل دهید.

10 ،11 ،15 ،23 ،27 ،28 ،38 ،38 ،39 ،39

40 ،41 ،44 ،45 ،46 ،46 ،52 ،58 ،65

برای تهیه نمودار ساقه و برگ ارقام داده شده را به دو بخش تقسیم می‌کنیم. ساقه شامل یک یا چند رقم اولیه و برگ که با فاصله کمی جلوتر از ساقه به شکل صعودی قرار می‌گیرد شامل ارقام باقی‌مانده اعداد است. بنابراین در این‌جا ارقام اول داده‌ها یعنی 1، 2، 3، 4، 5 و 6 را به‌عنوان ساقه و بقیه داده‌ها را به عنوان شاخه (برگ) در نظر می‌گیریم.

ساقه	برگ					
1	0	1	5			
2	3	7	8			
3	8	8	9	9		
4	0	1	4	5	6	6
5	2	8				
6	5					

مثال: نمودار ساقه و برگ داده‌های زیر را تشکیل دهید.

36/3 ،41/0 ،36/9 ،37/1 ،44/9 ،36/8 ،30/0 ،37/2 ،42/1

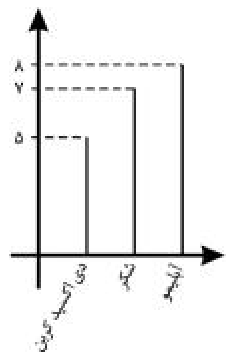
32/7 ،37/3 ،41/2 ،36/6 ،36/5 ،33/2 ،40/5 ،33/6 ،40/2

37/7 ،37/7 ،37/5 ،33/6 ،40/0 ،35/9 ،38/3 ،35/6 ،38/6

ابتدا دو رقم دهگان و یکان را به‌عنوان ساقه و ارقام بعد از ممیز را به‌عنوان برگ برمی‌گزینیم.

ساقه	برگ					
30	0/0					
32	0/7					
33	0/2	0/6	0/6			
35	0/6	0/9				
36	0/3	0/5	0/6	0/8	0/9	
37	0/1	0/2	0/3	0/5	0/7	0/7
38	0/3	0/6				
40	0/0	0/2	0/5			
41	0/0	0/2				
42	0/1					
44	0/9					

تست: **نمودار میله‌ای مقابل مربوط به ترکیبات یک نوع نوشیدنی است درصد فراوانی نسبی شکر چقدر است؟**



1) 8   2) 35

3) 40   4) 25

$$\text{درصد فراوانی نسبی شکر} = \frac{\text{فراوانی شکر}}{\text{فراوانی کل}} \times 100 = \frac{7}{5+7+8} \times 100 = 35$$

تست: اگر از فراوانی‌های مطلق، یک هیستوگرام کشیده شده باشد، سطح زیر هیستوگرام برابر با مجموع کدام فراوانی است؟

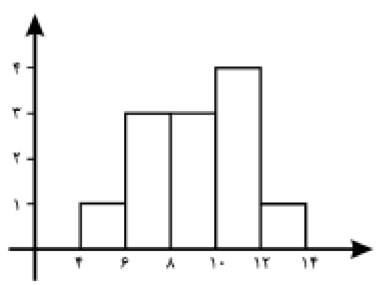
1) نسبی   2) تجمعی نسبی   3) تجمعی   4) مطلق

اگر برای رسم یک هیستوگرام از فراوانی‌های مطلق استفاده کنیم مساحت کل مستطیل‌ها با مجموع فراوانی‌های مطلق برابر می‌شود پس گزینه چهارم صحیح است.

تست: با توجه به نمودار مقابل فراوانی نسبی دسته سوم چه‌قدر است؟

1) 0/3   2) 0/25

3) 0/15   4) 0/2

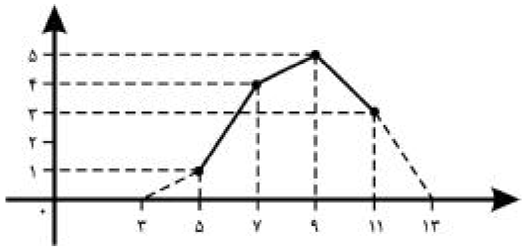


در هیستوگرام مقابل چون طول دسته‌ها یکسان است پس با توجه به ارتفاع مستطیل‌ها داریم:

$$N = \sum f_i = 1+3+4+1 = 12$$

$$\text{فراوانی نسبی دسته سوم} = \frac{3}{12} = 0/25$$

تست: با توجه به نمودار چند ضلعی مقابل، سطح زیر نمودار هیستوگرام متناظر با آن کدام است؟



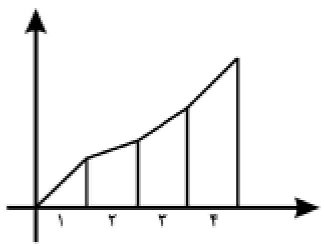
1) 22   2) 25

3) 26   4) 27

می‌دانیم که سطح زیر نمودار یک چندضلعی فراوانی، با مساحت نمودار هیستوگرام (مستطیلی) برابر است پس:

$$S = (2\times1) + (2\times4) + (2\times5) + (2\times3) = 26$$

تست: **نمودار تجمعی یک جدول فراوانی با چهار طبقه به‌صورت زیر است، کدام طبقه کمترین فراوانی مطلق را دارد؟**

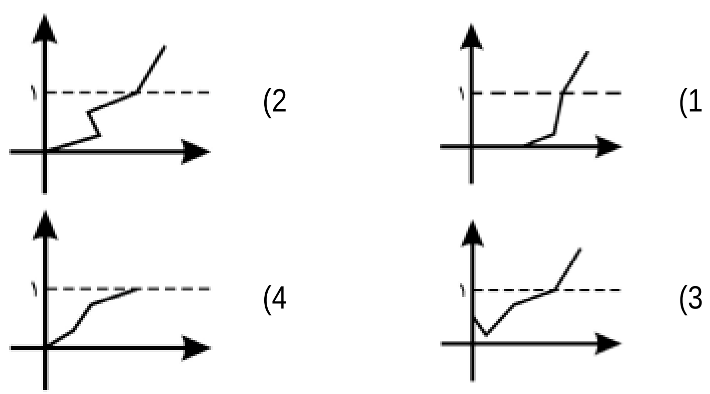


1) 2   2) 1

3) 4   4) 3

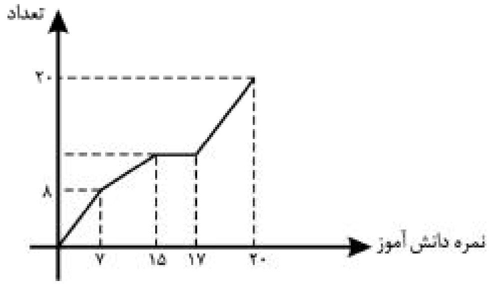
همان‌طور که اشاره شد، شیب خط نمودار فراوانی تجمعی در هر دسته با فراوانی همان دسته رابطه مستقیم دارد. چون شیب نمودار در طبقه دوم از همه شیب‌های دیگر کمتر است پس این طبقه کمترین فراوانی مطلق را دارد بنابراین گزینه اول صحیح است.

تست: **کدام‌یک از نمودارهای زیرنمودار فراوانی تجمعی نسبی است؟**



نمودار فراوانی تجمعی نسبی همواره صعودی بوده و سقف نمودار آن از 1 بیشتر نمی‌شود پس گزینه چهارم صحیح است.

تست: **با توجه به نمودار مقابل کدام از گزینه‌های زیر درست است؟**



1) 8 نفر نمره 6 گرفته‌اند.

2) 2 نفر نمره کمتر از 10 و بیشتر از 8/5 گرفته‌اند.

3) هیچ‌کس نمره بین 15 و 17 نگرفته است.

4) 15 نفر نمره‌ای کمتر از 15 گرفته‌اند.

از آن‌جا که شیب خط نمودار فراوانی تجمعی فوق بین فراوان‌های 15 تا 17 برابر صفر است بنابراین هیچ‌کس نمره‌ای بین 15 تا 17 نگرفته است و گزینه سوم صحیح است.

تست: **در انتخابات یک شهر ۵۴۰۰۰۰ نفر شرکت کرده‌اند اگر آنان را به ۵ گروه سنی تقسیم نموده و با نمودار دایره‌ای نشان دهیم، درصد شرکت‌کنندگان در یک گروه سنی با زاویه قطاع ۶۳ درجه نشان داده می‌شود، این تعداد کدام است؟**

1) 94500   2) 94800   3) 95000   4) 94000

$$a_i = \frac{f_i}{N} \times 360 \Rightarrow 63 = \frac{f_i}{540000} \times 360 \Rightarrow f_i = 94500$$

تست: **در نمودار دایره‌ای جدول زیر، زاویه مرکزی دسته به نمایندگی ۱۰، برابر 72^0 شده، چند درصد داده‌ها از ۱۱/۵ کوچکتر می‌باشند؟**

1) %45

2) %60

3) %40

4) %80

دسته‌ها	2/5_5/5	5/5_8/5	8/5_11/5	11/5_14/5
فراوانی‌ها	8	10	x	6

$$N = \sum_{i=1}^{4} f_i = 8 + 10 + x + 6 \Rightarrow N = 24 + x$$

$$a_i = \frac{f_i}{N} \times 360 \Rightarrow 72^0 = \frac{x}{24+x} \times 360 \Rightarrow 24 + x = 5x \Rightarrow x = 6$$

$$\Rightarrow N = 24 + 6 = 30$$

تعداد داده‌های کمتر از 11/5 برابر $8+10+6=24$ می‌شود بنابراین درصد داده‌های کمتر از 11/5 برابر است با:

$$\frac{24}{30} \times 100 = \%80$$

تست: توزیع گروه‌های خونی تعدادی از افراد به‌صورت

A	B	AB	O
24	14	10	12

است درصد مساحت مربوط به گروه خونی O در نمودار دایره‌ای کدام است؟

15 (1   20 (2   25 (3   40 (4

$$N = 24 + 14 + 10 + 12 \Rightarrow N = 60$$

$$S_i = \frac{f_i}{N} \times 100 \Rightarrow S_o = \frac{12}{60} \times 100 \Rightarrow S_o = \%20$$

تست: برای رسم داده‌هایی که تفاوت کوچکترین و بزرگترین داده آن‌ها از نظر تعداد ارقام کم باشد، کدام نمودار مناسب است؟

1) دایره‌ای   2) ساقه و برگ   3) میله‌ای   4) چندبرفراوانی

گزینه 2 صحیح است.

تست: در نمودار ساقه و برگ، ۵۰ داده بین ۱۰ و ۲۰ می‌باشد که برخی داده‌ها شامل یک رقم اعشار بوده، برگ‌های مربوط به اتصال ۱۴ روی این ساقه به‌صورت ۰۳۵۵۵۵۶۷۷۹ ۱۴ نمایش داده شده «چند درصد داده‌ها» درست عدد ۱۴/۵ بوده‌اند؟

4 (1   8 (2   10 (3   20 (4

با توجه به داده‌های مسئله می‌توان فهمید که تعداد عدد 14/5، 4تا می‌باشد و از طرفی تعداد کل داده‌ها برابر 50 است بنابراین داریم:

$$\text{درصد عدد } 14/5 = \frac{4}{50} \times 100 = 8$$

مشخص‌کننده‌های مرکزی

در هنگامی که داده‌ها طبقه‌بندی می‌شوند هر داده شخصیت فردی خود را از دست می‌دهد و شخصیت گروهی پیدا می‌کند و نماینده هر گروه بازگوکننده کلید داده‌های آن گروه می‌باشد. مشخص‌کننده‌های مرکزی (شاخص‌های مرکزی) معیارهایی هستند که محل تمرکز داده‌ها را معرفی می‌کنند. بنابراین برای محاسبه مشخص‌کننده‌های مرکزی مراکز داده‌ها به عنوان نماینده داده‌هایی که در هر دسته قرار دارند معرفی می‌شوند. مهمترین مشخص‌کننده‌های مرکزی عبارتند از: میانگین، میانه، و چندکها.

میانگین

مشهورترین مشخص‌کننده مرکزی میانگین است. میانگین انـواع گونـاگون دارد کـه شـش نـوع آن عبارتنـد از: میـانگین حسابی، میانگین وزنی (موزون)، میانگین هندسی، میانگین توافقی یا هارمونیک (همساز)، میانگین درجـه دوم، میـانگین درجه **r** ام. در آمار عمدتاً با میانگین حسابی سر و کار داریم و در بحث‌های پس از این بخش هر کجا از میانگین صحبت می‌شود منظور میانگین حسابی است.

۱ـ میانگین حسابی

تعدادی از مشاهدات که بیشتر داده‌ها حول آن متمرکز شده‌اند میانگین حسابی نامیده می‌شود. میانگین حسـابی مربـوط به جامعه را با نماد m و میانگین حسابی مربوط به نمونه را با نماد $\bar{x}$ نشان می‌دهیم. برای محاسـبه میـانگین بایـد دقـت کرد که آیا داده‌های مورد نظر طبقه‌بندی شده است یا خیر به همین منظور می‌توان نشان داد:

الف) میانگین در داده‌های طبقه‌بندی نشده: میانگین داده‌های $x_1, x_2, \ldots, x_n$ برابر است با:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ب) میانگین در داده‌های طبقه‌بندی شده: اگر نشان دسته‌ها (مرکز طبقات) $x_1, x_2, \ldots, x_n$ باشد و فراوانی مطلق طبقات $f_1, f_2, \ldots, f_n$ باشد (**n** تعداد طبقات است) می‌توان نشان داد:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{N} = \sum_{i=1}^{n} f_{pi} x_i$$

$$N = \sum_{i=1}^{n} f_i = f_1 + f_2 + \cdots + f_n$$

مثال: میانگین داده‌های زیر را به‌دست آورید.

1 , 7 , 3 , 4 , 11 , 17 , 2 , 8 , 10

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+7+3+4+11+17+2+8+10}{9} = 7$$

تست: معدل دانش‌آموزی با نمرات ۱۶، ۱۵، ۱۶/۵ و ۱۷/۵ برحسب درصد چند است؟

1) %81/5 2) %81/25 3) %80 4) %81

معدل همان میانگین حسابی است پس:

$$\bar{x}=\frac{16+15+16/5+17/5}{4}=16/25$$

با استفاده از تناسب معدل برحسب درصد محاسبه می‌شود:

$$\begin{array}{c|c} 20 & 100 \\ \hline 16/25 & x \end{array} \Rightarrow x=\frac{16/25\times 100}{20}=\%81/25$$

تست: میانگین داده‌های جدول زیر کدام است؟

حدود طبقات	11_13	14_16	17_19
فراوانی	4	6	2

1) 14/5 2) 14

3) 15 4) 15/5

ابتدا مرکز دسته‌ها (مرکز طبقات) را به‌دست می‌آوریم و سپس می‌توان نوشت:

مرکز دسته	12	15	18
فراوانی	4	6	2

$$\bar{x}=\frac{\sum_{i=1}^{n} f_i x_i}{N}=\frac{12\times 4+15\times 6+18\times 2}{4+6+2}=\frac{48+90+36}{12}=14/5$$

تست:اگر میانگین داده‌های جدول زیر برابر ۴ باشد درصد فراوانی نسبی طبقه آخر کدام است؟

حدود طبقات	0_2	2_4	4_6	6_8
فراوانی	5	7	4	x

1) 24/12

2) 27/27

3) 28/32

4) 29/05

پس از محاسبه مرکز طبقات داریم:

$$\bar{x}=\frac{\sum f_i x_i}{\sum f_i}$$

$$4=\frac{(1\times 5)+(3\times 7)+(5\times 4)+(7x)}{5+7+4+x} \Rightarrow 4=\frac{5+21+20+7x}{16+x}$$

$$\Rightarrow \quad 64+4x = 46+7x \quad \Rightarrow \quad 3x = 18 \quad \Rightarrow \quad x = 6$$

درصد فراوانی نسبی طبقه آخر $= \dfrac{6}{5+7+4+6} \times 100 = \dfrac{6}{22} \times 100 \,;\, 27/27$

نکات مهم در ارتباط با میانگین حسابی

1ـ در هر جامعه آماری فقط یک میانگین حسابی وجود دارد.

2ـ اگر داده‌های آماری $x_1 , x_2 , \ldots , x_n$ تشکیل یک تصاعد حسابی (عددی) دهند آنگاه:

$$\bar{x} = \frac{x_1 + x_n}{2}$$

3ـ میانگین تنها شاخص مرکزی است که اگر به جای کلیه داده‌ها قرار بگیرد مجموع داده‌ها تغییر نخواهد کرد.

4ـ مجموع انحرافات داده‌ها از میانگین برابر صفر است یعنی:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 \text{ یا } \sum_{i=1}^{n} f_i (x_i - \bar{x}) = 0$$

5ـ مجموع مربعات انحرافات داده‌ها از میانگین کمترین مقدار ممکن است یعنی:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \min \text{ یا } \sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \min$$

6ـ اگر همه داده‌های آماری با یک عدد جمع یا تفریق کنیم، میانگین نیز با آن عدد جمع یا تفریق می‌شود.

7ـ اگر همه داده‌های آماری را در یک عدد ضرب یا تقسیم کنیم، میانگین نیز در آن عدد ضرب یا تقسیم می‌شود.

تست: **معدل حدسی تعدادی نمره ۱۱ در نظر گرفته شده تفاوت آن از یکایک نمرات 5-، 1-، 3 و 7 گردیده معدل واقعی نمرات چه عددی است؟**

1) 7   2) 10   3) 12   4) 15

اگر x_4، x_3، x_2 و x_1 نمرات مورد نظر باشند خواهیم داشت:

$$\left.\begin{array}{lll} x_1 - 11 = -5 & \Rightarrow & x_1 = 6 \\ x_2 - 11 = -1 & \Rightarrow & x = 10 \\ x_3 - 11 = 3 & \Rightarrow & x_3 = 14 \\ x_4 - 11 = 7 & \Rightarrow & x_4 = 18 \end{array}\right. \quad \Rightarrow \quad \bar{x} = \frac{6+10+14+18}{4} = 12$$

تست: میانگین ۴ درس یک دانش‌آموز هر کدام با ضریب ۱ برابر ۱۵/۵ است، نمره درس پنجم وی که با ضریب ۲ منظور می‌گردد چه عددی باشد تا میانگین ۵ درس وی ۱۶/۵ گردد؟

1) 18/25   2) 18/5   3) 18/75   4) 19

فراوانی 4 درس اول یک و فراوانی درس پنجم برابر 2 است پس:

$$\bar{x} = \frac{\sum x_i F_i}{\sum F_i} \Rightarrow 16/5 = \frac{(4\times 15/5) + (2x)}{4+2} \Rightarrow 16/5\times 6 = 62+2x \Rightarrow x = 18/5$$

تست: میانگین ۵ داده آماری برابر ۳۷/۳ و میانگین ۶ داده آماری دیگر برابر ۴۵ می‌باشد، میانگین این ۱۱ داده آماری کدام است؟

1) 41   2) 41/25   3) 41/5   4) 41/75

نکته: اگر میانگین **n** داده برابر **x** و میانگین **m** داده برابر **y** باشد آنگاه میانگین کل ($m+n$ داده) برابر است با:

$$\bar{x} = \frac{nx+my}{n+m}$$

بنابراین داریم:

$$\bar{x} = \frac{(5\times 37/5) + (6\times 45)}{(5+6)} \Rightarrow \bar{x} = 41/5$$

تست: اگر میانگین داده‌های $x_1 , x_2 , \mathbf{K} , x_n$ برابر $2a$ باشد میانگین، $2x_1-a , 2x_2-a , \mathbf{K} , 2x_n-a$ کدام است؟

1) صفر   2) $4a$ 3) **a**   4) $3a$

تمام داده‌ها در عدد 2 ضرب شده و از همه آن‌ها عدد **a** کاسته شده است پس میانگین برابر است با:

$$\bar{X}_{\text{جدید}} = 2\bar{x}-a \Rightarrow \bar{X} = 2(2a)-a = 3a$$

تست: میانگین نمرات یک دانشجو برابر ۱۵ است، اگر به هر یک از نمرات این دانشجو به اندازه ۲۰ درصد آن نمره را اضافه کنیم میانگین داده‌های جدید کدام است؟

1) 15   2) 16   3) 17   4) 18

منظور از 20% همان 0/2 می‌باشد بنابراین:

$$\bar{X} = \bar{x} + 0/2\bar{x} \quad \Rightarrow \quad \bar{X} = 1/2\bar{x} \quad \Rightarrow \quad \bar{X} = 1/2 \times 15 = 18$$

تست: اگر میانگین داده‌های $x_1 - 1, x_2 - 2, \ldots, x_n - n$ **برابر** $\bar{x}$ **باشد، میانگین داده‌های** $x_1 + 1, x_2 + 2, \ldots, x_n + n$ **کدام است؟**

1) $\bar{x} + \frac{n+1}{2}$ 2) $2\bar{x}$ 3) $\bar{x} + \frac{n(n+1)}{2}$ 4) $\bar{x} + n + 1$

$$\bar{x} = \frac{(x_1 - 1) + (x_2 - 2) + \ldots + (x_n - n)}{n} = \frac{(x_1 + x_2 + \ldots + x_n) - (1 + 2 + \ldots + n)}{n}$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} - \frac{n(n+1)/2}{n} \quad \Rightarrow \quad \frac{x_1 + x_2 + \ldots + x_n}{n} = \bar{x} + \frac{n+1}{2}$$

$$\bar{X} = \frac{(x_1 + 1) + (x_2 + 2) + \ldots + (x_n + n)}{n} = \left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) + \frac{(n+1)}{2}$$

$$\bar{X} = \bar{x} + \frac{n+1}{2} + \frac{(n+1)}{2} \quad \Rightarrow \quad \bar{X} = \bar{x} + (n+1)$$

تست: اگر میانگین داده‌های $x_1, x_2, \ldots, x_n$ **برابر ۴ باشد** $\sum_{i=1}^{n} (3x_i - 6)$ **کدام است؟**

1) صفر 2) $6n$ 3) 3 4) $3n + 6$

عبارت $\sum_{i=1}^{n} (3x_i - 6)$ را می‌توان به‌صورت زیر هم نوشت:

$$\sum_{i=1}^{n} (3x_i - 6) = \sum_{i=1}^{n} (3x_i - 12 + 6) = 3\sum_{i=1}^{n} (x_i - 4) + \sum_{i=1}^{n} 6$$

از آن‌جا که مجموع انحرافات داده‌ها از میانگین برابر صفر است پس:

$$\sum_{i=1}^{n} (3x_i - 6) = 3(0) + \sum_{i=1}^{n} 6 = 6n$$

۲ـ میانگین وزنی (موزون)

چنانچه داده‌ها دارای ارزش نسبی یکسان نباشند میانگین حسابی نمی‌تواند مقدار واقعی متوسط را مشخص کنید در این گونه موارد برای مشخص کردن معدل یا مقدار متوسط از میانگین وزنی استفاده مـی‌کنـیم. اگـر $x_1 , x_2 , \ldots , x_n$ دارای ارزش‌های $w_1 , w_2 , \ldots , w_n$ باشند میانگین وزنی آن‌ها عبارت است از:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

تست: دانشجویی در یک ترم در درس ۳ واحدی آمار نمره ۱۸، در درس ۲ واحدی زبان نمـره ۱۷/۵، در درس ۳ واحدی ریاضی نمره ۱۹ و در درس تک واحدی تنظیم خانواده نمره ۱۲ گرفته است، میانگین یا معـدل ایـن دانشجو در این ترم کدام است؟

1) 16/62   2) 17/5   3) 18/25   4) 17/75

از آن‌جا که واحدهای تمام دروس یکسان نیست پس هر کدام از دروس دارای ارزش نسبی خاصی می‌باشند بنابراین:

$$\bar{x}_w = \frac{\sum_{i=1}^{4} w_i x_i}{\sum_{i=1}^{4} w_i} = \frac{(3\times18) + (2\times17/5) + (3\times19) + (1\times12)}{3+2+3+1} = 17/5$$

تست: معدل یک دانشجو در ۷ درس ۱۶/۷ بوده اگر نمره یک درس ۴ واحدی که ۱۸ بوده اسـت حـذف شـود معدل ۶ درس باقیمانده او کدام است؟ (سه درس باقیمانده ۳ واحدی و سه درس دیگر ۲ واحدی می‌باشند.)

1) 16/35   2) 16/6   3) 16/2   4) 16/15

$$\bar{x}_w = \frac{\sum_{i=1}^{7} w_i x_i}{\sum_{i=1}^{7} w_i} \Rightarrow 16/7 = \frac{\sum_{i=1}^{7} w_i x_i}{\sum_{i=1}^{7} w_i} = \frac{\sum_{i=1}^{7} w_i x_i}{19}$$

$$\sum_{i=1}^{7} w_i x_i = 16/7 \times 19 \Rightarrow \sum_{i=1}^{6} w_i x_i = 16/7 \times 19 - 4 \times 18 = 317/3 - 72 = 245/3$$

$$\bar{X}_w = \frac{\sum_{i=1}^{6} w_i x_i}{\sum_{i=1}^{6} w_i} = \frac{245/3}{(3\times3)+(3\times2)} = \frac{245/3}{15} = 16/35$$

۳ـ میانگین هندسی

اگر داده‌ها به‌صورت درصد، نسبت، شاخص نرخ رشد و ... بیان شده باشند آنگاه میانگین حسابی نمی‌توانـد مقـدار واقعـی متوسط داده‌ها را مشخص کند در این حالت باید از میانگین هندسی استفاده کرد. روش محاسبه میانگین هندسی بـرای داده‌های طبقه‌بندی شده و طبقه‌بندی نشده متفاوت است بنابراین می‌توان نشان داد:

الف) میانگین هندسی داده‌های طبقه‌بندی نشده: اگر $x_1, x_2, \ldots, x_n$ همگی مثبت باشند داریم:

$$(G)\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

ب) میانگین هندسی داده‌های طبقه‌بندی شده: اگر $x_1, x_2, \ldots, x_n$ همگی مثبت باشند داریم:

$$(G)\bar{x}_G = \sqrt[N]{x_1^{F_1} \cdot x_2^{F_2} \cdot \ldots\ x_n^{F_n}} = \sqrt[N]{\prod_{i=1}^{n} x_i^{F_i}}$$

نکته: اگر از میانگین هندسی در دو قسمت فوق لگاریتم (در مبنای 10) بگیریم و از خواص لگاریتم استفاده کنیم آنگاه خواهیم داشت:

$$\log(\bar{x}_G) = \frac{\sum_{i=1}^{n} \log x_i}{n} \qquad \log(\bar{x}_G) = \frac{\sum_{i=1}^{n} f_i \log x_i}{N}$$

مثال: میانگین هندسی اعداد 128، 125، 5 و 2 را به‌دست آورید.

$$\bar{x}_G = \sqrt[4]{2\times5\times125\times128} = \sqrt[4]{2\times5\times5^3\times2^7} = \sqrt[4]{2^8\times5^4} = 2^2\times5 = 20$$

مثال: میانگین هندسی داده‌های جدول زیر برابر کدام می‌باشد؟

حدود طبقات	1_3	5_7	8_10	11_13
فراوانی	1	2	1	2

ابتدا مراکز طبقات را پیدا می‌کنیم و سپس می‌نویسیم:

$$\bar{x}_G = \sqrt[N]{\prod_{i=1}^{4} x_i^{f_i}} = \sqrt[6]{2^1 \times 6^2 \times 9^1 \times 12^2} = \sqrt[6]{2^1 \times 2^2 \times 3^2 \times 3^2 \times 2^4 \times 3^2}$$

$$\bar{x}_G = \sqrt[6]{2 \times 2^6 \times 3^6} = 2 \times 3\ \sqrt[6]{2} = 6\sqrt[6]{2}$$

مثال: قیمت نوعی کالا در چهار سال گذشته برحسب تومان به ترتیب 100، 150، 300 و 800 بوده است قیمت این کالا به‌طور متوسط سالیانه چند برابر شده است؟

نسبت قیمت‌های هر سال به سال قبل عبارتند از $\frac{150}{100}$، $\frac{300}{150}$ و $\frac{800}{300}$ آنگاه با استفاده از رابطهٔ میانگین هندسی داریم:

$$\bar{x}_G = \sqrt[3]{\frac{150}{100} \times \frac{300}{150} \times \frac{800}{300}} = \sqrt[3]{8} = 2$$

یعنی قیمت این کالا به‌طور متوسط سالیانه 2 برابر شده است.

تست: **شاخص قیمت محصولی در سال ۱۳۷۱، ۴۰۰ و در سال ۱۳۷۳، ۹۰۰ تومان بوده است، متوسط نرخ تورم در این فاصله زمانی کدام بوده است؟**

1) %25 2) %125 3) %75 4) %50

$$\bar{x}_G = \sqrt{\frac{72\text{سال}}{71\text{سال}} \times \frac{73\text{سال}}{72\text{سال}}} = \sqrt{\frac{73\text{سال}}{71\text{سال}}} = \sqrt{\frac{900}{400}} = \sqrt{\frac{9}{4}} = \frac{3}{2} = 1/5$$

چون $\bar{x}_G = \frac{3}{2}$ است پس متوسط نرخ تورم این محصول در 2 سال ذکر شده برابر %50 می‌باشد.

تست: **با تغییر مدیریت در یک فروشگاه بزرگ، فروش در سال اول دو برابر سال قبل، در سال دوم سه برابر سال اول و در سال سوم چهار برابر سال اول شده است به‌طور متوسط، فروش از آغاز مدیریت چند برابر شده است؟**

1) سه برابر
2) بیش از سه برابر
3) قدری کمتر از سه برابر
4) دو برابر

$$\bar{x}_G = \sqrt[3]{\frac{\text{سال اول}}{\text{سال قبل}} \times \frac{\text{سال دوم}}{\text{سال اول}} \times \frac{\text{سال سوم}}{\text{سال دوم}}} = \sqrt[3]{2 \times 3 \times 4 \times \frac{1}{3}} = 2$$

دقت کنید که:

$$\frac{\text{سال سوم}}{\text{سال اول}}=4 \Rightarrow \frac{\text{سال سوم}}{\text{سال دوم}}=\frac{\text{سال سوم}}{\text{سال اول}}\times\frac{\text{سال اول}}{\text{سال دوم}}=4\times\frac{1}{3}$$

مثال: سرمایه یک شرکت $12/5\times10^6$ تومان است سرمایه شرکت بعد از یکسان ۱۸٪ نسبت به سال قبل اضافه می‌شود و اگر در سال سوم ۱۴٪ نسبت به سال دوم و در سال چهارم ۲۲٪ نسبت به سال سوم افزایش یابد مطلوب است متوسط رشد سرمایه این شرکت در سه سال گذشته.

$$\bar{x}_G=\sqrt[3]{x_1\cdot x_2\cdot x_3}=\sqrt[3]{\frac{\text{سال دوم}}{\text{سال اول}}\times\frac{\text{سال سوم}}{\text{سال دوم}}\times\frac{\text{سال چهارم}}{\text{سال سوم}}}=\sqrt[3]{1/18\times1/14\times1/22}$$

از آن‌جا که محاسبه ریشه سوم عدد فوق کاری دشوار است می‌توان از بحث لگاریتم نیز استفاده کرد یعنی:

$$\log\bar{x}_G=\frac{1}{3}(\log1/18+\log1/14+\log1/22) \Rightarrow \bar{x}_G \simeq 1/178$$

یعنی به‌طور متوسط 8/17 درصد افزایش داشته است.

۴ـ میانگین توافقی (هارمونیک)

چنانچه داده‌ها به‌صورت میزان تغییرات (مانند سرعت، شتاب و ...) بیان شده باشند در این‌صورت میانگین‌های حسابی و هندسی نمی‌تواند مقدار متوسط واقعی داده‌ها را مشخص کنند. برای این نوع داده‌ها میانگین توافقی مناسب است. میانگین توافقی نیز برای داده‌های طبقه‌بندی شده و طبقه‌بندی نشده متفاوت می‌باشد و می‌توان نشان داد:

الف) میانگین توافقی داده‌های طبقه‌بندی نشده: اگر $x_1, x_2, \ldots, x_n$ همگی هم علامت و مخالف صفر باشند داریم:

$$(H)\bar{x}_H=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}}$$

ب) میانگین توافقی داده‌های طبقه‌بندی شده: اگر $x_1, x_2, \ldots, x_n$ همگی هم‌علامت و مخالف صفر باشند داریم:

$$(H)\bar{x}_H=\frac{N}{\sum_{i=1}^{n}\frac{f_i}{x_i}}=\frac{f_1+f_2+\ldots+f_n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_n}{x_n}}$$

تذکر: اگر داده‌ها نسبت‌هایی باشند که صورت و مخرج آن‌ها دارای واحد متفاوت هستند ولی صورت نسبت‌ها مساوی است در این صورت از میانگین توافقی استفاده می‌کنیم ولی اما اگر مخرج نسبت‌ها مساوی باشند از میانگین حسابی استفاده می‌کنیم.

تذکر: اگر داده‌ها دارای وزن باشند میانگین توافقی آن‌ها به‌صورت زیر محاسبه می‌شود:

$$(H)\bar{x}_H = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}} = \frac{w_1 + w_2 + \cdots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \cdots + \frac{w_n}{x_n}}$$

مثال: اتومبیلی سه قطعه از یک تونل را با سرعت‌های 50، 60 و 70 کیلومتر در ساعت طی می‌کند:

الف) اگراین سه قطعه از تونل طول‌های یکسان داشته باشند سرعت متوسط این اتومبیل درون تونل چقدر است؟

ب) اگر به ترتیب $\frac{1}{5}$، $\frac{2}{5}$ و $\frac{2}{5}$ تونل را با سرعت‌های داده شده طی کند سرعت متوسط آن درون تونل چقدر است؟

الف) چون طول‌های هر سه قطعه از تونل یکسان است می‌توان نوشت:

$$\bar{x}_H = \frac{3}{\frac{1}{50} + \frac{1}{60} + \frac{1}{70}} \Rightarrow \bar{x}_H = 58/88$$

ب) چون طول‌های سه قطعه از تونل یکسان نمی‌باشد پس داده‌ها دارای ارزش نسبی یکسانی نمی‌باشند و می‌توان نوشت:

$$\bar{x}_H = \frac{\frac{1}{5} + \frac{2}{5} + \frac{2}{5}}{\frac{\frac{1}{5}}{50} + \frac{\frac{2}{5}}{60} + \frac{\frac{2}{5}}{70}} \Rightarrow \bar{x}_H = 61/05$$

تست: اتومبیلی زمان حرکت خود را به چهار قسمت مساوی تقسیم می‌کند، اگر در این چهار قسمت به ترتیب با سرعت‌های ۳۰، ۵۰، ۴۵ و ۷۵ کیلومتر در ساعت حرکت کند سرعت متوسط اتومبیل چقدر است؟

1) $45 \frac{km}{h}$ 2) $50 \frac{km}{h}$ 3) $55 \frac{km}{h}$ 4) $60 \frac{km}{h}$

از آن جا که واحد صورت و مخرج داده‌ها $\left(\frac{km}{h}\right)$ یکسان نیست و مخرج نسبت‌ها $\left(\frac{t}{4}\right)$ یکسان می‌باشد می‌توان از میانگین حسابی استفاده کرد و نوشت:

$$\bar{x} = \bar{V} = \frac{30 + 50 + 45 + 75}{4} = \frac{200}{4} = 5 \frac{km}{h}$$

تست: اتومبیلی مسافت بین دوشهر را به چهار قسمت مساوی تقسیم کرده است و در این چهار قسمت به ترتیب با سرعت‌های ۳۰، ۵۰، ۴۵ و ۷۵ کیلومتر در ساعت حرکت می‌کند، سرعت متوسط این اتومبیل چقدر است؟

1) $45\ \frac{km}{h}$ 2) $50\ \frac{km}{h}$ 3) $55\ \frac{km}{h}$ 4) $60\ \frac{km}{h}$

واحد صورت و مخرج داده‌ها $\left(\frac{km}{h}\right)$ یکسان نیست ولی مقدار صورت نسبت‌ها یکسان است بنـابراین از میـانگین تـوافقی استفاده می‌کنیم و می‌نویسیم:

$$\bar{x}_H = \bar{V} = \frac{4}{\frac{1}{30}+\frac{1}{50}+\frac{1}{45}+\frac{1}{75}} = 45\ \frac{km}{h}$$

تست: فرض کنید در یک مسافرت، ۱۰۰ کیلومتر را با قطار با سـرعت متوسـط ۷۰ کیلـومتر در سـاعت و ۷۰۰ کیلومتر را با کشتی با سرعت متوسط ۴۵ کیلومتر در ساعت و ۱۵۰ کیلومتر را با تاکسی با سرعت متوسط ۱۱۰ کیلومتر در ساعت و ۹۰۰ کیلومتر را با هواپیما با سرعت متوسط ۶۰۰ کیلومتر در ساعت طـی نمـوده باشـیم، سرعت متوسط در طول این مسافرت برای کل این مسافت چقدر است؟

1) 107 2) 206/25 3) 111/3 4) 93/2

$$\bar{x}_H = \frac{N}{\sum_{i=1}^{4}\frac{f_i}{x_i}} = \frac{100+700+150+900}{\frac{100}{70}+\frac{700}{45}+\frac{150}{110}+\frac{900}{600}}$$

$$\bar{x}_H = 93/2\ \frac{km}{h}$$

تست: سه ماشین کالای یکسانی را تولید می‌کنند، اولی یک کالا را در ۳ دقیقه، دومی در ۴ دقیقه و سـومی در ۱۲ دقیقه تولید می‌کنند، اگر این سه ماشین با هم کار کنند به‌طور متوسط یک کالا در چند دقیقه تولید می‌شود؟

1) 4 2) کمتر از 4 3) اندکی بیشتر از 4 4) 4/5

$$\bar{x}_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{12}} = \frac{3}{\frac{4+3+1}{12}} = \frac{3}{\frac{8}{12}} = \frac{36}{8}$$

$\bar{x}_H = 4/5$ دقیقه

۵ـ میانگین درجه ۲

میانگین درجه 2 برای هر نوع داده‌ای قابل محاسبه است و میانگین درجه دوم داده‌های $x_1, x_2, \mathbf{K}, x_n$ عبـارت است از:

$$(Q)\ \bar{x}_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)}$$

$$(Q)\ \bar{x}_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i\, x_i^2} = \sqrt{\frac{1}{N}\left(f_1 x_1^2 + f_2 x_2^2 + \cdots + f_n\, x_n^2\right)}$$

نکته: اگر داده‌ها دارای ارزش یکسانی نباشند میانگین درجه 2 آن‌ها به صورت زیر به‌دست می‌آید:

$$(Q)\ \bar{x}_Q = \sqrt{\sum_{i=1}^{n} w_i x_i^2} \quad , \quad 0<w<1 \quad , \quad \sum_{i=1}^{n} w_i = 1$$

مثال: میانگین درجه 2 داده‌های 7، 5، 5، 4 و 2 را به‌دست آورید.

$$\bar{x}_Q = \sqrt{\frac{2^2+4^2+5^2+5^2+7^2}{5}} \simeq 4/88$$

۶ـ میانگین درجه r

چنانچه **r** یک عدد صحیح غیرصفر باشد آنگاه میانگین درجه **r** ام به‌صورت زیر به‌دست می‌آید:

$$\bar{x}_r = \sqrt[r]{\frac{1}{n}\sum_{i=1}^{n} x_i^r} = \sqrt[r]{\frac{1}{n}\left(x_1^r + x_2^r + \cdots + x_n^r\right)}$$

$$\bar{x}_r = \sqrt[r]{\frac{1}{N}\sum_{i=1}^{n} f_i\, x_i^r} = \sqrt[r]{\frac{1}{N}\left(f_1 x_1^r + f_2 x_2^r + \cdots + f_n\, x_n^r\right)}$$

نکته: اگر داده‌ها دارای ارزش یکسانی نباشند میانگین درجه **r** ام آن‌ها به‌صورت زیر به‌دست می‌آید:

$$\bar{x}_r = \sqrt[r]{\sum_{i=1}^{n} w_i\, x_i^r} \quad , \quad 0<w_i<1 \quad , \quad \sum_{i=1}^{n} w_i = 1$$

نکته: میانگین درجه **r**، کاملترین نوع میانگین است. انواع دیگر میانگین‌ها حالت‌های خاص این میانگین می‌باشند.

نکته: رابطه زیر همواره بین میانگین های مختلف برقرار است:

$$\bar{x}_H \le \bar{x}_G \le \bar{x} \le \bar{x}_Q$$

میانه

میانه کمیتی است که در وسط صف منظم داده‌ها قرار دارد. یا به بیان دیگر میانه عـددی اسـت کـه نصـف داده‌هـا از آن بزرگتر و نصف دیگر داده‌ها از آن کوچکتر می‌باشند و آن را با نماد Md نمـایش مـی‌دهـیم. روش محاسـبه میانـه بـرای داده‌های طبقه‌بندی شده و طبقه‌بندی نشده متفاوت است و می‌توان نشان داد:

الف) میانه داده‌های طبقه‌بندی نشده: ابتدا داده‌های $x_1 , x_2 , \ldots , x_n$ را به‌صورت صعودی مرتب می‌کنیم ردیف میانه از رابطه $\frac{n+1}{2}$ به‌دست می‌آید و خود میانه را می‌توان با استفاده از فرمول‌های زیر به‌دست آورد:

$$Md = \begin{cases} x_{\frac{n+1}{2}} & : \quad n \text{ فرد} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & : \quad n \text{ زوج} \end{cases}$$

ب) میانه داده‌های طبقه‌بندی شده: ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم، اولین طبقه‌ای کـه فراوانـی تجمعـی آن بزرگتر یا مساوی $\frac{N}{2}$ باشد طبقه میانه‌دار نامیده می‌شود، پس از مشخص کردن طبقه میانه‌دار، میانه را از فرمول زیـر به‌دست می‌آوریم:

$$Md = L_i + \left(\frac{\frac{N}{2} - F_{ci-1}}{f_i} \right) \times C$$

L_i : کران پایین طبقه میانه‌دار

f_i : فراوانی مطلق طبقه میانه‌دار

F_{ci-1} : فراوانی تجمعی طبقه ماقبل میانه‌دار

C : فاصله طبقات (طول دسته)

مثال: میانه و ردیف میانه داده‌های زیر را به‌دست آورید.

1 2 0 3 3 4 5

ابتدا داده‌ها را به‌صورت صعودی مرتب می‌کنیم:

0 1 2 3 3 4 5

$\text{ردیف میانه} = \frac{n+1}{2} = \frac{7+1}{2} = 4$

یعنی چهارمین عدد در وسط صف منظم داده‌ها قرار دارد. چون تعداد داده‌ها $(n=7)$ فرد است بنابراین:

$$Md = x_{\frac{n+1}{2}} = x_4 \quad \Rightarrow \quad Md = 3$$

مثال: میانه و ردیف میانه داده‌های زیر را به‌دست آورید.

9 7 8 6 8 10 9 9 5 7

تعداد داده‌ها ($n = 10$) زوج می‌باشد پس بعد از مرتب کردن داده‌ها به‌صورت صعودی می‌توان نوشت:

5 6 7 7 8 8 9 9 9 10

$$\text{ردیف میانه} = \frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5/5$$

شماره ردیف میانه عددی اعشاری است که بین دو عدد صحیح متوالی 5 و 6 می‌باشد پس پنجمین و ششمین داده در صف منظم داده‌ها دو عددی می‌باشند که در وسط صف منظم قرار دارند. از آن‌جا که تعداد داده‌ها زوج است برای محاسبه میانه داریم:

$$Md = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) = \frac{1}{2}\left(x_{\frac{10}{2}} + x_{\frac{10}{2}+1}\right) = \frac{1}{2}(x_5 + x_6)$$

$$Md = \frac{1}{2}(8+8) = 8$$

تست: میانه داده‌های $\left\{2^n \mid x \in N \ , \ n \le 10\right\}$ **کدام است؟**

1) 32   2) 48   3) 36   4) 64

داده‌های مورد نظر عبارتند از:

$$2^1 \ , 2^2 \ , 2^3 \ , \ldots \ , 2^{10}$$

تعداد داده‌ها زوج می‌باشد پس میانه آن‌ها عبارت است از:

$$Md = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}\left(2^5 + 2^6\right) = \frac{1}{2}(32+64) = 48$$

تست: میانگین داده‌های $3 \ , 5 \ , x \ , 1 \ , 2$ **برابر ۳ است، میانه این داده‌ها کدام است؟**

1) 3/5   2) 2/5   3) 3   4) 4/5

$$\bar{x} = \frac{1+2+x+3+5}{5} \quad \Rightarrow \quad 3 = \frac{11+x}{5} \quad \Rightarrow \quad x = 4$$

$$Md = x_{\frac{n+1}{2}} \Rightarrow Md = x_3 = 3$$

چون بعد از مرتب کردن داده‌ها خواهیم داشت: 5 , 4 , 3 , 2 , 1.

تست: **در یک جامعه آماری فواصل داده‌ها از یکدیگر ۴، بیشترین داده ۴۲ و تعداد داده‌ها برابر ۱۰ است میانه این داده‌ها چقدر است؟**

20 (1   22 (2   24 (3   26 (4

داده‌های آماری فوق تشکیل تصاعد عددی با جمله آخر $a_{10} = 42$ قدرنسبت $d = 4$ و تعداد جملات $n = 10$ است از آن‌جا که تعداد داده‌ها زوج می‌باشد میانگین دو داده وسطی برابر میانه است بنابراین:

$$Md = \frac{a_5 + a_6}{2} = \frac{(a_{10} - 5d) + (a_{10} - 4d)}{2} = \frac{2a_{10} - 9d}{2} = \frac{2(42) - 9(4)}{2}$$

$$Md = 42 - 18 = 24$$

نکته: اگر داده‌های گسسته همراه با فراوانی باشند برای محاسبه میانه ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم اگر تعداد داده‌ها فرد باشد $t = \frac{N+1}{2}$ و اگر تعداد داده‌ها زوج باشد $t = \frac{N}{2}$ می‌گیریم و مقدار **t** بین دو فراوانی تجمعی و یا مساوی یکی از فراوانی‌های تجمعی است در این‌صورت:

$$F_{ci-1} < t \le F_{ci} \Rightarrow Md = x_i \quad : \quad \text{فرد}$$

$$\left.\begin{array}{l} F_{ci-1} < t < F_{ci} \Rightarrow Md = x_i \\ t = F_{ci} \Rightarrow Md = \frac{1}{2}(x_i + x_{i+1}) \end{array}\right\} \quad : \quad \text{زوج}$$

مثال: جدول توزیع فراوانی زیر، میانه هر جدول را به‌دست آورید.

الف)

x_i	6	7	8	9	10
f_i	2	3	1	7	2
F_{ci}	2	5	6	13	15

ب)

x_i	6	7	8	9	10
f_i	2	3	2	3	2
F_{ci}	2	5	7	10	12

ج)

x_i	6	7	8	9	10
f_i	2	3	2	5	2
F_{ci}	2	5	7	12	14

الف) چون $N=2+3+1+7+2=15$ فرد است پس $t=\frac{N+1}{2}=8$ و در این‌صورت:

$$6<t<13 \quad \Rightarrow \quad Md=9$$

$x_i=9$ نشان دسته‌ای است که فراوانی تجمعی آن $F_{ci}=13$ است.

ب) چون $N=2+3+2+3+2=12$ زوج است پس $t=\frac{N}{2}=6$ و در این‌صورت:

$$5<t<7 \quad \Rightarrow \quad Md=8$$

$x_i=8$ نشان دسته‌ای است که فراوانی تجمعی آن $F_{ci}=7$ می‌باشد.

ج) چون $N=2+3+2+5+2=14$ زوج است پس $t=\frac{N}{2}=7$ در این صورت:

$$t=F_{ci}=7 \quad \Rightarrow \quad Md=\frac{1}{2}\left(x_i+x_{i+1}\right)=\frac{1}{2}(8+9)=8/5$$

تست: میانه در داده‌های جدول زیر کدام است؟

x_i	5	12	15	20
f_i	6	8	12	4

1) 10   2) 13/5

3) 15   4) هیچ‌کدام

ابتدا ردیف فراوانی تجمعی را به‌صورت زیر به‌دست می‌آوریم:

x_i	5	12	15	20
f_i	6	8	12	4
F_{ci}	6	14	26	30

از آن‌جا که $N = 6+8+12+4 = 30$ زوج است بنابراین:

$$t = \frac{N}{2} = \frac{30}{2} = 15 \quad \Rightarrow \quad 14 < t < 26 \quad \Rightarrow \quad Md = 15$$

مثال: میانه داده‌های زیر را به‌دست آورید.

حدود طبقات	20_40	40_60	60_80	80_100	100_120
فراوانی	6	9	11	14	20

ابتدا فراوانی تجمعی هر طبقه را به‌دست می‌آوریم:

حدود طبقات	20_40	40_60	60_80	80_100	100_120
فراوانی	6	9	11	14	20
فراوانی تجمعی	6	15	26	40	60

چون داده‌های مورد نظر طبقه‌بندی شده‌اند پس اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر از $\frac{N}{2} = \frac{60}{2} = 30$ است را طبقه میانه‌دار می‌گوییم، طبقه میانه‌دار طبقه چهارم خواهد بود سپس می‌توان نوشت:

$$Md = L_i + \left(\frac{\frac{N}{2} - F_{ci-1}}{f_i} \right) \times C$$

$$Md = 80 + \left(\frac{\frac{60}{2} - 26}{14} \right) \times 20 = 80 + \left(\frac{4}{14} \right) \times 20 = 85/71$$

مثال: میانه را از توزیع جدول زیر حساب کنید.

حدود طبقات	10_25	25_40	40_55	55_70	70_85	85_100
فراوانی‌ها	6	20	44	26	3	1

داده‌های مورد نظر طبقه‌بندی شده‌اند پس ابتدا فراوانی‌های تجمعی را می‌یابیم و سپس مانند قبل عمل می‌کنیم:

CL	$10-25$	$25-40$	$40-55$	$55-70$	$70-85$	$85-100$
f_i	6	20	44	26	3	1
F_{ci}	6	26	70	96	99	100

از آن‌جا که $\frac{N}{2} = 50$ پس طبقه سوم طبقه میانه‌دار می‌باشد و:

$$Md = L_i + \left(\frac{\frac{N}{2} - F_{ci-1}}{f_i} \right) \times C = 40 + \left(\frac{50-26}{44} \right)(15) = 40 + \left(\frac{24}{44} \right) 15 = 48/18$$

نکات مهم در ارتباط با میانه

1ـ در هر جامعه آماری فقط یک میانه وجود دارد.

2ـ اگر یک عدد ثابت به تمام داده‌ها اضافه یا کم شود آن عدد به میانه نیز اضافه یا کم می‌شود.

3ـ اگر یک عدد ثابت در تمام داده‌ها ضرب یا تقسیم شود آن عدد در میانه نیز ضرب یا تقسیم می‌شود.

4ـ اگر در بین داده‌ها، داده‌ای وجود داشته باشد که اختلاف آن با بقیه داده‌ها زیاد است از میانه به عنوان شاخص مرکزی استفاده می‌کنیم.

5ـ از نظر هندسی میانه خطی عمود به معادله $x = Md$ است که نمودار هیستوگرام را از نظر سطح به دو قسمت مساوی تقسیم می‌کند.

6ـ عبارت $\sum_{i=1}^{n} |x_i - Md|$ و $\sum_{i=1}^{n} f_i |x_i - Md|$ همواره می‌نیمم است.

تست: میزان سود شرکت سهامی گیتا در شش سال گذشته برحسب درصد فروش به ترتیب ۴، ۲، ۳، ۴، ۲۶ و ۱ می‌باشد کدام‌یک از کمیت‌های زیر به عنوان شاخص مرکزی وضع سودآوری شرکت را بهتر نشان می‌دهد؟

1) 3/5   2) 3   3) 4   4) 7

در بین داده‌های موجود چون اختلاف داده 26 با بقیه زیاد است پس باید از شاخص مرکزی میانه استفاده کنیم:

$$Md = \frac{3+4}{2} = 3/5$$

تست: میانه ۴ عدد x_1, x_2, x_3, x_4 برابر a است میانه اعداد $2-\frac{x_4}{3}$، $2-\frac{x_3}{3}$، $2-\frac{x_2}{3}$ و $2-\frac{x_1}{3}$ کدام است؟

1) $2-a$ 2) $-\frac{a}{3}$ 3) $\frac{a}{3}-2$ 4) $2-\frac{a}{3}$

تمام اعداد در عدد $-\frac{1}{3}$ ضرب و با عدد 2 جمع شده‌اند پس میانه آن‌ها برابر می‌شود با:

$$Md = 2-\frac{a}{3}$$

تست: با استفاده از داده‌های زیر اگر عبارت $\sum_{i=1}^{n} |a-x_i|$ می‌نیمم شود مقدار آن کدام است؟

$3, 1, 0, 1, 2, 2, 4, 3, 1$

1) 2   2) 1   3) -1   4) 9

عبارت $\sum_{i=1}^{n} |a-x_i| = \sum_{i=1}^{n} |x_i - a|$ زمانی می‌نیمم می‌شود که **a** میانه اعداد باشد. پس ابتدا اعداد را به‌صورت صعودی مرتب می‌کنیم و چون تعداد آن‌ها فرد می‌باشد می‌توان نوشت:

$0, 1, 1, 1, 2, 2, 3, 3, 4$

$$Md = a = \frac{x_{9+1}}{2} = x_5 \quad \Rightarrow \quad a = 2$$

حال می‌توان نشان داد که:

$$\sum_{i=1}^{9} |a-x_i| = |2-0| + |2-1| \times 3 + |2-2| \times 2 + |2-3| \times 2 + |2-4|$$
$$= 2+3+0+2+2 = 9$$

مد (نما)

مد داده‌ای است که بیشترین فراوانی را در میان داده‌ها داشته باشد و آن را با نماد Mo نمایش می‌دهیم. روش محاسبه مد یا نما در داده‌های طبقه بندی شده و طبقه‌بندی نشده متفاوت است بنابراین خواهیم داشت:

الف) مد یا نمای داده‌های طبقه‌بندی نشده: در یک مجموعهٔ داده‌ها عددی که بیشترین فراوانی را داشته باشد نما یا مد داده‌ها است. اگر فراوانی همه داده‌ها یکسان باشد مد وجود ندارد و اگر چند داده دارای بیشترین فراوانی باشند همه آن‌ها را مد می‌نامیم.

ب) مد یا نمای داده‌های طبقه‌بندی شده: ابتدا طبقه‌ای که بیشترین فراوانی مطلق را دارد مشخص می‌کنیم این طبقه را طبقه مددار می‌نامیم سپس مد را با استفاده از فرمول زیر به‌دست می‌آوریم:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2} \right) \times C$$

L_i : کران پایین طبقه مددار

C : فاصله طبقات (طول دسته)

$d_2 = f_i - f_{i+1}$: تفاضل فراوانی مطلق طبقه مددار از فراوانی طبقه بعد

$d_1 = f_i - f_{i-1}$: تفاضل فراوانی مطلق طبقه مددار از فراوانی طبقه قبل:

تست: **در داده‌های آماری** 5 , 2 , 1 , 3 , 5 , 9 , 8 , 3 , 3 , 5 , 3 **مجموع میانه و مد چقدر است؟**

1) 4   2) 6   3) 7   4) 8

ابتدا داده‌ها را به‌صورت صعودی مرتب می‌کنیم:

1 2 3 3 3 3 5 5 5 8 9

$$Md = x_{\frac{11+1}{2}} = x_6 \Rightarrow Md = 3 \qquad Mo = 3$$

$$Md + Mo = 3+3 = 6$$

تست: در جدول داده‌های مقابل تفاوت مد از میانگین کدام است؟

1) 0/4

2) 0/7

3) 1/2

4) 1/4

داده‌ها	12	14	16	18	20
فراوانی	2	3	2	2	1

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{(2\times12)+(3\times14)+(2\times16)+(2\times18)+(1\times20)}{2+3+2+2+1} = \frac{24+42+32+36+20}{10}$$

$$\Rightarrow \bar{x} = \frac{154}{10} = 15/4 \quad \text{و} \quad Mo = 14$$

$$\bar{x} - Mo = 15/4 - 14 = 1/4$$

تست: مد در جدول مقابل کدام است؟

1) 13

2) 8/5

3) 17/5

4) 18

حدود دسته	2_4	5_7	8_10	11_13
f_i	2	5	8	5

بیشترین فراوانی مطلق مربوط به دسته سوم می‌باشد پس این دسته را دسته مدار می‌نامیم و با استفاده از فرمول زیر مد را محاسبه می‌کنیم:

حدود دسته	1/5_4/5	4/5_7/5	7/5_10/5	10/5_13/5
f_i	2	5	8	5

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \times C$$

$$Mo = 7/5 + \left(\frac{(8-5)}{(8-5)+(8-5)}\right) \times 2 = 7/5 + \left(\frac{3}{6}\right)(2) = 8/5$$

مثال: مد را از توزیع زیر حساب کنید.

نمرات	10_25	25_40	40_55	55_70	70_85	85_100
فراوانی‌ها	6	20	44	26	3	1

با توجه به این‌که دسته سوم بیشترین فراوانی مطلق را دارد می‌توان نوشت:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \times C$$

$$d_1 = f_i - f_{i-1} = 44 - 20 = 24$$

$$d_2 = f_i - f_{i+1} = 44 - 26 = 18$$

$$Mo = 40 + \left(\frac{24}{18+24}\right) \times (15) = 40 + \frac{24 \times 15}{42} = 48/57 \; ; \; 49$$

تست: هزینه ۱۰۰ خانوار در جدول زیر مندرج است، اگر مد توزیع برابر ۲۴ باشد فراوانی‌های مجهول کدامند؟

1) $x = 19 \quad , \quad y = 25$

2) $x = 25 \quad , \quad y = 19$

3) $x = 17 \quad , \quad y = 27$

4) $x = 27 \quad , \quad y = 17$

هزینه	0_10	10_20	20_30	30_40	40_50
تعداد خانوار	4	x	37	y	15

تعداد خانوارها برابر 100 می‌باشد بنابراین:

$$4 + x + 37 + y + 15 = 100 \Rightarrow x + y = 44$$

به دلیل $Mo = 24$، طبقه سوم طبقه مددار می‌باشد بنابراین داریم:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \times C$$

$$24 = 20 + \left(\frac{37-x}{(37-x)+(37-y)}\right) \times 10 \Rightarrow 4 = 10\left(\frac{37-x}{74-(x+y)}\right)$$

با قرار دادن $x + y = 44$ به‌دست می‌آید:

$$4 = 10\left(\frac{37-x}{74-44}\right) \Rightarrow 4 = \frac{37-x}{3} \Rightarrow \begin{matrix} x = 25 \\ y = 19 \end{matrix}$$

نکات مهم در ارتباط با مد یا نما

1ـ در یک جامعه آماری ممکن است مد منحصر به فرد نباشد.

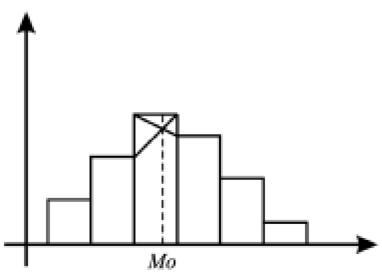
2ـ اگر جامعه چند مدی باشد مد شاخص معتبری نیست.

3ـ اگر به تمام داده‌ها یک مقدار ثابت اضافه یا کم کنیم به مد نیز همان مقدار ثابت اضافه یا کم می‌شود.

4ـ اگر تمام داده‌ها در یک مقدار ثابت ضرب یا تقسیم شوند مد نیز در آن ضرب یا تقسیم می‌شود.

5ـ در بین شاخص‌های مرکزی مد از اهمیت کمتری برخوردار است در صورتی که در داده‌های کیفی مد تنها شاخص مرکزی است.

6ـ برای تعیین مد به روش ترسیمی ابتدا منحنی هیستوگرام فراوانی را رسم می‌کنیم. مطابق شکل زیر دو انتهای بالایی بلندترین مستطیل را به انتهای رئوس مستطیل‌های مجاور وصل می‌کنیم پای عمود محل برخورد این دو خط همان مد یا نما می‌باشد.



تست: با توجه به داده‌های آماری ۱۴، ۱۰، ۱۰، ۹ و ۷ اندازه کدام شاخص زیر با تعداد اندازه داده ۹ بـه ۱۰ تغییـر خواهد کرد؟

1) دامنه تغییرات 2) میانه 3) میانگین 4) نما

با تغییر اندازه داده 9 به 10دامنه تغییرات، میانه و نما تغییر نمی‌کند ولی میانگین تغییر می‌کند پس گزینه سوم صحیح است.

تست: ششمین عددی که با قرار گرفتن در بین داده‌های ۶، ۵، ۴، ۳ و ۲ موجب می‌شود، میانگین، میانه و مـد آن‌ها برابر گردد چیست؟

3 (4   4 (3   5 (2   6 (1

اگر ششمین عدد را **a** در نظر بگیریم **a** حتماً باید یکی از اعداد 6، 5، 4، 3 و 2 باشد چون در غیر این‌صورت مد نخواهیم داشت بنابراین عدد **a** مد می‌شود و به دلیل این‌که مد، میانه و میانگین برابر هستند خواهیم داشت:

$$a\,,2\,,3\,,4\,,5\,,6 \;\Rightarrow\; Mo=a$$

$$\bar{x}=\frac{a+2+3+4+5+6}{6}=\frac{20+a}{6}$$

$$\bar{x}=Mo=a \;\Rightarrow\; \frac{20+a}{6}=a \;\Rightarrow\; 20+a=6a \;\Rightarrow\; a=4$$

حال اگر به جای **a** در داده‌ها عدد 4 را جایگذاری کنیم به وضوح دیده می‌شود که میانه، میانگین و مد با هم برابر هستند یعنی 4 می‌باشند.

مقایسه بین میانگین، میانه و مد

1ـ یک شاخص مرکزی هنگامی با ارزش است که بر کلیه داده‌ها متکی باشد. بزرگترین برتری میانگین، همین است که بر کلیه داده‌ها متکی می‌باشد در صورتی که میانه و مد این‌طور نمی‌باشند.

2ـ میانگین نسبت به میانه و مد پایدارتر است بدان معنی که اگر از جامعه‌ای نمونه‌های متفاوتی گرفته شود نوسات میانگین پس از محاسبه از نوسانات میانه و مد کمتر است.

3ـ اگر داده‌هایی که در ابتدا یا انتهای توزیع گرفته‌اند از سایر داده‌ها به‌طور قابل ملاحظه‌ای فاصله داشته باشند بهتر است از میانه به‌عنوان شاخص مرکزی استفاده کنیم چون میانگین از مقدار متوسط واقعی دور خواهد بود.

4ـ اگر به ابتدا یا انتهای جامعه دسترسی نداشته باشیم از میانه استفاده می‌کنیم.

چندک‌ها

اگر داده‌ها را به‌صورت صعودی مرتب کنیم چندک‌ها کمیت‌هایی هستند که دامنه تغییرات را به فواصل چندکی به گونه‌ای تقسیم می‌کنند که فراوانی‌ها در هر یک از فاصله‌ها درصد مشخصی از فراوانی کل را تشکیل می‌دهند.

چندک مرتبه **P** که در آن $0<P<1$، کمیتی است که **100P** درصد داده‌ها از آن کوچکتر یا با آن مساوی می‌باشند و آن را با نماد Q_P نمایش می‌دهیم چندک‌ها را برحسب مقادیر گوناگون **P** به سه گروه عمدهٔ زیر تقسیم می‌کنیم:

1ـ **چارک‌ها**: چارک‌ها دامنه تغییرات را به چهار قسمت مساوی تقسیم می‌کنند و به ازای $P=0/25\,,0/5\,,0/75$ به‌دست می‌آیند و آن‌ها را با Q_1، Q_2 و Q_3 نمایش می‌دهیم.

2ـ **دهک‌ها**: دهک‌ها دامنه تغییرات را به ده قسمت مساوی تقسیم می‌کنند و به ازای $P=0/1\,,0/2\,,\mathbf{K}\,,0/9$ به‌دست می‌آیند و آن‌ها را با $D_1\,,D_2\,,\mathbf{K}\,,D_9$ نمایش می‌دهیم.

3ـ **صدک‌ها**: صدک‌ها دامنه تغییرات را به صد قسمت مساوی تقسیم می‌کنند و به ازای $P=0/01\,,0/02\,,\mathbf{K}\,,0/99$ به‌دست می‌آیند و آن‌ها را با $P_1\,,P_2\,,\mathbf{K}\,,P_{99}$ نمایش می‌دهیم.

نکته: همواره می‌توان چارک‌ها و دهک‌ها را برحسب صدک‌ها بیان کرد به عنوان مثال: میانه همان چارک دوم یا دهک پنجم یا همان صدک پنجاهم است. چارک اول همان صدک بیست و پنجم است. دهک چهارم همان صدک چهارم است:

$$Md=Q_2=Q_5=P_{50}\quad,\quad Q_1=P_{25}\quad,\quad Q_4=P_{40}$$

نکته: از آن‌جا که چارک‌ها و دهک‌ها را می‌توان برحسب صدک‌ها بیان کرد بنابراین روش محاسبه صدک‌ها همان روش محاسبه چارک‌ها و دهک‌ها باشد که به ازای $\mathbf{P}$های خاص به‌دست می‌آیند. روش محاسبه چندک‌ها برای داده‌های طبقه‌بندی شده و طبقه‌بندی نشده متفاوت است و می‌توان نشان داد:

الف) چندک‌ها در داده‌های طبقه‌بندی نشده: اگر $\mathbf{n}$ داده را به‌صورت صعودی $x_1\le x_2\le \mathbf{L}\le x_n$ مرتب کنیم برای محاسبه Q_P (صدک $100P$ام) ابتدا، $(n+1)P$ را به‌دست می‌آوریم اگر حاصل این عبارت برابر عدد صحیح $\mathbf{Z}$ شود در این‌صورت $Q_P=x_Z$، یعنی چندک مورد نظر عدد Zام در صف داده‌های مرتب شده می‌باشد و اگر $(n+1)P$ عدد صحیح نباشد (عدد اعشاری باشد) در این صورت قسمت صحیح آن را با z و قسمت اعشاری آن را با $\mathbf{w}$ نشان می‌دهیم و چندک مورد نظر برابر می‌شود با میانگین وزنی zامین و $(z+1)$امین دادهٔ صف منظم داده‌ها و به‌صورت زیر به‌دست می‌آید: $Q_P=(1-w)x_z+wx_{z+1}$

تذکر: محل چندک مورد نظر در داده‌های طبقه‌بندی نشده با توجه به این‌که $P=0/01\,,0/02\,,\mathbf{K}\,,0/99$ می‌باشد از رابطه زیر به‌دست می‌آید:

$$\text{محل چندک}=np+\frac{1}{2}$$

ب) چندک‌ها در داده‌های طبقه‌بندی شده: روش محاسبه چندک‌ها در داده‌های طبقه‌بندی شده کاملاً مشابه روش محاسبه میانه برای این داده‌ها است. برای محاسبه Q_P (صدک $100P$ام) ابتدا با استفاده از **NP** اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا برابر **NP** باشد را مشخص می‌کنیم و به آن طبقه Q_P می‌گوییم و سپس با استفاده از فرمول زیر Q_P را به‌دست می‌آوریم:

$$Q_P = L_i + \left(\frac{NP - F_{ci-1}}{f_i} \right) \times C$$

کران پایین طبقه : Q_P L_i فراوانی مطلق طبقه : Q_P f_i

فراوانی تجمعی طبقه ماقبل : Q_P F_{ci-1} فاصله طبقات (طول دسته) : C

مثال: برای داده‌های 3، 4، 4، 5، 7، 6، 1، 2، 3، 4، 7، 2، 0 عبارت‌های زیر را محاسبه کنید.

الف) میانه (Q_2)، ب) چارک اول (Q_1)، ج) دهک سوم (D_3)، د) صدک 71 ام (P_{71}).

در ابتدا داده‌ها را به ترتیب صعودی مرتب می‌کنیم:

0 , 1 , 2 , 2 , 3 , 3 , 4 , 4 , 4 , 5 , 6 , 7 , 7

حال می‌توان نوشت:

الف) $P = 0/5 \Rightarrow (n+1)P = (13+1)(0/5) = 7$: عدد صحیح می‌باشد

$\Rightarrow Q_2 = x_7 = 4$

ب) $P = 0/25 \Rightarrow (n+1)P = (13+1)(0/25) = 3/5 = 3 + 0/5 \Rightarrow z = 3 \ , \ w = 0/5$

$Q_1 = (1-w)x_z + wx_{z+1} = (1-0/5)(2) + 0/5(2) = 2$

ج) $P = 0/3 \Rightarrow (n+1)P = (13+1)(0/3) = 4/2 = 4 + 0/2 \Rightarrow z = 4 \ , \ w = 0/2$

$D_3 = (1-0/2)(2) + 0/2(3) = 2/2$

د) $P = 0/71 \Rightarrow (n+1)P = (13+1)(0/71) = 9/94 = 9 + 0/94 \Rightarrow z = 9 \ , \ w = 0/94$

$\Rightarrow P_{71} = (1-0/94)(4) + 0/94(5) = 4/94$

مثال: برای داده‌های زیر:

64 , 23 , 18 , 33 , 47 , 92 , 81 , 5 , 10

محل و مقدار هر یک از عبارت‌های خواسته شده را به‌دست آورید.

الف) چارک سوم　　　ب) دهک ششم　　　ج) صدک سی و نهم

ابتدا داده‌ها را به‌صورت صعودی مرتب می‌کنیم:

5 , 10 , 18 , 23 , 33 , 47 , 64 , 81 , 92

سپس به محاسبه عبارت‌های فوق می‌پردازیم:

الف) چارک سوم = صدک هفتاد و پنجم $\Rightarrow \quad P=0/75$

محل چندک $= nP+\frac{1}{2} \quad \Rightarrow$ محل چارم سوم (Q_3) $=0/75(9)+\frac{1}{2}=7/25$

$(n+1)P=(9+1)(0/75)=7/5=7+0/5 \quad \Rightarrow \quad z=7 \quad , \quad w=0/5$

$Q_3=(1-w)x_z+wx_{z+1}=(1-0/5)(64)+(0/5)(81)=72/5$

ب) دهک ششم = صدک شصتم $\Rightarrow \quad P=0/6$

محل دهک ششم (D_6) $=(0/6)(9)+\frac{1}{2}=5/9$

$(n+1)P=(9+1)(0/6)=6 \quad \Rightarrow \quad z=6 \quad , \quad w=0 \quad \Rightarrow \quad D_6=x_6=47$

ج) صدک سی و نهم $\Rightarrow \quad P=0/39$

محل صدک سی و نهم (P_{39}) $=(0/39)(9)+\frac{1}{2}=4/01$

$(n+1)P=(9+1)(0/39)=3/9 \quad \Rightarrow \quad z=3 \quad , \quad w=0/9$

$P_{39}=(1-0/9)x_3+0/9(x_4)=0/1(18)+0/9(23)=22/5$

تست: با توجه به نمودار ساقه و برگ زیر اختلاف چارک سوم و اول کدام است؟

ساقه	برگ
0	3　6　9
1	0　1　2　5　5　8
2	1　5　6

1) 7　　　2) 11

3) 7/5　　　4) 11/5

با توجه به نمودار فوق می‌توان داده‌ها را به‌صورت زیر مرتب کرد:

$$3\ ,\ 6\ ,\ 9\ ,\ 10\ ,\ 11\ ,\ 12\ ,\ 15\ ,\ 15\ ,\ 18\ ,\ 21\ ,\ 25\ ,\ 26$$

صدک بیست و پنجم = چارک اول $\Rightarrow\ P = 0/25$

صدک هفتاد و پنجم = چارک سوم $\Rightarrow\ P = 0/75$

چارک اول $\Rightarrow\ (n+1)P = (12+1)(0/25) = 3/25 \Rightarrow z = 3\ ,\ w = 0/25$

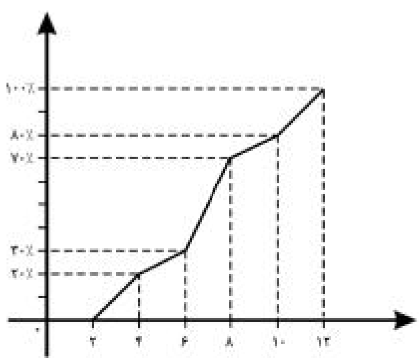
$$Q_1 = (1-0/25)(9) + (0/25)(10) = 9/25$$

چارک سوم $\Rightarrow\ (n+1)P = (13)(0/75) = 9/75 \Rightarrow z = 9\ ,\ w = 0/75$

$$Q_3 = (1-0/75)(18) + (0/75)(21) = 20/25$$

$$Q_3 - Q_1 = 20/25 - 9/25 = 11$$

تست: در نمودار فراوانی تجمعی نسبی زیر صدک ۸۴ ام کدام است؟



1) 10/5

2) 11

3) 10/4

4) 11/1

صدک هشتاد و چهارم عددی است که 84% داده‌ها از آن کوچکتر می‌باشند بنابراین روی محور **y**ها عدد 84% را مشخص می‌کنیم و از آن خطی می‌کشیم که نمودار را قطع کند تصویر این نقطه روی محور **x**ها همان صدک 84 ام می‌باشد که برابر است با:

$\%84 \Rightarrow$ (80−100) قسمت بازه 4 $\rightarrow \dfrac{4}{20} = \dfrac{1}{5}$

$\dfrac{1}{5}$ بازه (10−12) $\rightarrow 2 \times \dfrac{1}{5} = 0/4$

$$P_{84} = 10 + 0/4 = 10/4$$

مثال: از داده‌های زیر چارک سوم، دهک هفتم و صدک چهل و هفتم را محاسبه کنید.

حدود طبقات	20_40	40_60	60_80	800_100	100_120	120_140	140_160	160_180	180_200
فراوانی‌ها	6	9	11	14	20	15	10	8	7

در ابتدا ردیف فراوانی‌های تجمعی را به‌دست می‌آوریم:

$C-L$	20_40	40_60	60_80	800_100	100_120	120_140	140_160	160_180	180_200
F_i	6	9	11	14	20	15	10	8	7
F_{ci}	6	15	26	40	60	75	85	93	100

الف) $P=0/75$ در چارک سوم $\Rightarrow \quad NP=100\times 0/75=75$

پس طبقه ششم طبقه Q_P می‌باشد و داریم:

$$Q_P=L_i+\left(\frac{NP-F_{ci-1}}{f_i}\right)\times C \quad \Rightarrow \quad Q_{0/75}=120+\left(\frac{75-60}{15}\right)(20)$$

(چارک سوم) $Q_{0/75}=140$

ب) $P=0/7$ در دهک هفتم $\Rightarrow \quad NP=100\times 0/7=70$

پس طبقه ششم طبقه Q_P می‌باشد بنابراین:

$$Q_{0/7}=120+\left(\frac{70-60}{15}\right)(20) \quad \Rightarrow$$ (دهک ششم) $Q_{0/7}\ ;\ 133/33$

ج) $P=0/47$ در صدک چهل و هفتم $\Rightarrow \quad NP=100\times 0/47=47$

پس طبقه پنجم طبقه Q_P می‌باشد بنابراین:

$$Q_{0/47}=100+\left(\frac{47-40}{20}\right)(20) \quad \Rightarrow$$ (صدک چهل و هفتم) $Q_{0/47}=107$

مثال: در داده‌های جدول زیر مقادیر چارک اول، دهک هفتم و صدک هشتاد و یکم را تعیین کنید.

حدود طبقات	1_4	4_7	7_10	10_13
فراوانی مطلق	6	9	4	11

ابتدا ردیف فراوانی‌های تجمعی را تشکیل می‌دهیم:

$C-L$	$1-4$	$4-7$	$7-10$	$10-13$
f_i	6	9	4	11
F_{ci}	6	15	19	30

$P=0/25$ در چارک اول $\Rightarrow \quad NP=30\times 0/25=7/5$

پس طبقه دوم طبقه Q_P می‌باشد و:

$$Q_1 \text{ یا } Q_{0/25} = 4 + \left(\frac{7/5 - 6}{9}\right)(3) \Rightarrow Q_{0/25} = 4/5$$

ب) $P = 0/7$ در دهک هفتم $\Rightarrow NP = (0/7)(30) = 21$

پس طبقه چهارم طبقه Q_P است و:

$$D_7 \text{ یا } Q_{0/7} = 10 + \left(\frac{21-19}{11}\right)(3) \simeq 10/55$$

ج) $P = 0/81$ در صدک هشتاد و یکم $\Rightarrow NP = (0/81)(30) = 24/3$

پس طبقه چهارم طبقه Q_P است بنابراین:

$$P_{81} \text{ یا } Q_{0/81} = 10 + \left(\frac{24/3 - 19}{11}\right)(3) \simeq 11/44$$

تست: درآمد کارمندان یک فروشگاه زنجیره‌ای به شرح جدول زیر است:

$x_i - x_{i+1}$	10–20	20–30	30–40	40–50	50–60	60–70
N_i	10	40	60	60	40	10

حاصل عبارت $\frac{Q_3 - Q_1}{2}$ تقریباً کدام است؟

1) 9/16   2) 10   3) 18/32   4) 20

ابتدا ردیف فراوانی تجمعی را به‌دست می‌آوریم:

F_{ci}	10	50	110	170	210	220

$Q_1 \Rightarrow NP = (0/25)(220) = 55 \Rightarrow$ طبقه سوم طبقه Q_1 است.

$$Q_1 = 30 + \left(\frac{55-50}{60}\right)(10) \Rightarrow Q_1 \simeq 30/84$$

$Q_3 \Rightarrow NP = (0/75)(220) = 165$ طبقه چهارم طبقه Q_3 است

$$Q_3 = 40 + \left(\frac{165-110}{60}\right)(10) \Rightarrow Q_3 \simeq 49/17$$

$$\frac{Q_3 - Q_1}{2} = \frac{49/17 - 30/84}{2} \simeq 9/16$$

پس گزینه اول صحیح است.

مشخص کننده‌های پراکندگی

پس از تعیین شاخص‌های مرکزی مایلیم بدانیم که تجمع داده‌ها حول شاخص مرکزی چگونه است. آیا داده‌ها خیلی به هم نزدیک هستند یا از یکدیگر دور می‌باشند، شاخص‌هایی که این ویژگی داده‌ها را اندازه‌گیری می‌کنند مشخص کننده‌های پراکندگی (شاخص‌های پراکندگی) نام دارند. برای نشان دادن این که مشخص‌کننده‌های مرکزی به تنهایی قادر به توصیف داده‌ها نیستند به مثال زیر می‌پردازیم:

مثال: فرض کنید اعداد زیر نمرات دو کلاس در یک آزمون آمار می‌باشند:

کلاس **A**: 1, 2, 3, 4, 6, 8

کلاس **B**: **o**, 1, 2, 4, 4, 8, 9

مشاهده می‌شود که میانگین، میانه و مد هر دو جمعیت برابر عدد 4 است، در حالی که اختلاف و پراکندگی داده‌های کلاس **B** بیشتر است و شاخص‌های مرکزی این تفاوت را بیان نمی‌کنند. بنابراین معرفی مشخص کننده‌های پراکندگی که بیانگر این نوع اختلاف‌ها هستند ضرورت دارد. معروف‌ترین مشخص‌کننده‌های پراکندگی عبارتند از: دامنه تغییرات، دامنه چارک‌ها، انحراف چارک‌ها (نیم دامنه چارکی)، انحراف از میانگین، واریانس (پراش)، انحراف معیار (انحراف استاندارد) و گشتاورها که به شرح زیر می‌باشند.

دامنه تغییرات

دامنه تغییرات عبارت است از تفاضل کوچکترین داده از بزرگترین داده که با حرف **R** نمایش می‌دهیم، کم اهمیت‌ترین پارامتر پراکندگی دامنه تغییرات است، چون تنها تابعی از دو مقدار کوچک و بزرگ می‌باشد و وضعیت اعداد وسط را مشخص نمی‌کند یا به عبارت دیگر چگونگی توزیع اعداد را مشخص نمی‌کند.

در داده‌های گسسته: **R** = بزرگترین داده – کوچکترین داده

در داده‌های پیوسته: **R** = بزرگترین داده – کوچکترین داده **+y**

y = تقریب مورد نظر

تست: کوچکترین و بزرگترین داده‌های آماری به ترتیب ۱۱ و α می‌باشند، اگر به این داده‌ها چهار عدد ۱۹، ۶، ۳۳ و ۸۷ اضافه شود دامنه تغییرات ۱۲ واحد اضافه می‌گردد در این صورت α کدام است؟

1) 66   2) 86   3) 80   4) 65

حل: اختلاف بین عدد 11 و 6، پنج می‌باشد در صورتی که به دامنه تغییرات 12 واحد افزوده شده است بنابراین عدد α

از 87 کوچکتر است پس:

$R=\alpha-11$ دامنه تغییرات اولیه

$R'=87-6=81$ دامنه تغییرات ثانویه

$\Rightarrow R+12=R'\Rightarrow \alpha-11+12=81\Rightarrow \alpha=80$

تست: در یک کلاس درس طول قد ۱۵ نفر از دانشجویان اندازه‌گیری شده و نتایج زیر برحسب سانتی‌متر به دست آمده است دامنه تغییرات اطلاعات مربوط به قد دانشجویان برابر است با:

1) 34/01　　2) 34　　3) 34/02　　4) 34/11

$1-150/05$	$5-148$	$9-177/3$
$2-161/17$	$6-182/01$	$10-179/36$
$3-173/56$	$7-163/5$	$11-169/4$
$4-174/1$	$8-169/21$	$12-157/83$

حل: داده‌های موردنظر از نوع داده‌های پیوسته می‌باشند و همچنین با استفاده از اعداد داده شده می‌توان فهمید که تقریب موردنظر $\gamma=0/01$ است بنابراین:

$R=182/01-148+0/01\Rightarrow R=34/02$

نکات مهم در ارتباط با دامنه تغییرات

1- اگر همه داده‌های آماری را با یک عدد جمع یا تفریق کنیم دامنه تغییرات تغییر نمی‌کند.

2- اگر همه داده‌های آماری را در یک عدد ضرب یا بر آن عدد تقسیم کنیم، دامنه تغییرات نیز در آن عدد ضرب یا بر آن عدد تقسیم می‌شود.

3- اگر همه داده‌ها با هم برابر باشند دامنه تغییرات برابر صفر می‌شود و برعکس، یعنی اگر دامنه تغییرات برابر صفر باشد همه داده‌ها با هم برابرند.

تست: اگر دامنه تغییرات داده‌های آماری $x_1, x_2, \dots, x_n$ برابر صفر باشد، میانگین داده‌های آماری $2x_1+1, 2x_2+1, \dots, 2x_n+1$ کدام است؟

1) $\frac{x_1+x_n}{2}$　　2) $(x_1, x_2, \dots, x_n)+1$　　3) x_n　　4) $2x_1+1$

حل: دامنه تغییرات داده‌های آماری برابر صفر است پس: $x_1=x_2=\dots=x_n$ از آن جا که همه داده‌های آماری برابر هستند نتیجه می‌گیریم که میانگین آن‌ها با خود داده‌ها برابر است $\bar{x}=x_1=x_2=\dots=x_n$ در نتیجه میانگین داده‌های آماری خواسته شده برابر می‌شود با:

$2\overline{x}+1=2x_1+1$

تست: اگر دامنه تغییرات داده‌های ۴، $c-1$ ،$2b$ و a^3+3 برابر صفر باشد، دامنه تغییرات داده‌های $16c-3$، $16b-3$ و $16a-3$ کدام است؟

1) 64   2) 61   3) صفر   4) 48

حل: دامنه تغییرات داده‌های موردنظر صفر می‌باشد بنابراین:

$$a^3+3=2b=c-1=4\Rightarrow\begin{cases}a^3+3=4\Rightarrow a=1\\2b=4\Rightarrow b=2\\c-1=4\Rightarrow c=5\end{cases}$$

دامنه تغییرات داده‌ها **c**، **b** و **a** برابر 4 می‌باشد پس دامنه تغییرات داده‌ها $16c-3$، $16b-3$ و $16a-3$ برابر $R=16\times4=64$ است.

دامنه چارک‌ها

اگر میانه را معیار قرار دهیم و دامنه تغییرات را از پایین تا $\frac{1}{4}$ از بالا تا $\frac{3}{4}$ گسترش دهیم، دامنه چارک‌ها به دست می‌آید که آن را با نماد **IQR** نمایش می‌دهیم و برابر است با:

$$IQR=Q_3-Q_1$$

نکته: اگر داده‌ها خیلی بزرگ یا خیلی کوچک باشند، دامنه تغییرات وضعیت دو داده بزرگ و کوچک را بیان می‌کند در صورتی که (Q_3-Q_1) 25% داده‌هایی بالایی و 25% داده‌های پایین را حذف می‌کند (شامل 50% داده‌های متوسط می‌باشد) تا به داده‌های متعادل‌تری برسیم.

انحراف چارک‌ها (نیم دامنه چارکی)

انحراف چارک‌ها را با نماد **QD** نمایش می‌دهیم و برابر است با:

$$QD=\frac{IQR}{2}=\frac{Q_3-Q_1}{2}$$

تست: تعداد فرزندان ۱۶ خانواده تهرانی به ترتیب عبارتند از:

$0,1,1,2,2,2,3,3,3,3,4,4,4,5,6,7$

انحراف چارکی کدام است؟

1 (4   1/5 (3   3 (2   2 (1

حل: $p=0/25: Q_1$ برای محاسبه $\Rightarrow (n+1)P=(16+1)(0/25)=4/25$

$$\Rightarrow z=4 \;,\; w=0/25$$

$$Q_1=(1-w)x_4+wx_5=(0/75)(2)+(0/25)(2)=2$$

$P=0/75: Q_3$ برای محاسبه $\Rightarrow (n+1)P=(17)(0/75)=12/75$

$$\Rightarrow z=12 \;,\; w=0/75$$

$$Q_3=(0/25)x_{12}+(0/75)x_{13}=(0/25)(4)+(0/75)(4)=4$$

$$Q_D=\frac{Q_3-Q_1}{2}=\frac{4-2}{2}=1$$

تست: انحراف چارک‌ها برای داده‌های جدول زیر کدام است؟

اندازه	4-8	8-12	12-16	16-20	20-24
فراوانی	7	10	18	5	15

9/63 (1   4/815 (2

15/515 (3   7/76 (4

حل: در ابتدا جدول توزیع فراوانی تجمعی را به صورت زیر به دست می‌آوریم:

f_i	7	10	18	5	15
F_{ci}	7	17	35	40	55

سپس طبقه **Qp** را به صورت زیر مشخص می‌کنیم:

$P=0/25: Q_{0/25}=Q_1$ برای محاسبه $\Rightarrow NP=55\times 0/25=13/75$

طبقه دوم طبقه Q_1 است.

$P=0/75: Q_{0/75}=Q_3$ برای محاسبه $\Rightarrow NP=55\times 0/75=41/25$

طبقه پنجم طبقه Q_3 است.

$$Q_1=8+\left(\frac{13/75-7}{10}\right)(4)\Rightarrow Q_1=10/7$$

$$Q_3=20+\left(\frac{41/25-40}{15}\right)(4)\Rightarrow Q_3=20/33$$

$$QD=\frac{Q_3-Q_1}{2}=\frac{20/33-10/7}{2}=4/815$$

انحراف از میانگین

بزرگترین عیب دامنه تغییرات و انحراف چارکی این است که به تمام داده‌ها بستگی ندارد، یک پارامتر خوب که این عیب را نـدارد میـانگین انحـراف داده‌هـا از میـانگین اسـت امـا از آن جـا کـه ایـن مقـدار همـواره صـفر اسـت. $\left(\frac{1}{n}\sum x_i - \bar{x}) = \frac{1}{N}\sum f_i(x_i - \bar{x}) = \mathbf{o}\right)$ نمی‌توان از آن به عنوان یک پارامتر پراکندگی اسـتفاده کـرد لـذا یـا مـی‌تـوان میانگین قدرمطلق این انحراف‌ها را حساب کرد یا میانگین مربع این انحراف‌ها را به دست آورد بنابراین انحراف از میانگین یا انحراف متوسط که آن را با نماد **D.A** نمایش می‌دهیم به صورت‌های زیر به دست می‌آید:

برای داده‌های طبقه بندی نشده: $A.D = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

برای داده‌های طبقه بندی شده: $A.D = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - \bar{x}| = \sum_{i=1}^{n} f_{pi}|x_i - \bar{x}|$

تست: انحراف از میانگین داده‌های زیر کدام است؟

$4, 8, 12, 16, 20, 24, 28, 32, 36$

1) 9،1   2) 8   3) 9   4) 8،9

حل: ابتدا میانگین داده‌های زیر را به دست می‌آوریم:

$$\bar{x} = \frac{4+8+\ldots+36}{9} = 20$$

سپس:

$$A.D = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{9}(|4-20| + |8-20| + \ldots + |36-20|)$$

$$A.D = \frac{1}{9}(16+12+8+4+\mathbf{o},4+8+12+16) = \frac{80}{9} \Rightarrow A.D\ ;\ 8/9$$

تست: انحراف از میانگین داده‌های جدول زیر کدام است؟

x_i	4	5	6	7
f_i	20	40	30	10

1) 2   2) صفر

3) 0/76   4) 1/3

حل: در ابتدا میانگین داده‌های فوق را به صورت زیر محاسبه می‌کنیم:

$$\overline{x}=\frac{\sum_{i=1}^{n} f_i x_i}{N}=\frac{f_1x_1+f_2x_2+f_3x_3+f_4x_4}{f_1+f_2+f_3+f_4}=\frac{(4\times 20)+(5\times 40)+(6\times 30)+(7\times 10)}{20+40+30+1}$$

$$\overline{x}=\frac{530}{100}=5/3$$

سپس داریم:

$$A.D=\frac{1}{N}\sum_{i=1}^{n} f_i\left|x_i-\overline{x}\right|$$

$$=\frac{1}{100}\left(20\left|4-5/3\right|+40\left|5-5/3\right|+30\left|6-5/3\right|+\left|7-5/3\right|\right)$$

$$=\frac{1}{100}(26+12+21+17)=0/76$$

نکات مهم در ارتباط با انحراف از میانگین

1- اگر همه داده‌ها با هم برابر باشند انحراف از میانگین آن‌ها برابر صفر است و بالعکس.

2- اگر به تمام داده‌ها عدد ثابتی کم یا ا ضافه کنیم انحراف از میانگین تغییر نمی‌کند.

3- اگر تمام داده‌ها را در عدد ثابتی مانند **a** ضرب یا تقسیم کنیم انحراف از میانگین آن‌ها در قدرمطلق a($|a|$) ضرب یا تقسیم می‌شود.

4- اگر داده‌های $x_1, x_2, ..., x_n$ تشکیل یک تصاعد حسابی (عددی) با قدرنسبت **d** بدهند آنگاه انحراف از میانگین آن‌ها را به صورت زیر نیز می‌توان به دست آورد:

$$A.D=\begin{cases}\dfrac{nd}{4} & \text{n زوج:}\\[2ex] \dfrac{(n^2-1)d}{4n} & \text{n فرد:}\end{cases}$$

تست: انحراف از میانگین اعداد a, b, c, d, e **۲ برابر صفر است، انحـراف از میـانگین اعـداد** $8d, 10e, 6c, 4b, 2a$ **کدام است؟**

1) صفر 2) 5 3) 2 4) 15

حل: از آن جا که انحراف از میانگین اعداد داده شده صفر است بنابراین:

$$a=b=c=d=e=2$$

حال می‌توان نشان داد که اعداد $10e=2$ $8d=16$ $6e=12$ $4b=8$ و $2a=4$ تشکیل یک تصاعد عددی با قدرنسب

$d = 4$ را می‌دهند و چون تعداد جملات این تصاعد فرد است پس:

$$A.D = \frac{nd}{4} = \frac{5 \times 4}{4} = 5$$

واریانس (پراش)

شاخص‌هایی که تاکنون برای سنجش پراکندگی داده‌های جامعه بیان شده‌اند، هیچ کدام قادر بـه بیـان تمـامی تغییـرات نیستند بنابراین باید به دنبال شاخص باشیم که تغییرات کل داده‌ها را اندازه گیری کند، طبیعی است تغییر زمانی مفهوم پیدا می‌کند که هر یک از داده‌ها نسبت به یک مبدأ مقایسه شوند بهترین مرکز برای داده‌ها میانگین است. ایـن شـاخص که یکی از مهمترین شاخص‌ها در علم آمار می‌باشد واریانس یا پراش نام دارد، آن را بـا نمـاد σ^2 (بـرای جامعـه) یـا S^2 (برای نمونه) نمایش می‌دهیم و به صورت‌های زیر به دست می‌آید:

برای داده‌های طبقه بندی نشده:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\overline{x})^2$$

برای داده‌های طبقه‌بندی شده:

$$\begin{cases} \sigma^2 = \dfrac{1}{N}\sum_{i=1}^{n} f_i(x_i - \overline{x})^2 = \dfrac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - (\overline{x})^2 \\ \text{یا} \\ \sigma^2 = \sum_{i=1}^{n} f_{pi}(x_i - \overline{x})^2 = \sum_{i=1}^{n} f_{pi} x_i^2 - (\overline{x})^2 \end{cases}$$

تست: نسبت واریانس مقادیر $4,8,10,14,18$ **به واریانس مقادیر** $2,4,5,7,9$ **کدام است؟**

1) $\frac{1}{4}$ 2) 2 3) 4 4) $\frac{1}{2}$

حل: ابتدا میانگین اعداد داده شده را به دست می‌آوریم:

$$\overline{x}_1 = \frac{4+8+10+14+18}{5} = \frac{54}{5} = 10/8$$

$$\overline{x}_2 = \frac{2+4+5+7+9}{5} = \frac{27}{5} = 5/4$$

سپس به محاسبه واریانس آن‌ها می‌پردازیم:

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{5}\sum_{i=1}^{5}(x_i - 10/8)^2$$

$$S_1^2 = 0/2\left[(4-10/8)^2 + (8-10/8)^2 + (10-10/8)^2 + (14-10/8)^2 + (18-10/8)^2\right]$$

$$S_1^2 = 0/2(46/24 + 7/84 + 0/64 + 10/24 + 51/84) = 23/36$$

$$S_2^2 = \frac{1}{5}\sum_{i=1}^{5}(x_i - 5/4)^2$$

$$S_2^2 = 0/2\left[(2-5/4)^2 + (4-5/4)^2 + (5-5/4)^2 (7-5/4)^2 + (9-5/4)^2\right]$$

$$S_2^2 = 0/2(11/56 + 1/96 + 0/16 + 2/56 + 12/96) = 5/84 \Rightarrow \frac{S_1^2}{S_3^2} = \frac{23/36}{5/84} = 4$$

البته قابل ذکر است که به تست فوق می‌توان به سادگی با استفاده از روابط واریانس پاسخ داد که به آن خواهیم پرداخت.

تست: واریانس داده‌های زیر را به دست آورید.

x_i	-1	o	1	2
f_i	2	3	4	1

0/76 (1   0/78 (2

0/82 (3   0/84 (4

حل:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(-1\times 2) + (o\times 3) + (1\times 4) + (2\times 1)}{2+3+4+1} = 0/4$$

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N} = \frac{2(-1-0/4)^2 + 3(o-0/4)^2 + 4(1-0/4)^2 + 1(2-0/4)^2}{10}$$

$$\sigma^2 = \frac{2(1/96) + 3(0/16) + 4(0/36) + (2/56)}{10} \Rightarrow \sigma^2 = 0/84$$

تست: واریانس جدول زیر کدام است؟

اندازه	4-8	8-12	12-16	16-20	20-24	24-28
فراوانی	6	10	18	30	15	12

20/6 (1   41/2 (2

15/16 (3   30/3 (4

حل: در ابتدا مراکز دسته‌ها را مشخص می‌کنیم:

x_i	6	10	14	18	22	26
f_i	6	10	18	30	15	12

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{(6\times 6) + (10\times 10) + (14\times 18) + (18\times 30) + (22\times 15) + (26\times 12)}{6+10+18+30+15+12} = \frac{1570}{91}$$

$$\bar{x} = 17/25$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} f_i x_i^2 - (\bar{x})^2$$

$$= \frac{1}{91}(6\times 6^2 + 10\times 10^2 + 18\times 14^2 + 30\times 18^2 + 15\times 22^2 + 12\times 26^2) - (17/25)^2$$

$$\sigma^2 = \frac{1}{91}(216 + 1000 + 3528 + 9720 + 7260 + 8112) - (297/56) \Rightarrow \sigma^2 \; ; \; 30/3$$

نکات مربوط به واریانس

1- هرچه واریانس به صفر نزدیکتر باشد پراکندگی بین داده‌ها کمتر است.

2- اگر تمام داده‌ها با هم برابر باشند واریانس صفر است و بالعکس.

3- اگر همه داده‌ها را با یک عدد ثابت جمع یا تفریق کنیم واریانس تغییر نمی‌کند.

4- اگر تمام داده‌ها را در عدد ثابتی مانند **a** ضرب یا تقسیم کنیم واریانس آن‌ها در a^2 ضرب یا تقسیم می‌شود.

5- اگر داده‌های $x_1, x_2, \ldots, x_n$ تشکیل یک تصاعد حسابی (عددی) با قدر نسبت **d** دهند آنگاه:

$$S^2 = \frac{(n^2-1)d^2}{12}$$

تست: اگر مد صفتی در افراد یک جامعه مثبت باشد و این مقدار مد را از هر یک از داده‌ها کم کنیم:

1) واریانس افزایش می‌یابد. 2) واریانس تغییر نمی‌کند.

3) واریانس کاهش می‌یابد. 4) واریانس در مقدار مد ضرب می‌شود.

حل: با توجه به این که مد یک عدد ثابت است اگر از تمام داده‌ها مقدار مد را کم کنیم تأثیری روی واریانس نـدارد پـس گزینه «2» صحیح می‌باشد.

تست: واریانس داده‌های $x+4$ ، $x+4$ ، x **و** x **کدام است؟**

1) 5 2) 4 3) $\frac{\sqrt{5}}{2}$ 4) 25

حل: اگر از تمام داده‌ها عدد ثابتی مانند **x** را کم کنـیم واریـانس داده‌هـا تغییـر نمـی‌کنـد. داده‌هـای جدیـد عبارتنـد از $0,0,4,4$ که واریانس برابر است با:

$$\bar{x} = \frac{0+0+4+4}{4} = 2$$

$$S^2 = \frac{2(0-2)^2 + 2(4-2)^2}{4} = \frac{8+8}{4} = 4$$

البته قابل ذکر است که واریانس را می‌توان از روش کلی که قبلاً بیان شده به دست آورد.

تست: نسبت واریانس مقادیر $4,8,10,14,18$ **به واریانس مقادیر** $3,5,6,8,10$ **کدام است؟**

1) $\frac{1}{4}$ 2) 2 3) 3 4) $\frac{1}{2}$

حل: با توجه به داده‌های فوق می‌توان نوشت:

$$\begin{matrix} x_i : 2,8,10,14,18 \\ y_i : 3,5,6,8,10 \end{matrix} \Rightarrow x_i = 2y_i - 2 \Rightarrow S_x^2 = (2)^2 S_y^2 \Rightarrow \frac{S_x^2}{S_y^2} = 4$$

همان طور که می‌دانید جمع یا تفریق عدد ثابت با داده‌ها تأثیری در واریانس ندارد در صورتی که ضرب یـا تقسـیم عـدد ثابت **a** در تمام داده‌ها باعث ضرب عدد $(a)^2$ در واریانس آن‌ها می‌شود بنابراین می‌توان نشان داد که:

$$y = ax + b \Rightarrow S_y^2 = (a)^2 S_x^2$$

تست: میانگین قد دانش آموزان مدرسه‌ای ۱۲۰cm و واریانس قد آن‌ها $100cm^2$ است اگر هر فرد ۱۴٪ قد خـود رشد کند، میانگین و واریانس قد دانش آموزان چقدر خواهد شد؟

1) 100، 120 2) 120، 114 3) 136/8، 114 4) 136/8، 129/96

حل: با توجه به این که هر فرد 14% قد خودش رشد کرده داریم:

قد جدید = قد قدیم + 0/14 (قد قدیم) $\Rightarrow y = x + 0/14x = 1/14x$

$$\bar{y} = 1/14(\bar{x}) = 1/14(120) = 136/8$$

$$\sigma_y^2 = (1/14)^2 \sigma_x^2 = (1/14)^2 (100) = 129/96$$

تست: هشت داده آماری با میانگین ۱۵ و واریانس ۴ مفروض است، اگر دو داده ۱۲ و ۱۸ به آن‌ها اضـافه شـود واریانس ۱۰ داده حاصل کدام است؟

1) 4 2) 4/5 3) 4/8 4) 5

حل:

میانگین داده‌های قدیم $x_1 = \frac{\sum x_i}{n} \Rightarrow 15 = \frac{\sum_{i=1}^{8} x_i}{8} \Rightarrow \sum_{i=1}^{8} x_i = 120$

واریانس داده‌های قدیم $S_1^2 = \frac{\sum (x_i - \bar{x})^2}{n} \Rightarrow 4 = \frac{\sum_{i=1}^{8} (x_i - \bar{x})^2}{8}$

$$\Rightarrow \sum_{i=1}^{8} (x_i - \bar{x})^2 = 32$$

مجموع داده‌های جدید $= 12 + 18 + \sum_{i=1}^{8} x_i = 12 + 18 + 120 = 150$

میانگین داده‌های جدید $\overline{x}_2=\frac{150}{10}=15$

واریانس داده‌های جدید $S_2^2=\frac{\sum_{i=1}^{10}(x_i-\overline{x})^2}{n}=\frac{\sum_{i=1}^{8}(x_i-\overline{x})^2+(18-15)^2+(12-15)^2}{10}$

$S_2^2=5$

تست: مجموع مجذورات ۱۱ داده آماری برابر ۲۲۰۰ و میانگین این داده‌ها برابر ۱۴ است واریانس کدام است؟

1) 2   2) 3   3) 4   4) 5

حل: $n=11\ ,\ \sum x_i^2=2200\ ,\ \overline{x}=14$

$S^2=\frac{1}{n}\sum x_i^2-(\overline{x})^2=\frac{1}{11}(2200)-(14)^2=4$

تست: واریانس اعداد $0,5,10,15,20,25$ کدام است؟

1) 36/46   2) 14/58   3) 7/29   4) 72/92

حل: 6 عدد فوق تشکیل یک تصاعد عددی با قدر نسبت $d=5$ را می‌دهند بنابراین:

$S^2=\frac{(n^2-1)d^2}{12}=\frac{(6^2-1)(5)^2}{12}=\frac{35\times5^2}{12}=72/92$

انحراف معیار

عیب بزرگ واریانس آن است که از جنس داده‌های اولیه نیست یعنی با آن هم واحد نمی‌باشد به عنوان مثال اگـر واحـد داده‌ها برحسب میلی‌متر باشد واحد واریانس میلی متر مربع است. بنابراین هیچ گاه واریانس را با خود متغیـر نمـی‌تـوان مقایسه کرد، برای رفع این مشکل از جذر مثبت واریانس استفاده می‌کنیم و آن را انحـراف معیـار یـا انحـراف اسـتاندارد می‌نامیم و با نماد σ (برای جامعه) S (برای نمونه) نمایش می‌دهیم پس:

σ یا $S=\sqrt{\text{واریانس}}$

تست: جمع نمرات و جمع مربعات نمرات دانشجویان یک کلاس ۲۵ نفری به ترتیب ۴۰۰ و ۶۴۰۰ است انحـراف معیار نمرات این دانشجویان چقدر است؟

1) 16   2) 5   3) صفر   4) 4

حل:
$$S^2=\frac{\sum x_i^2}{n}-(\bar{x})^2=\frac{\sum x_i^2}{n}-\left(\frac{\sum x_i}{n}\right)^2=\frac{640}{25}-\left(\frac{400}{25}\right)^2$$

$$S^2=256-256=\mathbf{0}\Rightarrow S=\mathbf{0}$$

تست: میانگین داده‌های $x_1,x_2,...,x_n$ برابر ۴ و انحراف معیار آن‌ها برابر $\sqrt{3}$ است. میانگین داده‌های $x_1^2,x_2^2,...,x_n^2$ کدام است؟

1) 16 2) 25 3) 7 4) 19

حل:
$$S^2=\frac{\sum x_i^2}{n}-(\bar{x})^2\Rightarrow(\sqrt{3})^2=\frac{\sum x_i^2}{n}-(4)^2\Rightarrow\frac{\sum x_i^2}{n}=3+16$$

$$\Rightarrow\frac{\sum x_i^2}{n}=19$$

نکات مهم در ارتباط با انحراف معیار

1- اگر تمام داده‌ها با هم برابر باشند انحراف معیار صفر است و بالعکس.

2- اگر همه داده‌ها را با یک عدد ثابت جمع یا تفریق کنیم انحراف معیار تغییر نمی‌کند.

3- اگر تمام داده‌ها در عدد ثابتی مانند **a** ضرب و تقسیم کنیم انحراف معیار آن‌ها در $|a|$ ضرب و تقسیم می‌شود.

4- اگر داده‌های $x_1,x_2,...,x_n$ تشکیل یک تصاعد حسابی (عددی) با قدر نسبت **d** دهند آنگاه:

$$S=d\sqrt{\frac{n^2-1}{12}}$$

تست: اگر میانگین و واریانس x به ترتیب ۳ و ۹ باشد، میانگین و انحراف معیار $y=\frac{1}{2}x+1$ به ترتیب کدام است؟

1) 1/5 و 2/5 2) 1/5 و 1/5 3) 2/5 و 2/5 4) 2/5 و 1/5

حل:
$$S_x^2=9\Rightarrow S_x=3$$

$$y=\frac{1}{2}x+1\Rightarrow\begin{cases}\bar{y}=\frac{1}{2}(\bar{x})+1=\frac{1}{2}(3)+1=2/5\\ S_y=\left|\frac{1}{2}\right|S_x=\frac{1}{2}(3)=1/5\end{cases}$$

تست: اگر قیمت اجناس با انحراف معیار $\frac{1}{2}$ طی سال اخیر ۲۰٪ افزایش یابد، واریانس قیمت‌های جدید چقــدر است؟

1) 0/18 2) 0/72 3) 0/36 4) 0/24

حل:

$$S_x=\frac{1}{2}\Rightarrow S_x^2=\frac{1}{4}$$

$$y=x+\%20x=\frac{6}{5}x\Rightarrow S_y^2=\left(\frac{6}{5}\right)^2 S_x^2=\left(\frac{6}{5}\right)^2\frac{1}{4}=0/36$$

تســت: اگــر واریــانس داده‌هــای $3x_1+4,3x_2+4,\ldots,3x_n+4$ برابــر ۸۱ باشــد انحــراف معیــار داده‌هــای $2-x_1,2-x_2,\ldots,2-x_n$ کدام است؟

1) 9 2) 27 3) 1 4) 3

حل: اگر x_i معرف داده‌های اولیه و y_i معرف داده‌های ثانویه باشد بین آن‌ها رابطه زیر برقرار است:

$$x_i=-3y_i+10\Rightarrow S_x^2=(-3)^2S_y^2\Rightarrow S_x^2=9S_y^2$$

$$\Rightarrow 81=9S_y^2\Rightarrow S_y^2=9\Rightarrow S_y=3$$

تست: در صورتی که واریانس و انحراف معیار مقادیر $x_1,x_2,\ldots,x_n,7$ با هم برابر باشند انحراف معیــار مقــادیر $(3x_1-1),(3x_2-1),\ldots,(3x_n-1),20$ کدام است؟

1) 2 2) 3 3) 5 4) 1

حل:

$$S_1^2=S_1\Rightarrow S_1^2-S_1=0\Rightarrow S_1(S_1-1)=0\Rightarrow\begin{cases}S_1=0\\S_1=1\end{cases}$$

تست: انحراف معیار داده‌های $4,7,10,13,16,19,22$ کدام است؟

1) 3 2) 9 3) 6 4) 36

حل: با کمی دقت متوجه می‌شویم که داده‌های فوق تشکیل تصاعد عددی با قدرنسبت $d=3$ را می‌دهد بنابراین:

$$S=d\sqrt{\frac{n^2-1}{12}}=3\sqrt{\frac{49-1}{12}}=3\sqrt{4}=6$$

تست: میانگین پنج مشاهده برابر ۴/۴ و واریانس آن ۸/۲۴ است اگر ۳ تا از این پنج مشاهده $1,2,6$ باشد دوتای دیگر کدامند؟

1) $y=4, x=4$ 2) $y=4, x=9$ 3) $y=9, x=9$ 4) $y=9, x=4$

حل:

$1,2,6,x,y$

$$\bar{x}=\frac{\sum x_i}{n}\Rightarrow 4/4=\frac{1+2+6+x+y}{5}\Rightarrow x+y=13$$

$$S^2=\frac{1}{n}\sum x_i^2-(\bar{x})^2\Rightarrow 8/24=\frac{1}{5}\sum x_i^2-(4/4)^2\Rightarrow \sum x_i^2=138$$

$$\Rightarrow 1^2+2^2+6^2+x^2+y^2=138\Rightarrow x^2+y^2=97$$

$$\begin{cases}x+y=13\\x^2+y^2=97\end{cases}\Rightarrow\begin{cases}x=4\\y=9\end{cases}$$

نکته: برای محاسبه واریانس نمونه از رابطه زیر استفاده می‌کنیم و به آن واریانس ناریب می‌گوییم:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

تست: انحراف‌های مقادیر مشاهده شده از میانگین در ۶ مورد از یک نمونه ۷ تایی به صورت اعداد $4,3,-1,-2,-4,-5$ محاسبه شده انحراف معیار کدام است؟

1) 2 2) 3 3) 4 4) 5

حل: به دلیل اینکه انحراف مقادیر از میانگین تعدادی نمونه داده شده است. پس باید از فرمول واریانس نمونه استفاده کنیم:

$$x_i-\bar{x}: 4,3,-1,-2,-4,-5,k$$

$$\sum(x_i-\bar{x})=\mathbf{o}\Rightarrow 4+3-1-2-4-5+k=\mathbf{o}\Rightarrow k=5$$

$$S^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2=\frac{1}{7-1}\left(4^2+3^2+(-1)^2+(-2^2+(-4)^2+(-5)^2+5^2\right)$$

$$=\frac{1}{6}(16+9+1+4+16+25+25)\Rightarrow S^2=16\Rightarrow S=\sqrt{16}=4$$

اندازه استاندارد

فرض کنید $x_1, x_2, \ldots, x_n$ به ترتیب با فراوانی‌های $f_1, f_2, \ldots, f_n$ یک سری داده‌های nتایی با میانگین $\bar{x}$ و انحراف معیار $\mathbf{S}$ باشند از تبدیل:

$$z_i = \frac{x_i - \bar{x}}{S} \quad : \quad i = 1, 2, \ldots, n$$

داده‌های جدید $z_1, z_2, \ldots, z_n$ به ترتیب با فراوانی‌های $f_1, f_2, \ldots, f_n$ به نام اندازه‌های استاندارد (نمره استاندارد) به‌دست می‌آیند. نمره استاندارد چنین بیان می‌کند که داده $\mathbf{x}$ برحسب این‌که علامت آن منفی یا مثبت باشد به اندازه چند انحراف معیار در پایین یا در بالای میانگین قرار دارد.

نکته: میانگین داده‌های استاندارد صفر و واریانس آن‌ها برابر یک است.

نکته: با توجه به این‌که داده‌های استاندارد، مستقل از واحد اندازه‌گیری می‌باشند از آن‌ها جهت مقایسه توزیع‌های مختلف استفاده می‌شود.

مثال: در آزمون کاردانی به کارشناسی داوطلبی در درس آمار نمره 68 و در درس ریاضی نمره 72 کسب کرده است. اگر میانگین و انحراف معیار درس آمار 58 و 10 و میانگین و انحراف معیار درس ریاضی 60 و 16 باشد وضعیت این داوطلب در کدام درس بهتر است؟

نمره استاندارد مربوط به درس آمار : $z_1 = \frac{68-58}{10} = 1$

نمره استاندارد مربوط به درس ریاضی : $z_2 = \frac{72-60}{16} = 0/75$

اگر چه نمره درس ریاضی او از نمره درس آمار بیشتر است ولی وضعیت وی در درس آمار بهتر از درس ریاضی است زیرا در درس آمار به اندازه یک انحراف معیار بالای میانگین قرار دارد. در حالی که در درس ریاضی به اندازه 0/75 انحراف معیار در بالای میانگین قرار دارد.

گشتاورها

در هر جامعه داده‌ها اغلب حول یک نقطه به نام مرکز جامعه پراکنده می‌شوند میانگین توان rام x_iها و $\left(x_i - \bar{x}\right)$ها را به ترتیب گشتاور مرتبه rام و گشتاور مرکزی مرتبه rام داده‌ها می‌نامند و به‌صورت زیر محاسبه می‌شوند:

الف) گشتاور مرتبه r ام حول عدد a

داده‌های طبقه‌بندی نشده: $m_r(a) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - a\right)^r$ یا $m_r(a)$

داده‌های طبقه‌بندی شده: $m_r(a) = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - a\right)^r$ یا $m_r(a)$

ب) گشتاور مرکزی مرتبه r ام

داده‌های طبقه‌بندی نشده: $m_r = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^r$ یا m_r

داده‌های طبقه‌بندی شده: $m_r = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \overline{x}\right)^r$ یا m_r

نکته: با توجه به تعریف گشتاور، گشتاور مرتبه **r** ام حول مبدأ به‌صورت زیر تعریف می‌شود:

داده‌های طبقه‌بندی نشده: $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ یا m'_r

داده‌های طبقه‌بندی شده: $m'_r = \frac{1}{N}\sum_{i=1}^{n} f_i\, x_i^r$ یا m'_r

روابط مهم در ارتباط با گشتاورها

1) $r=1 \;\Rightarrow\; m_1 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right) = 0 \quad , \quad m_1 = \overline{x}$

2) $r=2 \;\Rightarrow\; m_2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = S^2 \quad , \quad m_2 = m'_2 - \left(m'_1\right)^2$

3) $r=3 \;\Rightarrow\; m_3 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3 \;\Rightarrow\; m_3 = m'_3 - 3m'_2\, m'_1 + 2\left(m'_1\right)^3$

مثال: برای داده‌های زیر گشتاورهای مرکزی مرتبه اول، دوم و سوم و همچنین گشتاورهای مرکزی مرتبه اول و دوم و سوم حول مبدأ مختصات را به‌دست آورید. $4\,,6\,,9\,,11\,,4\,,8$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(4+4+6+8+9+11) = 7$$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right) = \frac{1}{6}\left((4-7)+(4-7)+(6-7)+(8-7)+(9-7)+(11-7)\right)$$

$$=\frac{1}{6}(-3-3-1+1+2+4)=0$$

$$m_2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{1}{6}\left((4-7)^2+(4-7)^2+(6-7)^2+(8-7)^2+(9-7)^2+(11-7)^2\right)$$

$$=\frac{1}{6}(9+9+1+1+4+16)=\frac{40}{6}=6/67$$

$$m_3=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^3=\frac{1}{6}\left((4-7)^3+(4-7)^3+(6-7)^3+(8-7)^3+(9-7)^3+(11-7)^3\right)$$

$$=\frac{1}{6}(-27-27-1+1+8+64)=\frac{18}{6}=3$$

$$m'_1=\overline{x}=7$$

$$m'_2=\frac{1}{n}\sum_{i=1}^{n}x_i^2=\frac{1}{6}\left(4^2+4^2+6^2+8^2+9^2+11^2\right)=55/67$$

$$m'_3=\frac{1}{n}\sum_{i=1}^{n}x_i^3=\frac{1}{6}\left(4^3+4^3+6^3+8^3+9^3+11^3\right)=486$$

تست: **در جدول توزیع فراوانی زیر، گشتاور مرتبه اول حول مبدأ کدام است؟**

x_i	3	5	7	9
f_i	10	10	50	30

1) 5
2) 6
3) 8
4) 7

$$\overline{x}=\frac{1}{N}\sum_{i=1}^{n}f_i x_i=\frac{1}{100}(3\times10+5\times10+7\times50+9\times30)=7$$

گشتاور مرتبه اول حول مبدأ همان میانگین است؛ پس: $m'_1=7$

تست: **گشتاور مرکزی مرتبه دوم داده‌های جدول زیر کدام است؟**

حدود دسته‌ها	10_16	16_22	22_28	28_34	34_40
فراوانی‌ها	4	8	16	10	2

1) 24/7
2) 37/71
3) 50/28
4) 30/19

ابتدا مرکز طبقات را به‌دست می‌آوریم و سپس میانگین داده‌ها را حساب می‌کنیم:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i\, x_i = \frac{1}{40}\left(4\times13 + 8\times19 + 16\times25 + 10\times31 + 2\times37\right)$$

$$= \frac{1}{40}\left(52 + 152 + 400 + 310 + 74\right) = 24/7$$

$$m_r = \frac{1}{n}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^r \quad \Rightarrow \quad m_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2$$

$$m_2 = \frac{1}{40}\left(4(13-24/7)^2 + 8(19-24/7)^2 + 16(25-24/7)^2 + 10(31-24/7)^2 + 2(37-24/7)^2\right)$$
$$= 37/71$$

تست: **سه گشتاور اول یک توزیع حول متغیری برابر با** 2**، به ترتیب** 1، 16 **و** 40− **می‌باشد. گشتاور مرکزی مرتبه سوم کدام است؟**

1) 15   2) -88   3) 10   4) -86

با توجه به فرمول‌های گشتاور می‌توان نوشت:

$$m'_1 = 1 \quad , \quad m'_2 = 16 \quad , \quad m'_3 = -60$$

$$m_3 = m'_3 - 3m'_1 m'_2 + 2\left(m'_1\right)^3 = -40 - 3(1)(16) + 2(1)^3 = -40 - 48 + 2$$

$$m_3 = -86$$

تست: **چهار گشتاور اول یک متغیر حول ۵ به ترتیب برابر ۲، ۲۰، ۴۰ و ۴۰ می‌باشند گشتاور مرکزی مرتبه چهارم کدام است؟**

1) 152   2) 160   3) 144   4) 168

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 = \frac{1}{n}\sum\left(x_i^4 - 4x_i^3\bar{x} + 6x_i^2\bar{x}^2 - 4x_i\bar{x}^3 + \bar{x}^4\right)$$

$$m_4 = \frac{1}{n}\sum x_i^4 - \frac{4}{n}\sum x_i^3\bar{x} + \frac{6}{n}\sum x_i^2\bar{x}^2 - \frac{4}{n}\sum x_i\bar{x}^3 + \frac{1}{n}\sum \bar{x}^4$$

$$m_4 = m'_4 - 4m'_1 m'_3 + 6\left(m'_1\right)^2 m'_2 - 4\left(m'_1\right)^3 m'_1 + \left(m'_1\right)^4$$

$$= 40 - 4(2)(40) + 6(2)^2(20) - 4(2)^4 + (2)^4 = 152$$

تست: گشتاورهای اولیه حول مبدأ برای یک توزیع صفت متغیر به شرح زیر موجود است:

$m'_1 = 2 \quad , \quad m'_2 = 8 \quad , \quad m'_3 = 41$

گشتاورهای مرکزی مراتب اول، دوم و سوم برای توزیع فوق چقدر است؟

1) $m_1 = 0 , m_2 = 6 , m_3 = 8$
2) $m_1 = 0 , m_2 = 6 , m_3 = 0$
3) $m_1 = 0 , m_2 = 4 , m_3 = 9$
4) $m_1 = 1 , m_2 = 2 , m_3 = 3$

$m_1 = 0$

$m_2 = m'_2 - \left(m'_1\right)^2 = 8 - 4 = 4$

$m_3 = m'_3 - 3m'_1 m'_2 + 2\left(m'_1\right)^3 = 41 - 48 + 16 = 9$

مشخص‌کننده‌های نسبی پراکندگی

اگر مایل باشیم پراکندگی بین داده‌های دو یا چند جامعه آماری را که واحد اندازه‌گیری آن‌ها یکسان نیست با هم مقایسه کنیم مقایسه واریانس و انحراف معیار آن‌ها امکان‌پذیر نمی‌باشد یا اگر بخواهیم پراکندگی داده‌های دو یا چند جامعه‌ای که انحراف معیار آن‌ها تقریباً یکسان است ولی میانگین‌های آن‌ها متفاوت می‌باشد را با هم مقایسه کنیم روش‌های قبلی نتایج مطلوبی را در اختیار ما قرار نمی‌دهند. در چنین شرایطی باید از ضریبی استفاده کنیم که یا بعد نداشته باشد یا پراکندگی داده‌ها را نسبت به بزرگی آن‌ها تعدیل کند. برای این منظور از پارامترهای نسبی پراکندگی استفاده می‌کنیم که معروفترین آن‌ها عبارتند از: ضریب تغییرات (ضریب پراکندگی)، ضریب چولگی، ضریب کشیدگی.

ضریب تغییرات

شاخص پراکندگی است که میزان پراکندگی داده‌ها را نسبت به میانگین حسابی مشخص می‌کند. ضریب تغییرات را با نماد **V.C** نشان می‌دهیم و عبارت است از:

ضریب تغییرات $\quad C \cdot V = \dfrac{S}{\overline{x}}$

درصد ضریب تغییرات $\quad C \cdot V = \dfrac{S}{\overline{x}} \times \%100$

تست: برای این‌که بدانیم در سال گذشته میزان درآمد افراد کشورهای کانادا و اسپانیا نسبت به هم چگونه بوده است استفاده ار کدام شاخص آماری مناسبتر است؟

1) ضریب تغییرات　　2) واریانس　　3) انحراف متوسط　　4) انحراف معیار

از آن‌جا که واحد پول این دو کشور با هم متفاوت است پس برای مقایسه آن‌ها باید از ضریب بدون بعد استفاده کنیم در این صورت گزینه صحیح، گزینه 1 می‌باشد.

مثال: در ماه‌های گذشته دونده‌ای به طور میانگین 20 کیلومتر در هفته با انحراف معیار 4 کیلومتر در هفته دویده در حالی که دونده دیگری به طور میانگین 40 کیلومتر در هفته با انحراف معیار 5 کیلومتر در هفته دویده است کدام دونده در میزان دوندگی هفتگی خود نسبتاً بهتر عمل کرده است؟

ضریب تغییرات این دو دونده به ترتیب عبارتند از:

دونده اول : $C \cdot V = \frac{4}{20} \times \%\ 100 = \%20$

دونده دوم : $C \cdot V = \frac{5}{40} \times \%\ 100 = \%12/5$

ضریب تغییرات دونده دوم از دونده اول کوچکتر می‌باشد پس میزان پراکندگی دونده دوم کمتر است در نتیجه دونده دوم در دوندگی هفتگی خود استوارتر بوده است.

تست: ضریب تغییرات داده‌های ۱۴، ۱۲، ۱۰ و ۸ کدام است؟

1) $\frac{\sqrt{5}}{11}$　　2) $\frac{11}{\sqrt{5}}$　　3) $\frac{5}{11}$　　4) $\frac{11}{5}$

$$\overline{x} = \frac{8+10+12+14}{4} = 11$$

$$S^2 = \frac{(8-11)^2 + (10-11)^2 + (12-11)^2 + (14-11)^2}{4} = \frac{9+1+1+9}{4} = 5 \quad \Rightarrow \quad S = \sqrt{5}$$

$$C \cdot V = \frac{S}{\overline{x}} = \frac{\sqrt{5}}{11}$$

نکات مهم در ارتباط با ضریب تغییرات

1ـ اگر همه داده‌ها با هم برابر باشند ضریب تغییرات صفر است.

2ـ اگر همه داده‌ها را در یک عدد مثبت ضرب کنیم ضریب تغییرات تغییر نمی‌کند.

3ـ اگر به همه داده‌ها عددی مثبت اضافه کنیم چون میانگین بزرگتر می‌شود ولی انحراف معیار تغییر نمی‌کند. ضریب تغییرات داده‌های جدید، کوچکتر از ضریب تغییرات داده‌های اولیه می‌شود.

4ـ اگر از تمام داده‌ها عددی مثبت را کم کنیم چون میانگین کوچکتر می‌شود ولی انحراف معیار تغییر نمی‌کند، ضریب تغییرات داده‌های جدید، بزرگتر از ضریب تغییرات داده‌های اولیه می‌شود.

5ـ ضریب تغییرات یک عدد ثابت برابر با صفر است.

6ـ اگر همه داده‌ها را در عددی منفی ضرب کنیم ضریب تغییرات نیز منفی خواهد شد.

تست: **میانگین ۲۰ داده آماری ۱۵ و واریانس آن‌ها برابر ۲/۲۵ است درصد ضریب تغییرات آن‌ها چقدر است؟**

1) 10   2) 12   3) 15   4) 20

$C \cdot V = \frac{S}{\bar{x}} = \frac{\sqrt{2/25}}{15} = \frac{1/5}{15} = 0/1 \quad \Rightarrow$ درصد ضریب تغییرات $= 0/1 \times 100 = 10$

تست: **اگر $\sum x_i = 60$ و $\sum x_i^2 = 400$ و تعداد داده‌ها برابر ۱۰ باشد ضریب تغییرات تقریباً کدام است؟**

1) 0/62   2) 0/33   3) 0/7   4) 0/4

$$\bar{x} = \frac{\sum x_i}{n} = \frac{60}{10} = 6$$

$$S^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \frac{400}{10} - (6)^2 = 40 - 36 = 4 \quad \Rightarrow \quad S = 2$$

$$C \cdot V = \frac{S}{x} = \frac{2}{6} = \frac{1}{3} \; ; \; 0/33$$

تست: **اگر واریانس داده‌های a,b,c,d برابر ۴ و میانگین آن‌ها برابر ۹ باشد ضریب تغییرات $2a,2b,2c,2d$ چقدر است؟**

1) $\frac{2}{9}$ 2)   3) $\frac{8}{9}$ 4) $\frac{4}{3}$

اگر میانگین و واریانس داده‌های اولیه را به ترتیب $\overline{x}$ و S_x^2 فرض کنیم و میانگین و واریانس داده‌های جدید را به ترتیب $\overline{y}$ و S_y^2 در نظر بگیریم می‌توان نوشت:

$$\overline{x}=9 \quad , \quad S_x^2=4 \quad \Rightarrow \quad S_x=2$$

$$y=2x \quad \Rightarrow \quad \begin{cases} \overline{y}=2(\overline{x})=2(9)=18 \\ S_y=|2|\,S_x=2(2)=4 \end{cases}$$

$$C\cdot V_{\text{جدید}}=\frac{S_y}{\overline{y}}=\frac{4}{18}=\frac{2}{9}$$

تست: ضریب تغییرات داده‌های جدول زیر کدام است؟

حدود دسته‌ها	3_5	6_8	9_11
فراوانی	2	1	2

1) $\frac{6}{7\sqrt{5}}$

2) $\frac{\sqrt{5}}{6}$

3) $\frac{7}{6\sqrt{5}}$

4) $\frac{6}{\sqrt{5}}$

ابتدا مرکز دسته‌ها را به‌دست آورده و سپس ضریب تغییرات را محاسبه می‌کنیم:

مراکز دسته‌ها	4	7	10
فراوانی	2	1	2

$$\overline{x}=\frac{(2\times4)+(1\times7)+(2\times10)}{2+1+2}=7$$

$$s^2=\frac{2(4-7)^2+1(7-7)^2+2(10-7)^2}{2+1+2}=\frac{36}{5} \quad \Rightarrow \quad s=\frac{6}{\sqrt{5}}$$

$$C\cdot V=\frac{s}{\overline{x}}=\frac{6/\sqrt{5}}{7}=\frac{6}{7\sqrt{5}}$$

تست: اگر میانگین و واریانس داده‌های $5x_1-3\,,\,5x_2-3\,,\,\mathbf{K}\,,\,5x_n-3$ برابر ۹۷ و ۶۲۵ باشد ضریب تغییرات داده‌های $x_1\,,\,x_2\,,\,\mathbf{K}\,,\,x_n$ کدام است؟

1) %20   2) %25   3) %22/7   4) %20/3

اگر میانگین و انحراف معیار داده‌های اولیه فوق به‌صورت زیر باشد:

$$5\overline{x}-3=97 \Rightarrow 5\overline{x}=100 \Rightarrow \overline{x}=20$$

$$S^2_{(5x-3)}=625 \Rightarrow S_{(5x-3)}=25 \Rightarrow S_x=\left|\frac{1}{5}\right|S_{(5x-3)}=\frac{1}{5}\times 25=5$$

برای محاسبه ضریب تغییرات داده‌های جدید می‌نویسیم:

$$C\cdot V=\frac{S_x}{\overline{x}}=\frac{5}{20}=\frac{1}{4}=0/25$$

تست: متوسط درآمد ماهیانه کارگران کارخانه A ، ۱۷ هزار تومان با واریانس ۴ می‌باشد، در کارخانه B متوسط درآمد ماهیانه ۲۵۰ هزار ریال با واریانس ۹۰۰ می‌باشد:

1) اختلاف درآمد در کارخانه **A** بیش از کارخانه **B** است.

2) اختلاف درآمد در کارخانه **B** بیش از کارخانه **A** است.

3) درآمد اکثر افراد کارخانه **A** کمتر از اکثر افراد کارخانه **B** است.

4) کمترین درآمد در کارخانه **A** بیش از کارخانه **B** است.

برای مقایسه دو جامعه بهتر است از ضریب تغییرات استفاده کنیم چون واحدهای اندازه‌گیری بین دو جامعه یکسان نمی‌باشند:

$$(C\cdot V)_A=\frac{S_A}{\overline{x}_A}\times 100=\frac{\sqrt{4}}{17}\times 100=11/76$$

$$(C\cdot V)_B=\frac{S_B}{\overline{x}_B}\times 100=\frac{\sqrt{900}}{250}\times 100=12$$

به دلیل این‌که $(C\cdot V)_A<(C\cdot V)_B$ می‌توان بیان کرد که پراکندگی در کارخانه **B** بیشتر است و اختلاف درآمد در کارخانه **B** بیش از کارخانه **A** است.

تست: میانگین و انحراف معیار حقوق در یک سازمان به ترتیب ۵۰ هزار تومان و ۲۰ هزار تومان است، اگر حقوق‌ها در این سازمان ۲۵٪ افزایش یابند، ضریب تغییرات حقوق چه خواهد شد؟

1) نصف می‌شود.

2) تغییر نخواهد کرد.

3) چهار برابر خواهد شد.

4) 25 درصد افزایش خواهد یافت.

اگر $\mathbf{x}$ و $\mathbf{y}$ به ترتیب نمایش حقوق اولیه و جدید این سازمان باشد داریم:

$$y = x + 0/25\,x = 1/25\,x$$

$$\begin{array}{l} \overline{y} = 1/25\,\overline{x} \\ S_y = \left|1/25\right| S_x \end{array} \Rightarrow (C \cdot V)_y = \frac{S_y}{\overline{y}} = \frac{1/25\,S_x}{1/25\,\overline{x}} = \frac{S_x}{\overline{x}} = (C \cdot V)_x$$

پس، ضریب تغییرات حقوق تغییری نخواهد کرد.

تست: اگر همه داده‌های آماری را در عدد $\left(-\frac{1}{2}\right)$ضرب کنیم آنگاه نسبت تغییرات داده‌های جدید بـه ضـریب تغییرات ساده‌های اولیه چقدر است؟

1) 2 2) $\frac{1}{2}$ 3) 1 4) -1

اگر همه داده‌ها را در عددی منفی ضرب کنیم میـانگین در همـان عـدد و انحـراف معیـار در قـدرمطلق آن عـدد ضـرب می‌شود؛ پس:

$$(C \cdot V)_{\frac{-1}{2}x} = -(C \cdot V)_x$$

چولگی

طبیعی‌ترین منحنی فراوانی، منحنی فراوانی نرمال استاندارد می‌باشد که معادله مختصاتی آن به‌صورت زیر می‌باشد:

$$y = \frac{1}{\sqrt{2p}} e^{-x^2/2}$$

این منحنی ناقوسی شکل است و از نظر تقارن، کشیدگی، پخی، تناسب، زیبایی خاصی دارد

یک توزیع را متقارن گوییم هرگاه هیستوگرام فراوانی آن به شکل زیر باشد یا نمودار فراوانی چنـد ضـلعی آن وقتـی کـه فاصله طبقات خیلی کوچک باشد به فرم ناقوسی شکل بیان شود. در توزیع متقارن میانگین، میانه و مد بـر هـم منطبـق هستند.

توزیع‌هایی که متقارن نباشند را توزیع‌های چوله می‌نامیم در توزیـع‌هـای چولـه شـاخص‌هـای مرکـزی بـر هـم منطبـق نمی‌باشند ولی همیشه میانه بین مد و میانگین است. اگر دم منحنی در سمت چپ باشد مـی‌گـوییم چولـه بـه چـپ یـا چولگی منفی وجود دارد و اگر مد منحنی در سمت راست باشد می‌گوییم چوله به راست یا چولگی مثبت وجود دارد.

تذکر: اگر توزیع متقارن باشد میانگین، میانه و مد با هم برابرند ولی عکس آن صادق نیست یعنی اگر میانگین، میانه و مد با هم برابر باشند ممکن است توزیع متقارن و یا ممکن است نامتقارن باشد.

تذکر: در چولگی به راست (چولگی مثبت) همواره داریم:

$$Mo < Md < \bar{x}$$

تذکر: در چولگی به چپ (چولگی منفی) همواره داریم:

$$\bar{x} < Md < Mo$$

ضریب چولگی

میزان عدم تقارن منحنی فراوانی را چولگی **(skewness)** یا انحراف از قرینگی می‌گویند. برای تعیین میزان چولگی از ضریب بدون بعدی به نام ضریب چولگی استفاده می‌شود که آن را با SK نمایش می‌دهیم. ضرایب چولگی پیرسون و گشتاوری به‌صورت زیر ارائه می‌شوند:

ضریب چولگی اول پیرسون : $SK_1 = \dfrac{\bar{x} - Mo}{S}$

ضریب چولگی دوم پیرسون : $SK_2 = \dfrac{3(\bar{x} - Md)}{S}$

ضریب چولگی گشتاوری : $SK = \dfrac{m_3}{S^3}$

m_3 : گشتاور مرکزی مرتبه سوم S : انحراف معیار

نکته: اگر چولگی خفیف باشد آنگاه ضریب چولگی اول و دوم پیرسون تقریباً یکسان هستند و رابطه زیر به‌طور تقریبی صادق است:

$$SK_1 = SK_2 \quad \Rightarrow \quad \frac{\bar{x} - Mo}{S} = \frac{3(\bar{x} - Md)}{S} \quad \Rightarrow \quad \bar{x} - Mo = 3(\bar{x} - Md)$$

نکته: اگر ضریب چولگی صفر باشد منحنی توزیع متقارن است و اگر ضریب چولگی مثبت باشد منحنی توزیع چوله به راست و اگر ضریب چولگی منفی باشد منحنی توزیع چوله به چپ است.

نکته: اگر $|SK| \le 0/1$، منحنی توزیع تقریباً متقارن است.

اگر $0/1 \le |SK| \le 0/5$، چولگی توزیع خفیف است.

اگر $|SK| > 0/5$، چولگی توزیع شدید است.

مثال: طول عمر 200 لامپ دارای میانگین، میانه، مد و انحراف معیار 250، 222، 190 و 112 ساعت می‌باشد ضرایب چولگی پیرسون را محاسبه کنید.

$$SK_1 = \frac{\bar{x} - Mo}{S} = \frac{250 - 190}{112} = 0/54$$

$$SK_2 = \frac{3(\bar{x} - Md)}{S} = \frac{3(250 - 222)}{S} = 0/75$$

ضرایب چولگی مثبت است پس چولگی منحنی به راست می‌باشد.

مثال: ضریب پیرسون برای چولگی یک توزیع 0/32 است اگر انحراف معیار آن 6/5 و میانگین آن 29/6 باشد مد و میانه توزیع را محاسبه کنید. اگر مد توزیع بالا 24/4 باشد انحراف معیار آن چقدر است؟

چون $0/1 \le |SK| \le 0/5$ پس چولگی خفیف می‌باشد و:

الف) $SK_1 = \frac{\bar{x} - Mo}{S} \Rightarrow 0/32 = \frac{29/6 - Mo}{6/5} \Rightarrow Mo = 29/6 - (6/5 \times 0/32) = 27/52$

$$SK_2 = \frac{3(\bar{x} - Md)}{S} \Rightarrow 0/32 = \frac{3(29/6 - Md)}{6/5} \Rightarrow 3Md = 3 \times 29/6 - 6/5 \times 0/32$$

$$Md = 28/91$$

ب) $SK_1 = \frac{\bar{x} - Mo}{S} \Rightarrow 0/32 = \frac{29/6 - 24/4}{S} \Rightarrow S = 16/25$

تست: اگر $n = 10$، $\sum (x_i - \bar{x})^2 = 40$ و $\sum (x_i - \bar{x})^3 = 80$ باشد مقدار ضریب چولگی چقدر است؟

1) -3   2) 3   3) 2   4) 1

$$SK = \frac{m_3}{S^3} = \frac{\frac{1}{n}\sum (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2}\right)^3} = \frac{\frac{80}{10}}{\left(\sqrt{\frac{40}{10}}\right)^3} = \frac{8}{2^3} = 1$$

تست: در یک توزیع با چولگی خفیف میانگین حسابی $\bar{x} = 58$ و میانه $Md = 60$ به‌دست آمده است مد توزیع کدام است؟

1) 54   2) 64   3) 52   4) 62

$$(\bar{x} - Mo) = 3(\bar{x} - Md) \Rightarrow 58 - Mo = 3(58 - 60) \Rightarrow 58 - Mo = -6 \Rightarrow Mo = 64$$

تست: گشتاورهای مرتبه اول، دوم و سوم نسبت به مبدأ $a=10$ برای صفت متغیر x به‌صورت زیر محاسبه شده است: $M_1=2 \quad , \quad M_2=20 \quad , \quad M_3=70$ **ضریب چولگی توزیع کدام است؟**

1) 0/53   2) -0/53   3) 0/35   4) -0/35

$S^2 = M_2 - (M_1)^2 = 20-4=16 \quad \Rightarrow \quad S=4 \quad \Rightarrow \quad S^3=64$

$m_3 = M_3 - 3M_1M_2 + 2M_1^3 = 70-3\times2\times20+2\times8=-34$

$SK = \frac{-34}{64} = -0/53$

قضیه چبی‌شف: برای هر مجموعه از داده‌ها با شرط $K \geq 1$، حداقل $100 \times \left(1-\frac{1}{K^2}\right)$ درصد داده‌ها در فاصله $(\bar{x}-KS\ ,\ \bar{x}+KS)$ قرار دارند.

نکته: با استفاده از دستور چبی‌شف بطور تجربی می‌توان بیان کرد که برای داده‌هایی که منحنی فراوانی آن‌ها ناقوسی شکل است داریم:

الف) تقریباً 68 درصد داده‌ها به فاصله $(\bar{x}-S\ ,\ \bar{x}+S)$ تعلق دارند.

ب) تقریباً 95 درصد داده‌ها به فاصله $(\bar{x}-2S\ ,\ \bar{x}+2S)$ تعلق دارند.

ج) تمام و یا تقریباً 99/7 درصد داده‌ها به فاصله $(\bar{x}-3S\ ,\ \bar{x}+3S)$ تعلق دارند.

مثال: یک دستگاه اتوماتیکی بسته‌بندی زعفران به گونه‌ای تنظیم شده است که درون هر بسته به‌طور متوسط 16 گرم زعفران و با انحراف معیار 0/02 گرم زعفران قرار می‌دهد، دست‌کم چند درصد بسته‌های تولیدی‌اش دارای وزنی بین 15/95 گرم تا 16/05 گرم هستند؟

ملاحظه می‌شود که کران‌های فاصله $(15/95\ ,\ 16/05)$ عبارتند از میانگین به اضافه و منهای 0/05، بنابراین باید 0/05 را به‌عنوان **K** انحراف معیار تعبیر شود یعنی $SK=0/05$؛ پس، $K(0/02)=0/05$ و در نتیجه $K=2/5$؛ بنابراین، بنا بر قضیه چبی‌شف حداقل $\left(1-\frac{1}{(2/5)^2}\right)\times 100 = 84$ درصد بسته‌های زعفران وزنی بین 15/95 و 16/05 گرم خواهند داشت.

مثال: برای داده‌های زیر فاصله‌های $(\bar{x}-S\ ,\ \bar{x}+S)$ و $(\bar{x}-2S\ ,\ \bar{x}+2S)$ و $(\bar{x}-3S\ ,\ \bar{x}+3S)$ را مشخص کنید و درصد داده‌های متعلق به هر فاصله را با دستور تجربی به‌دست آورید.

1 , 1 , 1 , 2 , 3 , 3 , 4 , 6 , 6 , 7 , 8 , 12 , 14 , 18 , 18 , 20

$$\bar{x}=\frac{1+1+1+2+3+3+4+6+6+7+8+12+14+18+18+20}{16}=\frac{124}{16}=7/75$$

$$S^2=\frac{1}{n}\sum_{i=1}^{n} x_i^2-(\bar{x})^2$$

$$=\frac{1}{16}\left(3\times1^2+2^2+2\times3^2+4^2+2\times6^2+7^2+8^2+12^2+14^2+2\times18^2+20^2\right)-(7/75)^2$$

$$=40/81 \quad\Rightarrow\quad S=6/39$$

حدود 68 درصد داده‌ها به بازه روبه‌رو تعلق دارند: $(\bar{x}-S\ ,\ \bar{x}+S)\ ;\ (1/36\ ,\ 14/14)$

حدود 95 درصد داده‌ها به بازه روبه‌رو تعلق دارند: $(\bar{x}-2S\ ,\ \bar{x}+2S)\ ;\ (-5/03\ ,\ 20/53)$

حدود 99/7درصد داده‌ها به بازه روبه‌رو تعلق دارند: $(\bar{x}-3S\ ,\ \bar{x}+3S)=(-11/42\ ,\ 26/92)$

جالب است بدانید که حدود تجربی فوق تقریباً با حدود واقعی حاصل از محاسبه یکسان است چون توزیع فراوانی داده‌های فوق ناقوسی شکل است.

کشیدگی

کوتاهی یا بلندی توزیع فراوانی را نسبت به توزیع نرمال استاندارد کشیدگی می‌نامند. مقدار ماکزیمم منحنی نرمال استاندارد برابر $\frac{1}{\sqrt{2p}}$ است، اگر مقدار ماکزیمم منحنی از $\frac{1}{\sqrt{2p}}$بزرگتر باشد به آن منحنی، منحنی با کشیدگی مثبت می‌گویند و اگر مقدار ماکزیمم منحنی از $\frac{1}{\sqrt{2p}}$ کوچکتر باشد به آن منحنی، منحنی با کشیدگی منفی می‌گویند.

ضریب کشیدگی

ضریب کشیدگی شاخصی است که پراکندگی جامعه را نسبت به توزیع نرمال استاندارد نشان می‌دهد. ضریب کشیدگی را با نماد **E** نمایش می‌دهیم و عبارت است از:

$E=\frac{m_4}{S^4}-3$ $m_4:$ گشتاور مرکزی مرتبه چهارم $S:$ انحراف معیار

نکته: اگر ضریب کشیدگی صفر باشد کشیدگی منحنی استاندارد است اگر ضریب کشیدگی مثبت باشد کشیدگی منحنی مثبت و اگر ضریب کشیدگی منفی باشد کشیدگی منحنی منفی است.

نکته: اگر $|E|\le 0/1$، توزیع تقریباً نرمال است.

اگر $0/1 < |E| \le 0/5$، تفاوت کشیدگی توزیع با کشیدگی توزیع نرمال استاندارد اندک است.

اگر $|E| > 0/5$، تفاوت کشیدگی توزیع با کشیدگی توزیع نرمال استاندارد بسیار است.

مثال: برای جدول توزیع فراوانی زیر ضریب کشیدگی را محاسبه کنید.

حدود واقعی	20_24	24_28	28_32	32_36	36_40
فراوانی	2	8	16	14	10

ابتدا مرکز دسته‌ها را به‌دست آورده و سپس در فرمول زیر جایگذاری می‌کنیم:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{(2\times22) + (8\times26) + (16\times30) + (14\times34) + (10\times38)}{2 + 8 + 16 + 14 + 10} = 31/76$$

$$s^2 = \frac{1}{N}\sum f_i x_i^2 - (\bar{x})^2 = \frac{1}{50}\left(2\times22^2 + 8\times26^2 + 16\times30^2 + 14\times34^2 + 10\times38^2\right) - (31/76)^2$$

$$s^2 = 19/3 \quad \Rightarrow \quad s^4 = 372/58$$

$$m_4 = \frac{1}{N}\sum f_i (x_i - \bar{x})^4 = \frac{1}{50}\Big(2(22-31/76)^4 + 8(26-31/76)^4 + 16\,(30-31/76)^4 + 14(34-31/76)^4 + 10(38-31/76)^4\Big)$$

$$= \frac{1}{50}(18148/02 + 8806/02 + 153/52 + 352/47 + 15161/3) = 852/43$$

$$E = \frac{m_4}{s^4} - 3 = \frac{852/43}{372/58} - 3 = -0/71$$

تست: در جامعه‌ای به حجم $n = 25$ رابطه‌های $\sum_{i=1}^{25}(x_i - \bar{x})^2 = 50$ و $\sum_{i=1}^{25}(x_i - \bar{x})^4 = 275$ برقرار هستند ضریب کشیدگی و نوع کشیدگی کدام است؟

1) $E = 0/25$، اختلاف منحنی با منحنی نرمال کم است.

2) $E = -0/25$، اختلاف منحنی با منحنی نرمال کم است.

3) $E = 0/25$، اختلاف منحنی با منحنی نرمال زیاد است.

4) $E = -0/25$، اختلاف منحنی با منحنی نرمال زیاد است.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{25}\sum_{i=1}^{25}\left(x_i - \overline{x}\right)^2 = \frac{50}{25} = 2 \quad \Rightarrow \quad S^4 = 4$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4 = \frac{1}{25}\sum_{i=1}^{25}\left(x_i - \overline{x}\right)^4 = \frac{275}{25} = 11$$

$$E = \frac{m_4}{S^4} - 3 = \frac{11}{4} - 3 = -0/25$$

چون $0/1 < |E| \le 0/5$ است بنابراین اختلاف منحنی با منحنی نرمال کم است. و گزینه دوم صحیح می‌باشد.

تست: گشتاورهای مرکزی مرتبه دوم، سوم و چهارم برای توزیع صفت متغیر به قرار زیر می‌باشد:

$m_2 = 2 \quad , \quad m_3 = 0 \quad , \quad m_4 = 12$

ضریب کشیدگی کدام است؟

1) 10   2) 6   3) 0   4) 4

با توجه به این‌که $m_2 = S^2$ است داریم:

$$E = \frac{m_4}{S^4} - 3 = \frac{12}{(2)^2} - 3 = 3 - 3 = 0$$

تست: اگر چهار گشتاور اول یک متغیر حول مبدأ به ترتیب برابر ۶، ۴۵/۲، ۳۷۸ و ۳۳۱۸/۸ و واریانس برابر ۵ باشد ضریب کشیدگی کدام است؟

1) صفر   2) 0/976   3) 24/4   4) 4/88

به‌طور کلی برای محاسبه m_r می‌توان از فرمول زیر استفاده کرد:

$$m_r = \sum_{i=0}^{r}(-1)^i\binom{r}{i}m'_{r-i}\cdot\left(m'_1\right)^i$$

$$m_4 = \sum_{i=0}^{4}(-1)^i\binom{4}{i}m'_{4-i}\cdot\left(m'_1\right)^i$$

$$m_4 = m'_4 - 4m'_3\,m'_1 + 6m'_2\left(m'_1\right)^2 - 4m'_1\left(m'_1\right)^3 + \left(m'_1\right)^4$$

$$m_4 = 3318/8 - 4(378)(6) + 6(45/2)(6)^2 - 3(6)^4 = 122$$

$$S^2 = 5 \quad \Rightarrow \quad S^4 = 25$$

$$E = \frac{m_4}{S^4} = \frac{122}{25} = 4/88$$

تستهای فصل اول

۱- میزان سود شرکت سهامی بتا در پنج سال گذشته بر حسب درصد فروش به ترتیب ۳، ۴، ۴، ۳، ۲ می‌باشد. کدام یک از کمیت‌های زیر به عنوان شاخص مرکزی وضع سودآوری شرکت را بهتر نشان می‌دهد؟

1) 2/5 2) 3/2 3) 4 4) 3/1

۲- واریانس کدام یک از چهار مجموعه‌ی زیر بیش‌تر است؟

1) 1,1,2,4,5,7,8,8 2) 1,1,1,4,5,8,8,8 3) 1,1,1,1,8,8,8,8 4) 1,2,3,4,5,6,7,8

۳- با تغییر مدیریت در یک فروشگاه بزرگ، فروش در سال اول دو برابر سال قبل، در سال دوم سه برابر سال اول و در سال سوم چهار برابر سال دوم شده است. به طور متوسط، فروش از آغاز مدیریت چند برابر شده است؟

1) بیش از سه برابر 2) سه برابر 3) قدری کم‌تر از سه برابر 4) 2/5 برابر

۴- حقوق پرداختی به کارمندان شرکت آلفا به طور متوسط ۱۵ هزار تومان با انحراف معیار ۳ هزار تومان است. اگر ۲۰٪ به میانگین حقوق کارمندان اضافه شود، به ترتیب میانگین و انحراف‌معیار حقوق پرداختی چقدر خواهد شد؟

1) 15/3 و 3 2) 15/3 و 3/6 3) 18 و 3 4) 18 و 3/6

۵- جدول مقابل توزیع فراوانی فروش یک شرکت را نشان می‌دهد. میانگین و انحراف معیار فروش به ترتیب چقدر است؟

فروش به هزار تومان	تعداد روزها
20 تا کم‌تر از 30	10
30 تا کمتر از 40	25
40 تا کمتر از 50	15

1) 35 و 8/6 2) 36 و 5/7 3) 36 و 7 4) 35 و 9

۶- اگر کمیت‌های $X_1, X_2, ..., X_n$ با حجم n به دست آمده باشند، کدام یک از روابط زیر صادق است؟

1) $\sum X_i = n\bar{X}^2$ 2) $\sum(X_i - \bar{X}) = \mathbf{o}$ 3) $\sum(X_i - \bar{X})^2 = \mathbf{o}$ 4) $\sum(X_i - me) = \mathbf{o}$

۷- اگر در طی دوره‌ی دو ساله نرخ تورم به ترتیب ۲۱٪ و ۴۴٪ باشد، میانگین نرخ تورم در طی دوره‌ی مزبـور برابر است با:

1) ٪33 2) ٪32 3) ٪30 4) ٪32/5

۸- اگر به مقادیر متغیر X (با انحراف معیار ۱) عددی مانند a اضافه کنیم و آن را در عـددی ماننـد $\frac{c}{b}$ ضـرب کنیم، انحراف معیار آن برابر خواهد شد با:

1) $a + \frac{c}{b}$ 2) $a + \frac{c^2}{b^2}$ 3) $\frac{c}{b}$ 4) $a^2 + \frac{c}{b}$

۹- از روی توزیع صفت متغیر X در جامعه کمیت‌های زیر محاسبه شده است:

$$n = 100, \sum n_i (X_i - \bar{X})^2 = 36, \sum n_i X_i = 1200$$

ضریب تغییرات صفت متغیر X کدام است؟

1) ٪5 2) ٪4 3) ٪2 4) ٪6

۱۰- در جامعه‌ای به حجم ۲۰ = n پس از محاسبات لازم کمیت‌های زیر به دست آمده است:

$$\sum(X_i - \bar{X})^2 = 1620, \sum(X_i - \bar{X})^3 = -328$$

ضریب چولگی توزیع کدام است؟

1) ٪449 2) -٪449 3) 0/022 4) -0/02

۱۱- کدام یک از عبارت‌های زیر می‌تواند یکی از خواص مهم میانه را در توزیع جامعه بیان کند؟

1) $\sum(X_i - me)^2$ 2) $\sum(X_i - me)$ 3) $\sum|X_i - me|$ 4) $\sum(\bar{X} - me)$

۱۲- نظر گروهی از سوادآموزان راجع به زمان پخش برنامه‌ی نهضت سوادآموزی از سیمای جمهوری اسـلامی جمع‌آوری شده است کدام شاخص مرکزی برای آن داده‌ها مناسب‌تر است؟

1) میانگین 2) میانه 3) نما 4) چارک اول

۱۳- در صورتی که جامعه‌ای دارای چولگی مثبت باشد:

1) میانه در وسط، میانگین سمت راست و نما سمت چپ آن قرار دارد.

2) میانگین در وسط، میانه سمت راست و نما سمت چپ آن قرار دارد.

3) میانه در وسط، نما سمت چپ و میانگین سمت چپ آن قرار دارد.

4) میانگین در وسط، نما سمت چپ و میانه سمت راست آن قرار دارد.

۱۴- با توجه به جدول زیر، سرمایه‌گذاری در کدام شرکت مناسب‌تر است؟

شاخص	الف	ب	ج
میانگین سود	7	5	7
انحراف معیار سود	2	1	2
ضریب چولگی پیرسون	4	o	-4

1) الف 2) ب 3) ج 4) تفاوتی ندارد

۱۵- ضریب تغییرات (coefficient of Variation) عدد ۵ برابر است با:

1) 5 2) 1 3) ∞ 4) **o**

۱۶- کدام یک از روابط زیر بین میانگین حسابی ($\bar{x}$) میانگین هندسی ($\bar{x}_G$) و میانگین هارمونیک ($\bar{x}_H$) برقرار است؟

1) $\bar{X}_H < \bar{X}_G < \bar{X}$ 2) $\bar{X}_G < \bar{X} < \bar{X}_H$ 3) $\bar{X}_G < \bar{X}_H < \bar{X}$ 4) $\bar{X} < \bar{X}_G < \bar{X}_H$

۱۷- در یک توزیع با چولگی خفیف میانگین حسابی ۵۲/۴ و میانه ۵۱/۸ = med =md به دست آمده است، مد یا نمای توزیع کدام است؟

1) 53/6 2) 50/6 3) 54/2 4) 51/6

۱۸- سه ماشین که به تولید یک کالا مشغول‌اند، اولی یک کالا را در ۲، دومی در ۳ و سومی در ۶ دقیقه تولید می‌کنند. اگر این سه ماشین با هم کار کنند به طور متوسط یک کالا در چند دقیقه تولید می‌شود؟

1) 3/67 2) 3/3 3) 3 4) 3/1

۱۹- در صورتی که انحراف معیار ۱۲ عدد مساوی ۲/۴ باشد و به هر یک از اعداداین توزیع، عدد ۴ را اضافه کنیم، انحراف معیار جدید چقدر خواهد شد؟

1) $2/4\sqrt{4}$ 2) 9/6 3) 6/4 4) 2/4

۲۰- برای رسم هیستوگرام (نمودار مستطیلی) محورهای x , y بر اساس کدام اندازه‌ها مدرج می‌شوند؟

1) کرانه‌های طبقات و فراوانی طبقات
2) حدود طبقات و چگالی
3) حد وسط طبقات و فراوانی مطلق
4) مقادیر متغیر **x** و فراوانی‌های تجمعی

۲۱- قیمت سهام یک کارخانه از ۱۰۰ ریال در سال ۱۳۷۰ به ۱۶۰۰ ریال در سال ۱۳۷۴ رسیده است. متوسط نرخ افزایش قیمت سهام در این دوره چقدر است؟

1) ٪100 2) ٪200 3) ٪10 4) ٪20

۲۲- میانگین یک جامعه‌ی آماری یک بار با استفاده از داده‌های خام μ_1 و بار دیگر با استفاده از جدول توزیع فراوانی μ_2 محاسبه شده است. کدام یک از موارد زیر معمولاً صادق است؟

1) $\mu_1 = \mu_2$ 2) $\mu_1 > \mu_2$ 3) $\mu_1 \neq \mu_2$ 4) $\mu_1 < \mu_2$

۲۳- برای مقایسه دو توزیع فراوانی مربوط به حقوق پرداختی به کارگران مرد و زن در یک کارخانه، کدام یک از نمودارهای زیر مناسب‌تر است؟

1) پلی‌گون (چند گوش) فراوانی نسبی
2) نمودار میله‌ای فراوانی مطلق
3) نمودار تجمعی (**Ogive**) فراوانی مطلق
4) هیستوگرام (بافت‌نگار) فراوانی نسبی

۲۴- برای محاسبه متوسط نرخ رشد تولید ناخالص ملی در پنج سال گذشته، از کدام شاخص استفاده می‌شود؟

1) میانگین حسابی ساده
2) میانگین حسابی موزون
3) میانگین هارمونیک (هم‌سازه)
4) میانگین هندسی

۲۵- چارک سوم حقوق در یک سازمان ۶۵ هزار تومان است. یعنی سه چهارم کارکنان .... هزار تومان حقوق می‌گیرند؟

1) تا 65
2) 65
3) بیش‌تر از 65
4) 65 هزار تومان و بقیه کارکنان کم‌تر از 65

۲۶- توزیع صفت متغیر X در جامعه‌ای به صورت جدول زیر به دست آمده است.

فاصله طبقاتی $x_i - x_{i+1}$	2-4	4-6	6-8	8-10
فراوانی n_i	1	1	5	3

اگر انحراف معیار توزیع فوق $\sigma = 1/8$ باشد، ضریب چولگی پیرسن و تفسیر آن کدام است؟

1) $A_s = -0/87$، توزیع دارای چولگی شدید و چپ است.
2) $A_s = -0/27$، توزیع دارای چولگی غیر قابل اغماض و راست است.
3) $A_s = -0/33$، توزیع دارای چولگی غیر قابل اغماض و چپ است.
4) $A_s = -0/87$، توزیع دارای چولگی شدید و راست است.

۲۷- قیمت کالایی در سال گذشته ۲۰٪ کاهش و امسال ۲۰٪ افزایش داشته است. متوسط نرخ رشد قیمت این کالا در این دو سال چیست؟

1) صفر　　2) $\sqrt{0/96}$　　3) $(0/96)^{1/2}-1$　　4) $\frac{1}{2}(\log 0/8-\log 0/02)$

۲۸- در جامعه‌ای پس از محاسبات لازم کمیت‌های زیر به دست آمده است. $n=10, \sum x_i=20, \sum x_i^2=200$

اگر مد (نما) توزیع برابر ۳ به دست آمده باشد، ضریب چولگی پیرسن و تفسیر آن کدام است؟

1) $A_s=-0/5$ توزیع دارای چولگی شدید و چپ است.

2) $A_s=-0/25$، توزیع دارای چولگی غیر قابل اغماض و راست است.

3) $A_s=-0/25$، توزیع دارای چولگی غیر قابل اغماض و چپ است.

4) $A_s=-0/5$، توزیع دارای چولگی شدید و راست است.

۲۹- در جامعه‌ای به حجم $N=50$ برای صفت متغیر X کمیت‌های زیر به دست آمده است:

$$\sum x_i=250, \sum x_i^2=2500$$

ضریب تغییرات صفت X کدام است؟

1) $CV=\%25$　　2) $CV=\%50$　　3) $CV=\%75$　　4) $CV=\%100$

۳۰- میانگین و انحراف معیار حقوق در یک سازمان به ترتیب ۵۰ هزار تومان و ۲۰ هزار تومان است اگر حقوق‌ها در این سازمان ۲۵ درصد افزایش یابند، ضریب تغییرات حقوق چه خواهد شد؟

1) نصف خواهد شد.　　2) تغییر نخواهد کرد

3) چهار برابر خواهد شد　　4) 25 درصد افزایش خواهد یافت

۳۱- فروش بنگاهی در سال‌های ۱۳۷۳ تا ۱۳۷۷ به ترتیب ۵، ۱۰، ۲۰، ۴۰ و ۸۰ میلیون تومان بوده است. متوسط نرخ رشد سالیانه فروش بنگاه .... و توزیع آن دارای چولگی .... است.

1) 100%، مثبت　　2) 100% منفی　　3) 200%، مثبت　　4) 200%، منفی

۳۲- واریانس نمونه متشکل از سه عدد ۵۶۷۹۲۱۱۲۰ و ۵۶۷۹۲۱۱۲۴ و ۵۶۷۹۲۱۱۲۲ کدام است؟

1) $\frac{8}{3}$　　2) 4　　3) $\frac{25124}{3}$　　4) 25112

۳۳- کدام گزینه، یکی از خواص مهم مشخصه میانگین می‌باشد؟

1) برای هر توزیع، میانگین حسابی از نما کوچک‌تر است.

2) در هر توزیع، حاصل جمع قدر مطلق انحرافات مقادیر متغیر از میانگین به حداقل می‌رسد.

3) در هر توزیع، مجموع مجذورات تفاضل‌های مقادیر از میانگین به حداقل می‌رسد.

4) در هر توزیع، مجموع توان سوم انحرافات مقادیر متغیر از میانگین حسابی به حداقل می‌رسد.

۳۴- بنا به نابرابری چبی‌شف (Chebyshev) احتمال این که قدر مطلق یک متغیر استاندارد شده، حداقل k باشد:

1) برابر $\frac{1}{k^2}$ است.

2) برابر k^2 است به شرط آن که توزیع نرمال باشد.

3) حداکثر $\frac{1}{k^2}$ است.

4) حداکثر $\frac{1}{k^2}$ به شرط آن که توزیع نرمال باشد.

۳۵- در صورتی که به بزرگترین عدد یک سری داده مقدار ثابتی اضافه گردد، این افزایش بر کدام معیار تأثیر نمی‌گذارد؟

1) ضریب پراکندگی
2) میانه
3) میانگین
4) واریانس

۳۶- اگر واریانس ۱۰ مشاهده برابر ۵۰ باشد و مشاهدات را در ۴ ضرب و با ۷ جمع کنیم، واریانس جدید برابر است با:

1) 200
2) 228
3) 800
4) 912

۳۷- جدول توزیع فراوانی زیر را در نظر بگیرید. اگر $\mu = 2$ و $N = 28$ باشد، مقادیر a و b عبارتند از:

x_i	0	1	2	3	4
F_i	3	a	10	b	3

1) $a = b = 6$
2) $a = 5$ و $b = 7$
3) $a = b = 7$
4) $a = 4$ و $b = 8$

۳۸- اتومبیلی ۶۰ کیلومتر اول از مسافتی را با سرعت ۸۰ کیلومتر در ساعت و ۶۰ کیلومتر دوم را با سرعت ۱۰۰ کیلومتر در ساعت طی کرده است. برای محاسبه سرعت متوسط اتومبیل، کدام یک از میانگین‌های زیر مناسب است؟

1) حسابی
2) موزون
3) همساز (هارمونیک)
4) هندسی

۳۹- میانگین نمرات آمار و واریانس آن‌ها در دو کلاس به صورت زیر است:

کلاس	1	2
N_i تعداد دانشجو	20	3
μ_i میانگین نمرات	15	10
σ_i^2 واریانس نمرات	17	12

میانگین و واریانس نمرات کل دانشجویان دو کلاس چقدر است؟

1) 12/5 و 20 2) 12 و 20 3) 12/5 و 35 4) 12 و 35

۴۰- در توزیعی با چولگی منفی انتظار می‌رود که .... کم‌ترین مقدار را داشته باشد.

1) دامنه تغییرات 2) میانه 3) میانگین 4) نما

۴۱- برای تعیین آن که در ۳۰ روز گذشته، به نسبت، قیمت دلار از ثبات بیش‌تری برخوردار بوده است یا یورو، استفاده از کدام شاخص آماری مناسب‌تر است؟

1) انحراف متوسط 2) ضرایب پراکندگی 3) ضریب چولگی 4) واریانس

۴۲- فرض کنید شاخص قیمت خرده‌فروشی از ۲۰۰ در سال ۱۳۷۸ به ۴۵۰ در سال ۱۳۸۰ رسیده باشد. متوسط نرخ تورم سالانه در این فاصله زمانی چقدر بوده است؟

1) 50% 2) 62/5% 3) 125% 4) 150%

۴۳- در رسم هیستوگرام (بافت نگار) محور xها را بر اساس کدام اندازه مندرج می‌کنند؟

1) فراوانی های نسبی 2) کرانه‌های طبقات 3) حد وسط طبقات 4) فراوانی‌های تجمعی

۴۴- طی بررسی‌هایی که از پرسش‌نامه‌ی طرح هزینه و درآمد خانوارهای شهری برای چهار سال متوالی به عمل آمده، معلوم شد، که هزینه‌ی سوخت یک خانوار برای خرید نفت سفید به ترتیب برابر با ۱/۶، ۱/۸، ۲/۱، ۲/۵ ریال به ازای هر لیتر شده است. اگر خانواری ۲۰۰۰۰ ریال برای هر سال هزینه کرده باشد، متوسط هزینه‌ی سوخت خانوار به ازای یک لیتر در ۴سال برابر است با:

1) 1/75 2) 2/25 3) 2/5 4) 1/94

۴۵- با فرض در اختیار داشتن عبارت $\sum |x_i - a|$ اگر a، میانه توزیع باشد، این عبارت:

1) حداکثر است. 2) مقدار بین حداقل و حداکثر دارد.

3) حداقل است. 4) هیچ‌کدام از موارد فوق

۴۶- واریانس عدد ۶، ۵، ۴، ۳، ۲ در جامعه‌ای برابر است با:

1) 2   2) $\sqrt{3}$ 3) 5/5   4) 3

۴۷- اگر ضریب چولگی توزیع یک جامعه ۰/۶۶- باشد، آن‌گاه جامعه مورد مطالعه:

1) نرمال است.
2) با جامعه‌ی نرمال تفاوت مختصری دارد.
3) با جامعه‌ی نرمال تفاوت فاحش دارد.
4) با اطلاعات داده شده نمی‌توان قضاوت کرد.

۴۸- اگر $V(x)=9$ باشد، در این صورت انحراف معیار ۴x-۳ برابر است با:

1) 6   2) 36   3) 12   4) 33

۴۹- در یک توزیع با چولگی خفیف، میانگین حسابی $\bar{X}=52/4$ و میانه $Me=51/8$ به دست آمده است. مد توزیع کدام است؟

1) 53/6   2) 50/6   3) 54/2   4) 51/6

۵۰- کدام یک از روابط زیر بین میانگین حسابی $(\bar{x})$، میانگین هندسی $(\bar{x}_G)$ و میانگین هم‌ساز $(\bar{x}_H)$ در توزیع‌های صفت متغیر برقرار است؟

1) $\bar{X}_H < \bar{X}_G < \bar{X}$ 2) $\bar{X}_G < \bar{X} < \bar{X}_H$ 3) $\bar{X}_G > \bar{X}_H > \bar{X}$ 4) $\bar{X}_H > \bar{X}_G > \bar{X}$

۵۱- در جامعه‌ای به حجم $n=10$، کمیت‌های زیر محاسبه شده است:

$\sum n_i x_i = 40, \sum n_i x_i^2 = 400, \sum n_i x_i^3 = 600$ گشتاوری مرکزی مرتبه سوم (μ_3) کدام است؟

1) -292   2) 292   3) -548   4) 485

۵۲- کدام تعریف برای صفت مشخصه صحیح است؟

1) صفتی است که اندازه‌گیری آن از فردی به فرد دیگر تغییر می‌کند.
2) صفتی است که قابل اندازه‌گیری است.
3) صفت مشترک برای افراد جامعه است.
4) صفتی که قابل شمارش باشد.

۵۳- اگر μ_x میانگین $x_1, x_2, \dots, x_n$ باشد، آنگاه $\sum (x_i - \mu_x)$ مساوی خواهد شد با:

1) μ_x 2) **N**   3) 1   4) صفر

۵۴- برای یک جامعه $\sum x_i = 100$ و $\sum x_i^2 = 580$ می‌باشد، اگر $N = 20$ باشد، ضریب پراکندگی نسبی (ضریب تغییرات) کدام است؟

1) %80
2) %40
3) %70
4) %50

۵۵- میانگین و میانه یک جامعه به ترتیب ۳۰ و ۵۰ است. فرض کنید توزیع جامعه از چولگی معقولی (ضعیف) برخوردار باشد، مد توزیع را محاسبه کنید.

1) 90
2) 40
3) مد ندارد
4) 25

۵۶- اگر ضریب کشیدگی توزیعی ۰/۷۱- به دست آمده باشد، می توان گفت که:

1) کشیدگی جامعه نسبت به توزیع نرمال بیش‌تر و تفاوت آن فاحش است.
2) کشیدگی جامعه نسبت به توزیع نرمال کم‌تر و تفاوت آن فاحش است.
3) کشیدگی جامعه نسبت به توزیع نرمال بیش‌تر و تفاوت آن ناچیز است.
4) کشیدگی جامعه نسبت به توزیع نرمال کم‌تر و تفاوت آن ناچیز است.

۵۷- چنانچه یک جامعه غیر نرمال باشد، فاصله‌ی $(\mu_x \pm 2\sigma_x)$ حداقل شامل چند درصد داده‌ها است؟

1) %95/45
2) %85
3) %95
4) %75

۵۸- میانه‌ی دادهای جدول زیر کدام است؟

CL (فاصله طبقات)	20-29	30-39	40-49
F_i (فراوانی)	3	6	7

1) 34/6
2) 34/5
3) 37/8
4) 37/3

۵۹- ضریب تغییرات برای مقایسه‌ی پراکندگی دو جامعه‌ای به کار می‌رود که:

1) دارای میانگین‌های متفاوت باشند.
2) دارای واحدهای اندازه‌گیری مختلف باشند.
3) دارای مقادیر منفی باشند.
4) یا واحدهای اندازه‌گیری‌شان مختلف باشد و یا میانگین‌های متفاوت داشته باشند.

۶۰- اگر $n = 50, \sum x_i^2 = 3250$ و $\mu_x = 7$ و $\sum (x_i - \mu_x)^3 = 96$ باشد، ضریب چولگی جامعه کدام است؟

1) %3
2) %6
3) %1/96
4) %2/3

۶۱- دستگاه A در اندازه‌گیری مکرر از شیء واحدی دارای واریانس $\sigma^2 = 9$ بوده است. دستگاه B در اندازه‌گیری مکرر از همان شیء دارای واریانس $\sigma^2 = 25$ بوده است.

1) دستگاه A دقیق‌تر است.

2) دستگاه B دقیق‌تر است.

3) دستگاه A اندازه‌گیری‌های بزرگ‌تری از دستگاه B به دست می‌دهد.

4) دستگاه B اندازه‌گیری‌های بزرگ‌تری از دستگاه A به دست می‌دهد.

۶۲- نمای اعداد ۹، ۱۴، ۱۰، ۹، ۵، ۳، ۲ کدام عدد می‌باشد؟

1) 5　　2) 7　　3) 9　　4) 14

۶۳- اگر میانگین x_1 تا x_N برابر μ_x و میانگین y_1 تا y_k برابر μ_y باشد و داشته باشیم $\mu_y = a\mu_x$ در آن صورت مقدار $\dfrac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{K} y_i}$ کدام است؟

1) $\dfrac{N}{ka}$　　2) $k\mu_y$　　3) $N\mu_x$　　4) $N\mu_y$

۶۴- داده‌های زیر در دست است: ۹۸۷۵۰، ۹۸۹۵۰، ۹۸۷۵۰ انحراف معیار این داده‌ها برابر است با:

1) 98750　　2) 98750　　3) 1　　4) صفر

۶۵- متوسط درآمد ماهانه کارگران کارخانه A، ۱۷ هزار تومان با واریانس ۴ می‌باشد. در کارخانه B متوسط درآمد ماهانه ۲۵۰ هزار ریال با واریانس ۹۰۰ می‌باشد.

1) اختلاف درآمد در کارخانه A بیش از کارخانه B است.

2) اختلاف درآمد در کارخانه B بیش از کارخانه A است.

3) درآمدهای اکثر افراد کارخانه A کم‌تر از اکثر افراد کارخانه B است.

4) کم‌ترین درآمد در کارخانه A بیش از کارخانه B است.

۶۶- کدام یک از معیارهای زیر، برای تعیین نوع کالایی که در بازار بیش‌ترین متقاضی را داشته باشد، مناسب‌تر است؟

1) میانه　　2) نما　　3) میانگین حسابی　　4) میانگین وزنی

۶۷- سرمایه شرکتی در سال ۱۳۶۴، ۲ میلیون تومان، در سال ۱۳۶۵، ۴ میلیون تومان و در سال ۱۳۶۶، ۳۲ میلیون تومان بوده است. به طور متوسط این شرکت هر سال نسبت به سال قبل چند برابر سود داشته است؟

1) $3/84$　　2) 4　　3) $\frac{38}{3}$　　4) $(256)^{\frac{1}{2}}$

۶۸- اگر واریانس متغیر تصادفی X برابر ۴ باشد، واریانس ۲x-۳ برابر است:

1) $\frac{1}{4}$　　2) 2　　3) 4　　4) 16

۶۹- کدام نمودار برای نمایش داده‌ها با مقیاس اسمی به کار می‌رود؟

1) نمودار مستطیلی　　2) نمودار چند ضلعی یا چند گوش

3) نمودار دایره‌ای　　4) نمودار تراکمی

۷۰- کشیدگی گشتاور یک توزیع آماری، ۳ است. کدام گزینه صحیح است؟

1) پراکندگی آن در حد توزیع نرمال است.　　2) پراکندگی آن زیاد است.

3) پراکندگی آن پایین است.　　4) جهت اظهار نظر به اطلاعات بیش‌تری نیاز است.

۷۱- کدام یک از نمودارهای زیر برای توصیف داده‌های فاصله‌ای مناسب‌تر است؟

1) مستطیلی　　2) پاره‌تو　　3) دایره‌ای　　4) ستونی

۷۲- داده‌های طبقه‌بندی شده زیر را در نظر بگیرید، مقدار مد را محاسبه کنید.

فاصله طبقاتی CL	-10-0	0-10	10-20
F_i (فراوانی)	30	20	20

1) -10　　2) $2/5$　　3) $-2/5$　　4) 0

۷۳- اگر میانگین و واریانس X به ترتیب ۳ و ۹ باشد، میانگین و انحراف معیار $y=\frac{1}{2}x+1$ کدام است؟

1) $\mu=1/5,\sigma=2/5$　　2) $\mu=1/5,\sigma=1/5$　　3) $\mu=2/5,\sigma=4/5$　　4) $\mu=2/5,\sigma=1/5$

۷۴- جامعه‌ای دارای چولگی ضعیف است (توزیع جامعه از چولگی معقولی برخوردار است). میانگین و میانه آن را به ترتیب ۱۰۰ و ۱۲۰ می‌باشد. مقدار مد چقدر است؟

1) 110　　2) 120　　3) 160　　4) 220

۷۵- جدول طبقه‌بندی زیر داده شده است :

CL فاصله طبقاتی	10-20	20-30	30-40
n_i (فراوانی)	10	30	20

مقدار چارک سوم کدام است؟

1) 25　　2) 30/5　　3) 32/5　　4) 45

۷۶- ضریب کشیدگی یک جامعه ۰/۶- است، جامعه‌ی مورد مطالعه:

1) دارای چولگی فاحش است.　　2) در مقایسه با جامعه نرمال تفاوتی ندارد.

3) در مقایسه با جامعه نرمال تفاوت اندکی دارد.　　4) در مقایسه با جامعه نرمال تفاوت فاحشی دارد.

۷۷- اگر واریانس ۵ عدد ۱۴ و $\sum x_i^2 = 390$ باشد، میانگین اعداد برابر است با:

1) 5/12　　2) 3/2　　3) 2/56　　4) 8

۷۸- اگر $N=100$ و $\sum x_i = 200$ و $\sum x_i^2 = 500$ باشد، مقدار ضریب تغییرت کدم است؟

1) %50　　2) %90　　3) %10　　4) %20

۷۹- چارک اول جدول زیر کدام است؟

حدود طبقات	2-5	6-9	10-13
فراوانی مطلق	10	30	20

1) 5　　2) 6/17　　3) 9/5　　4) 10

۸۰- اگر $N=1000$ و $\sum F_i(x_i-\mu_x)^4 = 5000$ و انحراف معیار جامعه ۲ باشد، مقدار ضریب کشیدگی کدام است؟

1) -2/69　　2) -2/53　　3) 0/31　　4) 2/53

۸۱- اگر ۳ اتومبیل مسیر ۶۰ کیلومتری بین دو منطقه را به ترتیب با سرعت ۱۲۰، ۶۰ و ۹۰ کیلومتر در ساعت طی نمایند، میانگین سرعت این سه اتومبیل برابر با چند کیلومتر در ساعت است؟

1) تقریباً 83　　2) تقریباً 86　　3) تقریباً 90　　4) 90

۸۲- میانگین سن یک گروه ۱۲ سال و ضریب تغییرات سن آنان ۲۰ درصد است. انحراف معیار سن آنان چقدر است؟

1) 0/6　2) 2/4　3) 60　4) 240

۸۳- اگر $N=10$, $\sum(x_i-\mu_x)^2=40$ و $\sum(x_i-\mu_x)^3=80$ باشد مقدار ضریب چولگی چقدر است؟

1) -3　2) 3　3) 2　4) 1

۸۴- در جدول زیر دهک دوم برابر چند است؟

CL	40-50	50-60	60-70
F_i	5	18	7

1) 48/2　2) 51/5　3) 50/55　4) 62/38

۸۵- با فرض در اختیار داشتن $\sum_{i=1}^{N}|y_i-a|$ به شرط آن که **a** میانه باشد، همواره حاصل این عبارت .... است.

1) حداقل　2) حداکثر

3) مقداری بین حداقل و حداکثر　4) نامشخص

۸۶- کشیدگی (**Kurtosis**) چندگی و گشتاوری توزیع نرمال به ترتیب (از چپ به راست) کدام است؟

1) (o,o)　2) (0/263و3)　3) (0/263 و 0/263)　4) (3و3)

۸۷- کدام نمودار از نوع «تحلیل اکتشافی داده‌ها» (**Exploratorg Data Analysis**) می‌باشد؟

1) دایره‌ای　2) هیستوگرام　3) پاره تو　4) شاخه و برگ

۸۸- کدام نمودار برای نمایش مشاهدات کمی طبقه‌بندی نشده به کار می‌رود؟

1) پاره‌تو　2) چند ضلعی　3) ریشه و برگ　4) بافت‌نگار

۸۹- در حالت چولگی مثبت $(SK>o)$، کدام رابطه بین میانگین (μ_x)، مد (**Mo**) و میانه (**Md**) برقرار است؟

1) $MO<Md<\mu_x$　2) $Mo\le\mu_x<Md$　3) $\mu_x<Md<Mo$　4) $\mu_x\le Md\le Mo$

۹۰- میانگین و انحراف معیار حقوق کارکنان در یک بنگاه به ترتیب ۸۰ هزار تومان و ۲۰ هزار تومان است. اگر حقوق‌ها در این بنگاه ۱۲/۵ درصد افزایش یابد، ضریب تغییرات جدید چقدر خواهد شد؟

1) 12/5 درصد　2) 20 درصد　3) 25 درصد　4) 40 درصد

۹۱- خاصیت مهم میانه آن است که:

1) مجموع انحرافات از میانه صفر است.
2) تعداد انحرافات از میانه حداقل است.
3) مجذور انحرافات از میانه حداقل است.
4) مجموع قدرمطلق انحرافات از میانه حداقل است.

۹۲- اتومبیلی مسیری را با سرعت ۱۰۰ کیلومتر در ساعت رفته و $\frac{1}{3}$ مسیر را با سرعت ۸۰ کیلومتر و باقیمانده را با سرعت ۱۲۰ کیلومتر برگشته است. متوسط سرعت این اتومبیل چقدر بوده است؟

1) 90
2) 100
3) 101/4
4) 102/8

۹۳- در توزیع زیر مد کدام است؟

L-C	3-5	6-8	9-11	جمع
F_i	4	20	12	36

1) 6/33
2) 6/53
3) 7/5
4) 20

۹۴- کدام دسته از فنون آماری زیر بر فرض آزاد از توزیع بنا شده‌اند؟

1) پارامتریک
2) ناپارامتریک
3) توصیفی
4) استنباطی

۹۵- کدام یک از این نمودارها برای تحلیل اکتشافی مشاهدات استفاده می‌شود؟

1) جعبه‌ای
2) پاره‌تو
3) دایره‌ای
4) بافت‌نگار

۹۶- اگر N_i تعداد جامعه آماری i، μ_i، σ_i^2 به ترتیب میانگین و واریانس باشد. واریانس کل این جدول کدام است؟

N_i	100	200	700
μ_i	80	90	100
σ_i^2	1600	2500	2500

1) 2200
2) 2233/3
3) 2410
4) 2454

۹۷- واریانس داده‌ها با فرض جدول فراوانی کدام است؟

x_i	-1	0	1	2
F_i	2	3	4	1

1) 0/76
2) 0/78
3) 0/82
4) 0/84

۹۸- میانگین ۲۰ داده آماری ۱۵ و واریانس آن‌ها برابر ۲/۲۵ است. درصد ضریب تغییرات آن‌ها چقدر است؟

10 (1   12 (2   15 (3   20 (4

۹۹- در تبدیل داده‌های آماری با میانگین ۳۲ و انحراف‌معیـار ۳ بـه متغیرهـای هنجـاری (نرمـال)، داده ۴۱ متناظر با کدام هنجاری است؟

-1 (1   1 (2   2 (3   3 (4

۱۰۰- اگر میانگین داده‌های: $2x_1-3, 2x_2-3, \ldots, 2x_{20}-3$ برابر ۲۹ باشد، آن‌گاه $\sum x_i$ کدام خواهد بود؟

280 (1   320 (2   480 (3   640 (4

پاسخ تشریحی تست‌های فصل اول

1- گزینه (4) صحیح است.

چون میزان سود شرکت بر حسب درصد فروش بیان شده است، از میانگین هندسی به عنوان شاخص مرکزی استفاده می‌کنیم.

$$\overline{x}_G = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} \Rightarrow \overline{x}_G = \sqrt[5]{3 \times 4 \times 4 \times 3 \times 2} = \sqrt[5]{288} = 3/104$$

2- گزینه (3) صحیح است.

چون دامنه‌ی تغییرات (**R**) در گزینه‌ی (3) بین 1 و 8 می‌باشد، پس میزان پراکندگی این متغیرها را از سایر مجموعه‌ها بیش‌تر خواهد بود.

3- گزینه (3) صحیح است.

متوسط فروش با میانگین هندسی قابل محاسبه است.

$$\overline{x}_G = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} \Rightarrow \overline{x}_G = \sqrt[3]{2 \times 3 \times 4} = \sqrt[3]{8 \times 3} = 2\sqrt[3]{3} = 2/88$$

4- گزینه (3) صحیح است.

اگر 20% به میانگین حقوق کارمندان اضافه شود، داریم: هزار تومان 3 = 20% × 15بنابر خواص میانگین و انحراف معیار داریم: $\overline{x+3} = \overline{x} + 3 = 15 + 3 = 18$

$$\sigma_{(x+3)} = \sigma_x = 3$$

5- گزینه (3) صحیح است.

CL	F_i	x_i	$F_i x_i$	$F_i x_i^2$
20–30	10	25	250	6250
30–40	25	35	875	30625
40–50	15	45	675	30375
–	50	–	1800	67250

$$\overline{x} = \frac{\sum F_i x_i}{\sum F_i} = \frac{1800}{50} = 36$$

$$\sigma_x^2=\frac{\sum F_i x_i^2}{N}-(\bar{x})^2=\frac{67250}{50}-(36)^2=49$$

$$\sigma_x=\sqrt{\sigma_x^2}=\sqrt{49}=7$$

6- گزینه (2) صحیح است.

$$\sum_{i=1}^{n}(X_i-\bar{X})=\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\bar{X}=n\bar{X}-n\bar{X}=0$$

7- گزینه (3) صحیح است.

$$\bar{x}_G=\sqrt[n]{x_1\times x_2\times\ldots\times x_n}\Rightarrow \bar{x}_G=\sqrt[3]{21\times 44}=\%30/39$$

8- گزینه 3 صحیح است.

بنا به خواص انحراف معیار داریم: $y=(x+a)\frac{c}{b}$

$$\sigma_y=|\frac{c}{b}|\sigma_x\xrightarrow{\sigma_x=1}\sigma_y=|\frac{c}{b}|$$

9- گزینه (1) صحیح است.

$$\bar{x}=\frac{\sum n_i X_i}{n}=\frac{1200}{100}=12$$

$$\sigma_x^2=\frac{\sum n_i(X_i-\bar{X})^2}{n}=\frac{36}{100}=0/36\Rightarrow \sigma_x=\sqrt{0/36}=0/6$$

$$CV=\frac{\sigma_x}{\bar{x}}\times 100=\frac{0/6}{12}\times 100=\%5$$

10- گزینه (4) صحیح است.

$$\mu_3=\frac{\sum(X_i-\bar{X})^3}{n}=\frac{-328}{20}=-16/4$$

$$\sigma_x=\frac{\sum(X_i-\bar{X})^2}{n}=\frac{1620}{20}=81\Rightarrow \sigma=\sqrt{81}=9$$

$$SK=\frac{\mu_3}{\sigma^3}=\frac{-16/4}{(9)^3}=\frac{-16/4}{729}=-0/022$$

11- گزینه (3) صحیح است.

$$\sum|x_i-me|=\min$$

12- گزینه (3) صحیح است.

13- گزینه (1) صحیح است.

برای جامعه با چولگی مثبت داریم $\overline{x} > me > mo$.

14- گزینه (2) صحیح است.

سرمایه گذاری در شرکت (ب) مناسب‌تر است. زیرا دارای پراکندگی کم‌تر و ضریب چولگی صفر است.

15- گزینه (4) صحیح است.

می‌دانیم انحراف معیار عدد ثابت برابر صفر است و میانگین عدد ثابت برابر خود آن عدد است، بنابراین:

$$CV = \frac{\sigma}{\overline{x}} \times 100 = \frac{0}{5} \times 100 = 0$$

16- گزینه (1) صحیح است.

17- گزینه (2) صحیح است.

$$mo - \overline{x} = 3(me - \overline{x}) \Rightarrow mo - 52/4 = 3(51/8 - 52/4)$$

$$\Rightarrow mo = 50/6$$

18- گزینه (3) صحیح است.

از میانگین هارمونیک استفاده می‌کنیم.

$$\overline{x}_H = \frac{n}{\sum \frac{1}{x_i}} = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{6}} = 3$$

19- گزینه (4) صحیح است.

$$\sigma_{(x+4)} = \sigma_x = 2/4$$

20- گزینه (2) صحیح است.

21- گزینه (1) صحیح است.

$$\begin{aligned} P_n = P_0(1+r)^n \quad & \Rightarrow P_{1374} = P_{1370}(1+r)^4 \\ & \Rightarrow 1600 = 100(1+r)^4 \\ & \Rightarrow (1+r)^4 = 16 = 2^4 \Rightarrow 1+r = 2 \Rightarrow r = 1 \\ & \Rightarrow \%r = 1 \times 100 = 100 \end{aligned}$$

22- گزینه (2) صحیح است.

23- گزینه (1) صحیح است.

24- گزینه (4) صحیح است.

25- گزینه (1) صحیح است.

26- گزینه (3) صحیح است.

$$\bar{x}=\frac{\sum F_i x_i}{N}=\frac{(3\times1)+(5\times1)+(7\times5)+(9\times3)}{10}=7$$

$$me=L_i+\left(\frac{\frac{N}{2}-FC_{i-1}}{F_i}\right)C=6+\left(\frac{5-2}{5}\right)\times2=7/2$$

$$A_s=\frac{3(\bar{x}-me)}{\sigma}=\frac{3(7-7/2)}{1/8}=-0/33$$

27- گزینه (3) صحیح است.

28- گزینه (2) صحیح است.

$$\bar{x}=\frac{\sum x_i}{n}=\frac{20}{10}=2$$

$$\sigma^2=\frac{\sum x_i^2}{n}-(\bar{x})^2=\frac{200}{10}-2^2=16\Rightarrow\sigma_x=4$$

$$A_s=\frac{\bar{x}-mo}{\sigma}=\frac{2-3}{4}=-0/25$$

29- گزینه (4) صحیح است.

$$\bar{x}=\frac{\sum x_i}{n}=\frac{250}{50}=5$$

$$\sigma^2=\frac{\sum x_i^2}{n}-(\bar{x})^2=\frac{2500}{50}-5^2=25\Rightarrow\sigma_x=5$$

$$CV=\frac{\sigma_x}{\bar{x}}\times100=\frac{5}{5}\times100=\%100$$

30- گزینه (2) صحیح است.

$$x+0/25x=\bar{x}+0/25\bar{x}=50+(0/25\times50)=62/5$$

$$\sigma_{(x+0/25x)}=\sigma_x+0/25\sigma_x=20+(0/25\times20)=25$$

CV قبل از افزایش حقوق‌ها:

$$CV=\frac{\sigma_x}{\bar{x}}\times 100=\frac{25}{62/5}\times 100=\%40$$

بنابراین CV تغییری نخواهد کرد.

$$CV=\frac{\sigma_x}{\bar{x}}\times 100=\frac{20}{50}\times 100=\%40$$

31- گزینه (1) صحیح است.

32- گزینه (2) صحیح است.

می‌دانیم که اگر از همه‌ی داده‌ها مقدار ثابتی را کم کنیم، واریانس تغییری نخواهد کرد. بنابراین کافی است واریانس اعداد (0، 2، 4) را به دست آوریم.

$$\bar{x}=\frac{\sum x_i}{n}=\frac{0+2+4}{3}=2$$

$$V(x)=\frac{\sum (x_i-\bar{x})^2}{n-1}=\frac{4+0+4}{3-1}=\frac{8}{2}=4$$

33- گزینه (3) صحیح است.

$$\sum (x_i-\bar{x})^2=\min$$

34- گزینه (3) صحیح است.

35- گزینه (2) صحیح است.

36- گزینه (3) صحیح است.

$$\sigma^2_{(4x+7)}=4^2\times\sigma^2_x=16\times 50=800$$

37- گزینه (1) صحیح است.

$$\mu=\frac{\sum x_i}{N}\Rightarrow 2=\frac{0+a+20+3b+12}{28}\rightarrow a+3b=24$$

تساوی بالا به ازای $a=b=6$ برقرار است.

38- گزینه (2) صحیح است.

39- گزینه (3) صحیح است.

از میانگین هم‌ساز (هارمونیک) استفاده می‌کنیم زیرا مقیاس سنجش داده‌ها ترکیبی (کیلومتر در ساعت) است.

40- گزینه (2) صحیح است.

$$\mu = \frac{\sum N_i\mu_i}{N} = \frac{(20\times15)+(30\times10)}{50} = \frac{600}{50} = 12$$

$$\sigma^2 = \frac{\sum N_i\sigma_i^2}{N} + \frac{\sum N_i(\mu_i-\mu)^2}{N} = \frac{(20\times17)+(30\times12)}{50} + \frac{20(15-12)^2+30(10-12)^2}{50} = 20$$

41- گزینه (3) صحیح است.

در توزیع با چولگی منفی داریم $mo > me > \bar{x}$.

42- گزینه (2) صحیح است.

یکی از کاربردهای ضریب پراکندگی (ضریب تغییرات) در جایی است که در دو یا چنـد جامعـه‌ی آمـاری مـورد مقایسـه، مشاهدات نامتجانس وجود داشته باشد.

43- گزینه (2) صحیح است.

از میانگین هندسی استفاده می‌کنیم:

$$\bar{x}_G = \sqrt{\frac{P_{79}}{P_{78}}.\frac{P_{80}}{P_{79}}} = \sqrt{\frac{P_{80}}{P_{78}}} = \sqrt{\frac{450}{200}} = 1/5$$

$$1/5 - 1 = 0/5$$

که نشان دهنده‌ی 50% افزایش قیمت به طور متوسط است. راه حل دیگر.

$$A_1 = A_0(1+r)^t \Rightarrow 450 = 200(1+r)^2 \Rightarrow 1+r = \sqrt{\frac{450}{200}}$$

$$\Rightarrow 1+r = 1/5 \Rightarrow r = ۰/۵ \xrightarrow{\text{درصد}} \%50$$

44- گزینه (4) صحیح است.

میزان مصرف نفت سفید در چهار سال متوالی عبارت است از:

$$t_1 = \frac{20000}{1/6} = 12500 \Rightarrow t_2 = \frac{20000}{1/8} = 11111/11$$

$$t_3 = \frac{200000}{2/1} = 9523/81 \Rightarrow t_4 = \frac{20000}{2/5} = 80000$$

جمع کل مصرف برابر است با:

$$12500 + 11111/11 + 9523/81 + 80000 = 41134/92$$

کل مبلغ پرداختی برای هزینه سوخت 4 سال:

$$20000 \times 4 = 80000$$

متوسط هزینه سوخت خانوار به ازای یک لیتر در 4 سال:

$$80000 \div 41134/92 = 1/94$$

45- گزینه (3) صحیح است.

$$\sum |x_i - me| = \min$$

46- گزینه (1) صحیح است.

$$V(x) = d^2\left(\frac{N^2-1}{12}\right) = 1\left(\frac{25-1}{12}\right) = 2$$

47- گزینه (3) صحیح است.

48- گزینه (3) صحیح است.

$$\sigma^2_{(4x-3)} = 4^2 \times \sigma^2_x = 16 \times 9 = 144 \Rightarrow \sigma_{(4x-3)} = \sqrt{144} = 12$$

یا: $\sigma^2_x = 9 \Rightarrow \sigma_x = 3 \Rightarrow \sigma_{(4x-3)} = 4\sigma_x = 4 \times 3 = 12$

49- گزینه (2) صحیح است.

$$mo - \overline{x} = 3(me - \overline{x}) \Rightarrow mo - 52/4 = 3(51/8 - 52/4) \Rightarrow mo = 50/6$$

50- گزینه (1) صحیح است.

51- گزینه (1) صحیح است.

$$m_1 = \frac{\sum n_i x_i}{n} = \frac{40}{10} = 4$$

$$m_2 = \frac{\sum n_i x_i^2}{n} = \frac{400}{10} = 40$$

$$m_3 = \frac{\sum n_i x_i^3}{n} = \frac{600}{10} = 60$$

$$\mu_2 = m_3 - 3m_2m_1 + 2m_1^3 = 60 - 3(40)(4) + 2(4)^3 = -292$$

52- گزینه (3) صحیح است.

53- گزینه (4) صحیح است.

54- گزینه (2) صحیح است.

$$\overline{x} = \frac{\sum X_i}{N} = \frac{100}{20} = 5 \Rightarrow \sigma^2_x = \frac{\sum X_i^2}{N} - \overline{X}^2 = \frac{580}{20} - 25 = 4$$

$$CV = \frac{\sigma_x}{\overline{x}} \times 100 = \frac{2}{5} \times 100 = \%40$$

55- گزینه (1) صحیح است.

$$\bar{x}-mo=3(\bar{x}-me)\Rightarrow 30-mo=3(30-50)\Rightarrow mo=90$$

56- گزینه (2) صحیح است.

توزیع از توزیع نرمال کوتاهتر $E=-0/71<0\Rightarrow$

اختلاف توزیع و توزیع نرمال فاحش است $|E|=|-0/71|>0/5\Rightarrow$

57- گزینه (3) صحیح است.

58- گزینه (3) صحیح است.

$$N=\sum F_i=16\Rightarrow \frac{N}{2}=8$$

CL	20-29	30-39	40-49
F_i	3	6	7
FC_i	3	9	16

پس میانه در طبقه‌ی دوم قرار دارد. $me=L_i+\left(\dfrac{\frac{N}{2}-FC_{i-1}}{F_i}\right)C$

$$=29/5+\left(\frac{8-3}{6}\right)\times 10=37/83$$

59- گزینه (4) صحیح است.

60- گزینه (1) صحیح است.

$$sk=\frac{\mu_3}{\sigma^3}=\frac{1/92}{64}=0/03$$

$$\mu_3=\frac{\sum(x_i-\mu_X)^3}{n}=\frac{96}{50}=1/92$$

$$\sigma^2=\frac{\sum X_i^2}{n}-(\mu_X)^2=\frac{3250}{50}-49=16\Rightarrow\sigma=4$$

61- گزینه (1) صحیح است.

دستگاه A دارای واریانس کمتری است، بنابراین میزان پراکندگی کم‌تری دارد. پس دقیق‌تر است.

62- گزینه (3) صحیح است.

چون عدد 9 بیش‌ترین فراوانی را دارد.

63- گزینه (1) صحیح است.

$$\mu_X=\frac{\sum_{i=1}^{N}X_i}{N}\Rightarrow\sum_{i=1}^{N}X_i=N.\mu_X$$

$$\mu_Y=\frac{\sum_{i=1}^{K}Y_i}{K}\Rightarrow\sum_{i=1}^{K}Y_i=K.\mu_Y$$

$$\frac{\sum_{i=1}^{N}X_i}{\sum_{i=1}^{K}X_i}=\frac{N.\mu_X}{K.\mu_Y}=\frac{N.\mu_X}{K.a\mu_X}=\frac{N}{Ka}$$

64- گزینه (4) صحیح است.

هرگاه داده‌های آماری با هم برابر باشند، انحراف معیار برابر صفر است و بالعکس.

65- گزینه (2) صحیح است.

$$(CV)_A=\frac{\sigma_X}{\overline{x}_A}\times100=\frac{2}{17}\times100=\%11/76$$

$$(CV)_B=\frac{\sigma_B}{\overline{x}_B}\times100=\frac{30}{250}\times100=\%12$$

بنابراین میزان پراکندگی کارخانه B بیش از کارخانه A است.

66- گزینه (2) صحیح است.

67- گزینه (2) صحیح است.

$$x_1=\frac{4}{2}=2,x_2=\frac{32}{4}=8$$

$$\overline{x}_G=\sqrt{x_1.x_2}=\sqrt{2\times8}=\sqrt{16}=4$$

68- گزینه (4) صحیح است.

$$\sigma^2_{(2X-3)}=2^2\sigma^2_X=4\times4=16$$

69- گزیه (3) صحیح است.

70- گزینه (1) صحیح است.

71- گزینه (1) صحیح است.

72- گزینه (3) صحیح است.

مد در طبقه اول قرار دارد.

$$mo=L_i+\left(\frac{d_1}{d_1+d_2}\right)C=-10+\left(\frac{30}{30+10}\right)\times10=-2/5$$

73- گزینه (4) صحیح است.

$$\overline{y}=\frac{1}{2}\overline{x}+1=\frac{1}{2}(3)+1=2/5$$

$$\sigma_x^2=9\Rightarrow\sigma_x=3,\sigma_y=\frac{1}{2}\sigma_x=\frac{1}{2}\times3=1/5$$

74- گزینه (3) صحیح است.
$$\bar{x}-mo=3(\bar{x}-md)\Rightarrow 100-mo=3(100-120)\Rightarrow mo=160$$

75- گزینه (3) صحیح است.
$$N=\sum F_i=60, \frac{3N}{4}=45$$

CL	10-20	20-30	30-40
F_i	10	30	20
FC_i	10	40	60

پس چارک سوم در طبقه‌ی سوم قرار دارد.
$$Q_3=L_i+\left(\frac{\frac{3N}{4}-FC_{i-1}}{F_i}\right)C$$

$$Q_3=30+\left(\frac{45-40}{20}\right)\times 10=32/5$$

76- گزینه (4) صحیح است.

با جامعه نرمال تفاوت اندکی دارد $\Rightarrow$ توزیع از توزیع نرمال کوتاه‌تر
$$E=-0/6<\mathbf{0}\Rightarrow$$

77- گزینه (4) صحیح است.
$$\sigma_x^2=\frac{\sum X_i}{N}-(\bar{x})^2\Rightarrow 14=\frac{390}{5}-(\bar{x})^2\Rightarrow \bar{x}=8$$

78- گزینه (1) صحیح است.
$$\bar{x}=\frac{\sum X_i}{N}=\frac{200}{100}=2$$

$$\sigma_x^2=\frac{\sum X_i^2}{N}-(\bar{x})^2=\frac{500}{100}-4=1\Rightarrow \sigma_x=1$$

$$CV=\frac{\sigma_x}{\bar{x}}\times 100=\frac{1}{2}\times 100=\%50$$

79- گزینه (2) صحیح است.
$$N=\sum F_i=60. \frac{N}{4}=\frac{60}{4}=15$$

CL	2-5	6-9	10-12
F_i	10	30	20
FC_i	10	40	60

$$Q_1=L_i+\left(\frac{\frac{N}{4}-FC_{i-1}}{F_i}\right)C$$

$$Q_1=5/5+\left(\frac{15-10}{30}\right)\times 4=6/17$$

80- گزینه (1) صحیح است.
$$\mu_4=\frac{\sum F_i(x_i-\mu_x)^4}{N}=\frac{5000}{1000}=5$$

$$E=\frac{\mu_4}{\sigma^4}-3=\frac{5}{16}-3=-2/687$$

81- گزینه (1) صحیح است.

$$\bar{x}_H = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}} = \frac{3}{\frac{1}{120}+\frac{1}{60}+\frac{1}{90}} = \frac{1080}{13} = 83/076$$

82- گزینه (2) صحیح است.

$$CV = \frac{\sigma}{\bar{x}}\times 100 \Rightarrow \%20 = \frac{\sigma}{12}\times 100 \Rightarrow \sigma = 2/4$$

83- گزینه (4) صحیح است.

$$\mu_2 = \frac{\sum (x_i-\mu_x)^2}{N} = \frac{40}{10} = 4 \Rightarrow \sigma = 2$$

$$\mu_3 = \frac{\sum (x_i-\mu_x)^3}{N} = \frac{80}{10} = 8$$

$$sk = \frac{\mu_3}{\sigma^3} = \frac{8}{8} = 1$$

84- گزینه (3) صحیح است.

$$N = 30, \frac{2\times N}{10} = \frac{2\times 30}{10} = 6$$

CL	40-50	50-60	60-7
F_i	5	18	7
FC_i	5	23	30

پس دهک در طبقه دوم قرار دارد.

$$D_2 = L_i + \left(\frac{\frac{2N}{10} - FC_{i-1}}{F_i}\right)C$$

$$D_2 = 50 + \left(\frac{6-5}{18}\right)10 = 50/55C$$

85- گزینه (1) صحیح است.

86- گزینه 1 صحیح است.

87- گزینه (4) صحیح است.

88- گزینه (3) صحیح است.

89- گزینه (1) صحیح است.

90- گزینه (3) صحیح است.

$$x'_i = x_i + 0/125x_i \Rightarrow \bar{x}' = \bar{x} + 0/125\bar{x} = 80 + (0/125\times 80) = 90$$

$$\sigma_{(x_i+0/125x_i)} = \sigma_{(1/125x_i)} = 1/125\times\sigma_x = 1/125\times 20 = 22/5$$

$$CV = \frac{\sigma}{\bar{x}}\times 100 = \frac{22/5}{90}\times 100 = \%25$$

91ـ گزینه (4) صحیح است.

92- گزینه (3) صحیح است.

برای محاسبه‌ی میانگین هارمونیک موزون مشاهدات از رابطه زیر استفاده می‌کنیم.

$$\bar{x}_H = \frac{\sum \omega_i}{\frac{\omega_1}{x_1}+\frac{\omega_2}{x_2}+\dots+\frac{\omega_k}{x_k}}$$

بنابراین:

$$\bar{x}_H = \frac{1+\frac{1}{3}+\frac{2}{3}}{\frac{1}{100}+\frac{\frac{1}{3}}{80}+\frac{\frac{2}{3}}{120}} = 101/4$$

93- گزینه (3) صحیح است.

با توجه به جدول توزیع فراوانی مد در طبقه دوم است.

$$mo = L_i + \left(\frac{d_1}{d_1+d_2}\right)C = 5/5 + \left(\frac{16}{16+8}\right)(3) = 7/5$$

94- گزینه (2) صحیح است.

فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از توزیع نرمال است، ولی در آمار ناپارامتریک این فرض ضرورتی ندارد. اصطلاحاً گفته می‌شود آمار ناپارامتریک آزاد است.

95- گزینه (1) صحیح است.

96- گزینه (4) صحیح است.

$$\mu = \frac{\sum N_i\mu_i}{N} = \frac{(100\times 80)+(200\times 90)+(700\times 100)}{100+200+700} = 96$$

$$\sigma^2 = \frac{\sum N_i\sigma_i^2}{N} + \frac{\sum N_i(\mu_i-\mu)^2}{N}$$

$$\sigma^2 = \frac{(100\times 1600)+(200\times 2500)+(700\times 2500)}{1000} + \frac{100(-16)^2+200(-6)^2+700(4)^2}{1000} = 2454$$

97- گزینه (4) صحیح است.

$$\sum F_i x_i = -2+\mathbf{o}+4+2 = 4$$

$$\sum F_i x_i^2 = 2+\mathbf{o}+4+4 = 10$$

$$\sigma^2 = \frac{\sum F_i x_i^2}{N} - \left(\frac{\sum F_i x_i}{N}\right)^2 = \frac{10}{10} - \left(\frac{4}{10}\right)^2 = 1-0/16 = 0/84$$

98- گزینه (1) صحیح است.

$$\sigma_x^2 \Longrightarrow \sigma_x = \sqrt{2/25} = 1/5$$

$$CV = \frac{\sigma_x}{\overline{x}} \times 100 = \frac{1/5}{15} \times 100 = \%10$$

99- گزینه (4) صحیح است.

$$Z = \frac{x - \overline{x}}{\sigma} = \frac{41 - 32}{3} = 3$$

100- گزینه (2) صحیح است.

$$(2x_1 - 3) + (2x_2 - 3) + \ldots + (2x_{20} - 3) = \sum_{i=1}^{20}(2x_i - 3) = 2\sum_{i=1}^{20}(x_i - \frac{3}{2})$$

$$\overline{x} = \frac{2\sum_{i=1}^{20}(x_i - \frac{3}{2})}{20} = 29 \Rightarrow \frac{2\sum_{i=1}^{20}(x_i - \frac{3}{2})}{20} = 29 \Rightarrow$$

$$\frac{\sum x_i - \frac{3}{2} \times 20}{10} = 29 \Rightarrow \sum x_i = 290 + 30 = 320$$

$$y = 2x - 3 \Rightarrow \overline{y} = 2\overline{x} - 3 \Rightarrow 29 = 2\overline{x} - 3 \Rightarrow \overline{x} = 16$$

$$\overline{x} = \frac{\sum x_i}{N} \Rightarrow 16 = \frac{\sum x_i}{20} \Rightarrow \sum x_i = 320$$

آزمون خودسنجی فصل اول

۱- در یک توزیع، با تقسیم مقادیر صفت بر عدد ۵، واریانس ۱۰ به دست آمد. واریانس اصلی مقـادیر صفت چند است؟

50(1   100(2   250(3   500(4

۲- در یک توزیع نزدیک به نرمال، انحراف استاندارد(میانگین قدر مطلق انحرافات) برابـر ۸ اسـت. انحراف معیار توزیع برابر با چه عددی است؟

4(1   6/4(2   10(3   /36(4

۳- در یک توزیع چوله که چولگی آن به سمت راست است، کدام یک از روابط زیر حاکم است؟

1)میانگین = میانه = نما
2)میانگین>میانه>نما
3)میانگین< میانه< نما
4)میانگین>میانه< نما

۴- توزیع متقارن چه موقع اتفاق می‌افتد؟

1)زمانی که میانگین و میانه و نما یکی باشند.
2) زمانی که میانگین بزرگتر از میانه و نما باشد.
3) زمانی که نما بزرگتر از میانه و میانگین باشد.
4) زمانی که مقدار هر دو مساوی یک باشد.

۵- مورد استفاده میانه بیشتر در مورد کدام یک از مقیاس‌های اندازه‌گیری می‌باشد؟

1) رتبه‌ای   2) فاصله‌ای   3)اسمی   4) نسبی

۶- کم کردن فراوانی‌ها از وسط یک توزیع سبب کدام یک از موارد زیر می‌شود؟

1) باعث افزایش دامنه تغییر می‌گردد
2) تغییری در انحراف استاندارد ایجاد نمی‌کند.
3) سبب کاهش انحراف استاندارد می‌شود.
4) سبب افزایش انحراف استاندارد می‌شود.

۷- اگر همه عناصر یک سری نمره را در ۴ ضرب کنیم، واریانس آنها ۶۴ خواهد بود، واریانس نمـره-های اصلی چقدر است؟

8 (1   2 (2   4 (3   16 (4

۸- واریانس مقادیر(۱۰ و ۸ و ۶ و ۴ و ۲)کدام است؟

4(1 6(2 8(3 10(4

۹- برای مقایسه میزان پراکندگی دو متغیر سن و قد، کدام شاخص مناسب است؟

1)واریانس 2) انحراف معیار 3)میدان تغییرات 4) ضریب تغییرات

۱۰- در یک توزیع چوله، میانگین برابر با ۱۴ و میانه برابر با ۱۲ می‌باشد، نمای این توزیع کدام است؟

8 (1 10(2 12(3 26(4

۱۱- اگر انحراف معیار برابر با صفر باشد، می‌توان نتیجه گرفت...............

1)داده‌ها همگن می‌باشند.

2)میانگین، ابزار اندازه‌گیری خوبی برای بررسی حد متوسط است.

3)هیچ پراکندگی در داده‌ها وجود ندارد.

4)تمام موارد

۱۲- دانشکده هنر و علوم دانشگاه تهران ۹ گروه دارد، تعداد اعضای هیئت علمی در هر گروه عبارت است از ۷ و ۱۴ و ۱۳ و ۱۱ و ۱۷ و ۱۵ و ۹ و ۱۲ و ۸ تعداد میانه اعضای هیئت علمی در دو دانشکده کدام است؟

4/5(1 5(2 12(3 17(4

۱۳- مناسب‌ترین کاربرد شاخصی نما(مد)در مورد چه نوع مقیاس متغیرهاست؟

1)رتبه‌ای 2)اسمی 3)فاصله‌ای 4)نسبی

۱۴- از بین شاخص‌های مرکزی معرفی یافته‌های تحقیق«میانگین» می‌تواند چه معایبی داشته باشد؟

1)معرف تفکیک تمامی اعداد و نمونه‌ها نباشد. 2)تحت تأثیر اعداد اغراق آمیز باشد.

3)نیاز به تعیین انحراف معیار ندارد. 4)تحت تأثیر انحراف استاندارد باشد.

۱۵- نمره هوشی افراد در یک جامعه دارای توزیع نرمال با میانگین ۱۰۰ و انحراف استاندارد ۱۵ مـی-باشد، درصد افرادی که نمره آنها از ۸۵ کمتر باشد، تقریباً برابر است با چند درصد؟

1) 68 2) 32 3) 34 4) 16

۱۶- کدام یک از آماره‌های گرایش مرکزی برای هر سطحی از اندازه‌گیری قابل بکارگیری است؟

1) انحراف معیار 2) میانه 3) میانگین 4) نما

۱۷- میزان خودکشی در ۱۰۰ روستای بزرگ در جدول زیر ارائه شده است، میزان خودکشی از کـدام نوع توزیع برخوردار است؟

طبقات	فراوانی
0-2	5
3-5	10
6-8	15
9-11	20
12-14	24
15-17	26

1) متقارن

2) چولگی راست(مثبت)

3) چولگی چپ(منفی)

4) هموار

۱۸- با توجه به سوال ۱۷، در توزیع میزان خودکشی چه رابطه‌ای بین گـرایش‌هـای مرکـزی وجـود دارد؟

1) نما>میانگین>میانه 2) نما> میانه > میانگین

3) میانه>نما>میانگین 4) میانه>میانگین>نما

۱۹- میانگین نمرات درس «نقد و بررسی» دانشجویان ۱۶ و انحراف معیار ۱/۲۵ بوده است، نمره خـام دانشجویی که نمره استاندارد(۲-) در این درس داشته است، چند می‌باشد؟

1) 14/5 2) 13/5 3) 17/25 4) 18/25

۲۰- توزیعی داریم که میانگین۷۵/۷۰ ومیانه ۷۲/۲۳ می‌باشد، این توزیع چگونه می‌باشد؟

1) چولگی منفی 2) دارای پراکندگی کم

3) چولگی مثبت 4) نزدیک به توزیع نرمال

۲۱- اگر یک توزیع نمره استاندارد(z)یک فرد برابر با(۰/۸-) باشد، کدام عبارت صحیح است؟

1) نمره فرد از نمره میانگین کمتر است.

2)نمره فرد $\frac{1}{3}$نمره میانگین است.

3)نمره فرد تقریباً نصف نمره میانگین است.

4) نمره فرد به اندازه(0/8)از نمره میانگین کمتر است.

۲۲- اگر در یک توزیع، فراوانی نسبی دارا بودن یک صفت برابر با ۰/۷ و دارا نبودن آن برابر با ۰/۳ باشد، واریانس توزیع چقدر است؟

0/1(1   0/21(2   0/4(3   2/33(4

۲۳- در توزیع روبرو مقدار چولگی چقدر است؟

$x_i - \bar{x}$
-2
-1
0
1
2

1)صفر

1(2

2(3

6/8(4

۲۴- اگر میانگین برابر با ۱۰۰ و انحراف معیار یا انحراف استاندارد برابر با ۲۰ باشد، ضریب تغییر نسبی(CRV)، برابر با چه عددی است؟

20(1   80(2   120(3   500(4

۲۵- در صورتی که قصد محاسبه مناسب‌ترین شاخص مرکزی را باتوجه به نوع توزیع نمرات داشته باشیم و در توزیع مورد نظر، اعداد نوعی وجود داشته باشند، کدام یک از شاخص‌های زیر مناسب‌تر است؟

1) میانگین   2)میانه

3)نما   4)هر کدام از شاخص‌ها

۲۶- چنانچه توزیع نمرات زیر موجود باشد و بخواهیم مناسب‌ترین شاخص پراکندگی را محاسبه کنیم، کدام‌یک از شاخص‌های زیر مناسب‌تر است؟۱۹- ۱۸۰ – ۱۴۰- ۱۴- ۲۹- ۲۵- ۱۵- ۱۷- ۱۸

1)انحراف چارکی 2)انحراف استاندارد 3)واریانس 4)چارک سوم

۲۷- نمرات استاندارد دارای کدام‌یک از ویژگی‌های زیر هستند؟

1)میانگین صفر و انحراف استاندارد یک 2)انحراف استاندارد یک و میانگین صفر

3)واریانس یک و میانگین یک 4)واریانس بزرگ و انحراف استاندارد بزرگ

۲۸- اگر متوسط انحراف از میانه معادل ۴۰ و میانه ۸۰ باشد، ضریب تغییر نسبی(CRV)متوسط انحراف از میانه چقدر است؟

0/5(1 0/25(2 1(3 2(4

۲۹- واریانس نمرات ۵۰۵ و ۵۰۴ و ۵۰۳ و ۵۰۲ و ۵۰۱ چقدر است؟

5(1 2(2 502(3 502/5(4

۳۰- در صورتی که میانگین یک توزیع مساوی ۴۵ باشد و به هر یک از اعداد این توزیع عدد ۵ را اضافه کنیم، میانگین توزیع جدید مساوی کدام‌یک از مقادیر زیر است؟

50(1 45(2 40(3 47/5(4

۳۱- در صورتی که میانه به عنوان مناسب‌ترین شاخص مرکزی مورد استفاده قرار گیرد، کدام‌یک از شاخص‌های زیر مناسب‌ترین شاخص برای تعیین پراکندگی است؟

1)متوسط انحراف 2)انحراف استاندارد 3)واریانس 4)انحراف چارکی

۳۲- اگر ضریب کشیدگی مثبت باشد، تجمع افراد حول میانگین توزیع..............توزیع نرمال است.

1) کمتر از 2)همانند 3)بسیار کمتر از 4)بیشتر از

۳۳- اگر در یک توزیع، میانگین برابر ۲۰ و میانه برابر ۲۲ باشد، نما چه عددی است؟

18(1 36(2 26(3 21(4

۳۴- در یک آزمون با توزیع طبیعی میانگین و انحراف معیار به ترتیب مساوی ۵۰ و ۵۰ است، نمره خام معادل $z=-2$ چقدر است؟

1) 40 2) 60 3) +10 4) -10

۳۵- مجموع انحراف هرتوزیع از میانگین چقدر است؟

1) 1 2) 0<1 3) 0>1 4) صفر

۳۶- متغیر تصادفی X دارای توزیع طبیعی با میانگین ۱۰۰ و انحراف استاندارد ۵ است، نمره ۱۱۰ چند انحراف استاندارد بالاتر از میانگین است؟

1) 3 2) 2 3) -2 4) 1

۳۷- نمرات آزمون هوش دانشجویان، یک توزیع نرمال با میانگین ۷۲ و واریانس ۱۶ می‌باشد. نمرات چند درصد این دانشجویان بین ۶۸و ۷۶ قرار دارد؟ $(p(z<1)=0/1587)$

1) 66 2) 67 3) 68 4) 70

۳۸- در یک توزیع آماری ضریب کشیدگی برابر با $-0/7$ است، در نتیجه پراکندگی در اطراف میانگین............از توزیع نرمال است.

1- کمی کمتر 2- خیلی کمتر 3- خیلی بیشتر 4- کمی بیشتر

۳۹- میانه و مد در یک جامعه آماری به ترتیب ۴۲ و ۵۰ است، توزیع این جامعه از چولگی معقولی بر خوردار است. میانگین و نوع چولگی کدام است؟

1) 38، چوله به چپ 2) 38،چوله به راست 3) 40، چوله به چپ 4) 40، چوله به راست

۴۰- اگر $p(z\le -2)=0/023$ و متغیر x دارای توزیع نرمال با میانگین۲۹ و $p(x\ge 5)=0/977$ باشد. انحراف معیار xکدام است؟

1) 9 2) 10 3) 11 4) 12

۴۱- کدام مقیاس از ویژگی‌های بهتری برای اندازه‌گیری برخوردار است؟

1) اسمی 2) رتبه‌ای 3) نسبی 4) فاصله‌ای

۴۲- در توزیع ۲- ۱ – ۰ – ۴ – ۵- ۲۶۲ مناسب‌ترین شاخص پراکندگی عبارتست از:

1) واریانس 2) انحراف استاندارد 3) چارک‌ها 4) انحراف چارکی

۴۳- در صورتیکه توزیع نمرات درآمد در یک جامعه بصورت کمّی باشد، مناسب‌ترین شاخص مرکزی در ایـن شرایط عبارت است از :

1) میانگین 2)نما 3)میانه 4)نما و میانگین

پاسخنامه آزمون خودسنجی فصل اول

۱- گزینه ۳ صحیح است.

در رابطه با واریانس باید به نکات زیر توجه شود:

الف) اگر در یک توزیع، از تمامی مقادیر صفت، یک مقدار ثابت **a** کسر یا اضافه کنیم، واریانس این مقادیر تغییری نخواهد کرد.

ب) اگر در یک توزیع، تمامی مقادیر صفت را بر مقدار ثابت **a** تقسیم کنیم، واریانس متغیر اصلی a^2 برابر کوچکتر می‌شود. یعنی واریانس اعداد حاصله برابر است با واریانس قبلی تقسیم بر a^2

ج) اگر در یک توزیع، تمامی مقادیر صفت را در مقدار ثابت **a** ضرب کنیم، واریانس متغیر اصلی a^2 برابر بزرگتر می‌شود. یعنی واریانس اعداد حاصله برابر است با واریانس قبلی ضرب در a^2

در این سوال تمامی مقادیر صفت را بر مقدار ثابت**a** تقسیم شده است، واریانس اعداد حاصله برابر است با واریانس قبلی تقسیم بر a^2:

$$s^2 \div a^2 = x$$
$$s^2 \div 5^2 = 10 \Rightarrow s^2 \div 25 = 10$$
$$s^2 = 250$$

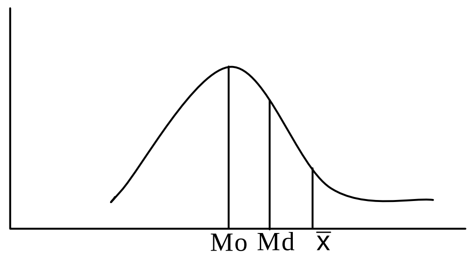
۲- گزینه ۳ صحیح است.

همواره انحراف معیار بزرگتر از انحراف متوسط(میانگین قدر مطلق انحرافات) است. در واقع انحراف معیار، 1/25برابر انحراف متوسط است.

$$s = 1/25md$$
$$md = 8 \Rightarrow 8 \times 1/25 = 10$$
$$s = 10$$

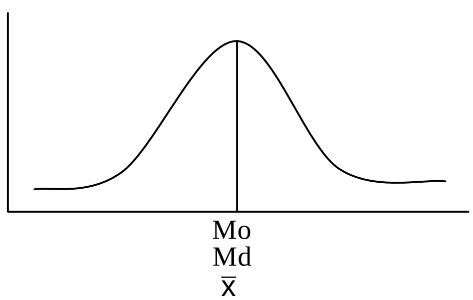
۳- گزینه ۲ صحیح است.

کجی(چولگی) زمانی مثبت است که میانگین بزرگتر از میانه و مد باشد. به بیان آماری: **M>Md>Mo**



۴- گزینه ۱ صحیح است.

نمودار توزیع فراوانی زمانی متقارن است که میانگین و میانه و مد با هم برابر باشند. در اینجا نمودار به صورت توزیع طبیعی می‌باشد. به بیان آماری: **M=Md=M**



۵- گزینه ۱ صحیح است.

برای توصیف یافته‌های متغیرهای ترتیبی از شاخص مرکزی میانه استفاده می‌شود.(ساعی:62)

۶- گزینه۴ صحیح است.

اگر به تمامی داده‌های یک توزیع عدد ثابت**a** را اضافه یا کسر کنیم، انحراف معیار هیچ تغییری نخواهد کرد.(دلاور،1387 :118)

۷- گزینه ۳ صحیح است.

در رابطه با واریانس باید به نکات زیر توجه شود:

الف) اگر در یک توزیع، از تمامی مقادیر صفت، یک مقدار ثابت **a** کسر یا اضافه کنیم، واریانس این مقادیر تغییری نخواهد کرد.

ب) اگر در یک توزیع، تمامی مقادیر صفت را بر مقدار ثابت **a** تقسیم کنیم، واریانس متغیر اصلی a^2 برابر کوچکتر می‌شود. یعنی واریانس اعداد حاصله برابر است با واریانس قبلی تقسیم بر a^2

ج) اگر در یک توزیع، تمامی مقادیر صفت را در مقدار ثابت **a** ضرب کنیم، واریانس متغیر اصلی a^2 برابر بزرگتر می‌شود. یعنی واریانس اعداد حاصله برابر است با واریانس قبلی ضرب در a^2

در این سوال تمامی مقادیر ضربدر 4 شده است، واریانس حاصله مساوی است با 64، یعنی a^2 برابر بزرگتر شده است، پس واریانس نمره‌های اصلی برابر است با:

$$s^2 = x \times 4^2 = 64$$
$$x \times 16 = 64 \Rightarrow x = \frac{64}{16} = 4$$
$$s^2 = 4$$

۸- گزینه ۳ صحیح است.

$$\bar{x} = \frac{\sum xi}{N} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$$
$$S^2 = \frac{\sum (xi-\bar{x})^2}{N} = \frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5}$$
$$= \frac{16+4+0+4+16}{5} = \frac{40}{5} = 8$$

۹- گزینه ۴ صحیح است.

ضریب پراکندگی یا ضریب تغییر(**V**)، شاخصی است که برای مقایسه پراکندگی داده‌های کمّی مورد استفاده قرار می‌گیرد. در این سوال سن و قد که مورد مقایسه قرار می‌گیرند هر دو متغیر کمّی هستند

۱۰- گزینه ۱ صحیح است.

برای محاسبه مد یا نما در این سوال از معادله پیرسون استفاده می‌شود:

$$\bar{X} = 14$$
$$Md = 12$$
$$Mo = ?$$

$$Md = \frac{Mo + 2\bar{x}}{3}$$
$$12 = \frac{Mo + 2(14)}{3} \Rightarrow 36 = Mo + 28 \Rightarrow Mo = 36 - 28 = 8$$

۱۱- گزینه ۳ صحیح است.

انحراف معیار، یکی از شاخص‌های پراکندگی است و وقتی صفر باشد، یعنی اینکه هیچ پراکندگی دیده نمی‌شود.

۱۲- گزینه ۳ صحیح است.

میانه مقداری است که تمام فراوانی‌ها را به دو گروه به اندازه مساوی تقسیم می‌کند. به بیان دیگر، مقداری است که حجم آماری یا نمونه آماری را به دو قسمت مساوی تقسیم می‌کند. بطوری که برای نصف افراد نمونه آماری، مقدار متغیر کوچکتر از میانه و برای نیم دیگر بزرگتر از میانه خواهد بود.

برای محاسبهٔ میانه در یک مجموعه اعداد، ابتدا داده‌ها از کوچک به بزرگ مرتب کرده، سپس از طریق فرمول $\frac{N+1}{2}$ ردیف میانه را پیدا می‌کنند. اگر تعداد اندازه‌ها **(N)** فرد باشد، میانه عددی خواهد بود که به فرد وسط تعلق دارد.

$$7-8-9-11-\boxed{12}-13-14-15-17$$

$$\text{ردیف میانه} = \frac{N+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5 \Rightarrow \text{میانه} = 12$$

۱۳- گزینه ۲ صحیح است.

متغیرهایی که سطح سنجش آنها اسمی است با استفاده از شاخص‌های آماری مد یا نما یافته‌های مربوط به آنها توصیف می‌گردد.

۱۴- گزینه ۲ صحیح است.

از ویژگی‌های عمده میانگین این است که این شاخص نسبت به تک تک اعداد توزیع فراوانی حسّاس است. امّا این مزیت در برخی از شرایط یکی از نقاط ضعف میانگین محسوب می‌شود. به عنوان مثال، کلاسی را در نظر بگیرید که یک یا دو نمره خیلی بالا در آن وجود دارد وهمین یک یا دو عدد، بر ارزش عددی میانگین تأثیر می‌گذارد. چنانچه پژوهشگری علاقه‌مند باشد تا نمره‌های خیلی بزرگ یا کوچک بر شاخص مرکزی تأثیری داشته باشند، میانگین شاخص مناسبی است در غیر این صورت توصیه نمی‌شود، چرا که میانه و نما مناسب‌تر هستند، چون تحت تأثیر اعداد خیلی کوچک یا خیلی بزرگ قرار نمی‌گیرند.

۱۵- گزینه ۲ صحیح است.

اگر نمره‌های انحراف از میانگین $(x-\overline{x})$ را بر انحراف استاندارد مجموعهٔ نمره‌ها **(S)** تقسیم کنید. نمره‌های استاندارد بدست می‌آیند.

فرمول آن عبارتست از:

$$Z_x = \frac{x_i - \overline{x}}{S_x}$$

در این فرمول Z_x نمره استاندارد، x_i نمره خام، $\overline{x}$ میانگین نمره‌های X و S_x انحراف استاندارد نمره‌های **x** است.

$$p(z < \frac{85-100}{15}) = p(z < -1) = 1 - 0/68 = 0/32$$

۱۶- گزینه ۳ صحیح است.

نما یک اندازه گرایش مرکزی است و صرفاً برای سطح سنجش اسمی که پایین‌ترین مقیاس اندازه‌گیری می‌باشد، بکار می‌رود. میانه یک شاخص ترتیبی است. در حالیکه از میانه هم در سطح مقیاس ترتیبی و هم در سطح فاصله‌ای می‌توان استفاده کرد. اما نمی‌توان از میانه در سطح اسمی استفاده کرد. میانگین صرفاً برای مقیاس‌های فاصله‌ای و نسبی بکار می‌رود؛ اما متغیرهای فاصله‌ای و نسبی برای هر سطحی از اندازه‌گیری قابل بکار گیری است، چون قابل تبدیل به مقیاس رتبه‌ای و اسمی است ولی مقیاس رتبه‌ای و اسمی قابل تبدیل به مقیاس فاصله‌ای و نسبی نمی‌باشد. بنابراین میانگین برای هر سطحی از اندازه‌گیری قابل بکارگیری است. میانگین بهترین شاخص گرایش مرکزی است، چون برای محاسبه آن از تمامی مقادیر توزیع استفاده می‌شود. در مقابل مد ضعیف‌ترین است چون در محاسبه آن تنها یک نمره نقش دارد آن هم نمره‌ای که بیشترین فراوانی را دارد

۱۷- گزینه ۳ صحیح است

طبقات	فراوانی(Fi)	حدوسط طبقات(Xi)	(Fi Xi)
0-2	5	$\frac{0+2}{2}=1$	5
3-5	10	$\frac{3+5}{2}=4$	40
6-8	15	$\frac{6+8}{2}=7$	105
9-11	20	$\frac{9+11}{2}=10$	200
12-14	24	$\frac{12+14}{2}=13$	312
15-17	26	$\frac{15+17}{2}=16$	416

$$\bar{x} = \frac{\sum FiXi}{N} = \frac{1078}{100} = 10/78$$

$$Mo = L + I(\frac{d_1}{d_1 + d_2})$$

$$Mo = 14/5 + (\frac{2}{2+0}) \times 3 = 17/5$$

جمع	100		1078

با توجه به مقدار میانگین(10/78) و مد(17/5)؛ چون مقدار مد از مقدار میانگین بیشتر است، بنابراین شکل توزیع میزان خودکشی، چوله به چپ(منفی) است. در این حالت مد یا نما از میانه و میانه از میانگین بزرگتر می‌باشد. $Mo > Md > \bar{x}$

۱۸- گزینه ۲ صحیح است.

همانطور که در سوال قبل ذکر شد، در چولگی منفی، مد یا نما از میانه و میانه از میانگین بزرگتر می‌باشد.

$Mo > Md > \bar{x}$

۱۹- گزینه ۲ صحیح است.

اگر نمره‌های انحراف از میانگین ($x-\bar{x}$) را بر انحراف استاندارد مجموعهٔ نمره‌ها **(S)** تقسیم کنید. نمره‌های استاندارد بدست می‌آیند.

فرمول آن عبارتست از:

$$Z_x = \frac{x_i - \bar{x}}{S_x}$$

در این فرمول Z_x نمره استاندارد، x_i نمره خام، $\bar{x}$ میانگین نمره‌های X و S_x انحراف استاندارد نمره‌های **x** است.

$$Z_x = \frac{x_i - \bar{x}}{S_x}$$

$$-2 = \frac{xi - 16}{1/25} \Rightarrow (-2 \times 1/25) = xi - 16 \Rightarrow xi = 16 - 2/5 = 13/5$$

۲۰- گزینه ۳ صحیح است.

زمانی که میانگین از میانه و نما یا مد بزرگتر باشد، چولگی مثبت است($\bar{x} > Md > Mo$). در این سوال هم میزان میانگین بزرگتر از میزان میانه است، پس توزیع دارای چولگی مثبت است($75/70 > 72/23$).

۲۱- گزینه ۱ صحیح است.

میانگین نمره‌های استاندارد معادل صفر و انحراف استاندارد آنها معادل یک است. در اینجا نمره استاندارد فرد(0/8-)است که از یک کمتر است.(ساعی:90)

۲۲- گزینه۲ صحیح است.

$0/7\times0/3=0/21$

۲۳- گزینه ۱ صحیح است.

$$2)A=\frac{\mu_3}{S^3}=\frac{M_3}{S^3}=\frac{\frac{\sum F(x_i-\overline{x})^3}{N}}{S^3}=\frac{\frac{\sum (x_i-\overline{x})^3}{N}}{S^3}$$

در این سوال چون مجموع انحراف نمره‌ها از میانگین برابر با صفر است(0=2+1+0+2−1−)صورت کسر صفر می-شود. بنابراین مقدار چولگی صفر است

۲۴- گزینه ۱ صحیح است.

$\overline{X}=100$
$S=20$
$CRV=?$

$$CRV=\frac{S}{\overline{X}}$$
$$CRV=\frac{20}{100}\times100=20$$
$$CRV=20$$

۲۵- گزینه ۲ صحیح است.

اگر در بین اعداد عادی یک عدد نابهنجاری پدید آید، همان عدد می تواند روی میانگین تأثیر بگذارد، به عنوان مثال، اگر نمره‌های یک کلاس بین 15 تا 20 توزیع شوند و ناگهانی یک نفر صفر بگیرد، تعداد میانگین به سمت نمره صفر کشیده می‌شود. در این موارد بهتر است از میانه استفاده شود. در واقع مناسب‌ترین شاخص مرکزی برای اعداد نوعی و پرت یا داده‌های خیلی بزرگ یا خیلی کوچک میانه می‌باشد، چرا که تحت تأثیر این اعداد قرار نمی‌گیرد

۲۶- گزینه ۱ صحیح است.

در این سوال بهتر است که از شاخص پراکندگی «انحراف چارکی» استفاده شود. با ترتیب کردن نمره‌ها متوجه می‌شویم که اعداد انتهایی اعداد پرت هستند، شاخص‌های پراکندگی دیگر مثل انحراف معیار، واریانس و چارک سوم این اعداد پرت را در نظر می‌گیرند، ولی اگر از انحراف چارکی استفاده شود، فقط 50 درصد وسط توزیع را در نظر می‌گیرد.

$(14-15-17-18-19-25-29-140-180)$

تفاضل چارک سوم بر چارک اول را دامنه تغییر بین چارک‌ها می‌گویند.

دامنه تغییر بین چارکی $= Q_3 - Q_1$

نصف فاصله بین چارک‌های اول و سوم را انحراف چارکی می‌گویند. این شاخص اندازه تغییرپذیری 50 درصد نمره‌های وسط توزیع را فراهم می‌کند.

$$Q = \frac{Q3 - Q1}{2}$$

۲۷- گزینه ۱ صحیح است.

میانگین نمره‌های استاندارد صفر و انحراف استاندارد آنها یک است.(ساعی:87 و دلاور 1387: 143)

۲۸- گزینه۱ صحیح است.

ضریب تغییر نسبی متوسط انحراف از میانه، مانند ضریب تغییر نسبی انحراف از میانگین است با تفاوت زیرکه بجای میانگین از میانه استفاده می‌شود.

$$Md = 80 \qquad CRV_{Ad} = \frac{AD_{Md}}{Md} \times 100$$
$$AD_{Md} = 40 \qquad CRV_{Ad} = \frac{40}{80} \times 100$$
$$CRV_{Ad} = ? \qquad CRV_{Ad} = 0/50$$

۲۹- گزینه۲ صحیح است.

همانطوری که قبلاً ذکر شد، در رابطه با واریانس، اگر در یک توزیع، از تمامی مقادیر صفت، یک مقدار ثابت **a** کسر یا اضافه کنیم، واریانس این مقادیر تغییری نخواهد کرد. در این مثال هم بجای 501 تا 505، 1تا 5 تـا در نظر می‌گیریم.

$$\bar{x} = \frac{\sum xi}{N} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$
$$S^2 = \frac{\sum (xi - \bar{x})^2}{N} = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}$$
$$s^2 = \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$$

۳۰- گزینه ۱ صحیح است.

میانگین به عنوان یک شاخص مرکزی، به ارزش عددی هر یک از مشاهده‌ها در توزیع فراوانی بسـتگی دارد. در صورتی که ارزش عدد تغییر کند(افزایش یا کاهش)یابد، مقدار میانگین نیز در همان جهت تغییر خواهـد کـرد.

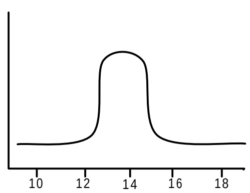
پس اگر به مقادیر یک متغیر عدد ثابتی مانند **a** را اضافه کنیم یا کم کنیم یا تقسیم کنیم یـا ضـرب، میـانگین اعداد حاصله نیز به همان اندازه تغییر خواهد کرد، یعنی میانگین اعداد حاصله برابر است با میـانگین قبلـی بـه اضافه **a** یا منهای **a** یا تقسیم بر **a** یا ضرب در **a**. پس در این مثال هـم بـرای بدسـت آوردن میـانگین جدیـد، عدد 5 را که به همه مقادیر اضافه شده است، به میانگین قبلی اضافه می‌کنیم. $\bar{x} = 45 + 5 = 50$

۳۱- گزینه ۴ صحیح است.

انحراف چارکی مانند میانه تحت تأثیر نمره‌های خیلی بزرگ یا خیلی کوچک (اعداد پرت) قرار نمی‌گیرد

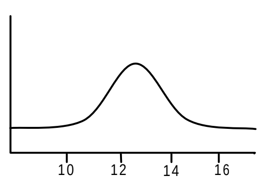
۳۲- گزینه ۴ صحیح است.

کشیدگی زمانی مثبت است که **K** بزرگتر از صفر ($K>0$) باشد. در اینجا منحنـی توزیـع نمـره‌هـا، از منحنـی نرمال، کشیده‌تر است. برآوردگی منحنی زمانی در نقطه اوج قرار مـی‌گیـرد کـه **K** آن بـیش از 0/5 ($K>0/5$) باشد. در این صورت کشیدگی خیلی دورتر از نرمال است:



کشیدگی زمانی نرمال است که **K** معادل صفر ($k=0$) باشد.

حتی اگر $K \le 1$ باشد، در این صورت نیز توزیع جامعه به نرمال نزدیکتر است. در کشیدگی طبیعی نمره‌ها، تقریباً به طور یکسان توزیع می‌شوند.



۳۳- گزینه ۳ صحیح است.

برای محاسبه مد یا نما در این سوال از معادله پیرسون استفاده می‌شود:

$\bar{X} = 20$
$Md = 22$
$Mo = ?$

$$Md = \frac{Mo + 2\bar{x}}{3}$$

$$22 = \frac{Mo + 2(20)}{3} \Rightarrow 66 = Mo + 40 \Rightarrow Mo = 66 - 40 = 26$$

۳۴- گزینه ۴ صحیح است.

فرمول نمره استاندارد: $Z_x = \frac{x_i - \bar{x}}{S_x}$

در این فرمول Z_x نمره استاندارد، x_i نمره خام، $\bar{x}$ میانگین نمره‌های X و S_x انحراف استاندارد نمره‌های $\mathbf{x}$ است.

$\bar{x} = 50$
$s = 50$
$xi = ?$
$z = -2$

$$(-2 = \frac{x_i - 50}{50} \Rightarrow -100 = x_i - 50$$
$$x_i = -100 + 50 = -50$$
$$x_i = -50$$

۳۵- گزینه ۴ صحیح است.

مجموع انحراف نمره‌ها یا مقادیر از میانگین برابر صفر است و فرمول آن عبارت است از: $\sum(x_i - \bar{x}) = 0$

مثال: اگر این اعداد را داشته باشیم، 11-17-15-13 طبق فرمول بالا هنگامی که میانگین را از نمره‌ها کم می‌کنیم، حاصل آن انحراف از میانگین نامیده می‌شود. نمره‌هایی که بزرگتر از میانگین هستند، دارای انحراف مثبت و نمره‌هایی که کوچکتر از میانگین هستند، انحراف منفی دارند. مجموع انحرافات منفی برابر مجموع انحرافات مثبت است. به همین دلیل میانگین، نقطهٔ تعادل توزیع نمره‌هاست

۳۶- گزینه ۲ صحیح است.

فرمول نمره استاندارد: $Z_x = \frac{x_i - \bar{x}}{S_x}$

در این فرمول Z_x نمره استاندارد، x_i نمره خام، $\bar{x}$ میانگین نمره‌های X و S_x انحراف استاندارد نمره‌های $\mathbf{x}$ است.

$\bar{x} = 100$
$s = 5$
$xi = 110$
$z = ?$

$$(z = \frac{110 - 100}{5} \Rightarrow \frac{10}{5} = 2)$$
$$z = 2$$

۳۷- گزینه ۳صحیح است.

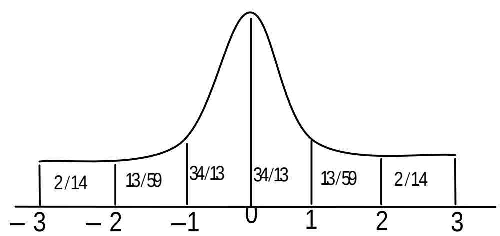
$$Z_x = \frac{x_i - \overline{x}}{S_x}$$

فرمول نمره استاندارد:

در این فرمول Z_x نمره استاندارد، x_i نمره خام، $\overline{x}$ میانگین نمره‌های X و S_x انحراف استاندارد نمره‌های $\mathbf{x}$ است.

$$\bar{x} = 72$$
$$s^2 = 16 \Rightarrow s = \sqrt{16} = 4$$
$$xi = 68, 76$$
$$z = ?$$

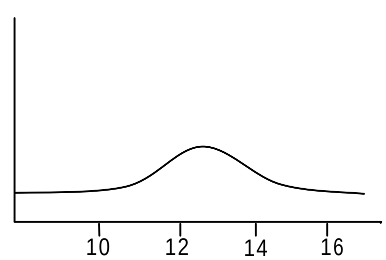
$$\left[(z = \frac{68-72}{4}) p ? (z = \frac{76-72}{4}\right]$$
$$(z = -1) > p < (z = 1)$$
$$p = 68\%$$



۳۸- گزینه ۳ صحیح است.

کشیدگی زمانی منفی است که **K** کوچکتر از صفر **(K<0)** باشد. که در این مثال هم این میزان منفی است.(-0/8)

در اینجا منحنی نمره‌ها از منحنی نرمال پایین‌تر خواهد بود. هم‌چنین منحنی به صورت پخی درمی‌آید.



در کشیدگی منفی دامنه تغییرات زیاد اطراف میانگین است. مثلاً اگر دانشجویان نمره بین 10 تا 17 را کسب نمایند منحنی به صورت پخی در می‌آید. به کشیدگی منفی منحنی پخی نیز می‌گویند.

۳۹- گزینه ۱ صحیح است.

برای محاسبه مد یا نما در این سوال از معادله پیرسون استفاده می‌شود:

$$\bar{X} = ?$$
$$Md = 42$$
$$Mo = 50$$

$$Md = \frac{Mo + 2\bar{x}}{3}$$

$$42 = \frac{50 + 2(\bar{x})}{3} \Rightarrow 126 = 50 + 2(\bar{x}) \Rightarrow 2(\bar{x}) = 126 - 50 = 76 \Rightarrow 2(\bar{x}) = 76$$

$$\bar{x} = \frac{76}{2} = 38$$

چون مد یا نما بزرگتر از میانگین است شکل توزیع چوله به چپ است.($Mo > Md > \bar{x}$)

۴۰- گزینه ۳صحیح است.

درمقیاس نسبی تمامی عملیات ریاضی قابل اجراست. و دارای نقطه صفر واقعی است.

۴۱- گزینه ۴ صحیح است.

انحراف چارکی مانند میانه تحت تأثیر نمره‌های خیلی بزرگ یا خیلی کوچک (اعداد پرت) قرار نمی‌گیرد.

(دلاور،1387: 123)

۴۲- گزینه ۱ صحیح است.

اگر توزیع نمرات کمّی باشد، میانگین بهترین شاخص مرکزی است.

فصل دوم: منحنی طبیعی (بهنجار)

در پرتاب **n** سکه، توزیع فراوانی شیرها یا خط ها هر چقدر که **n** افزایش یابد، به توزیع بهنجار شباهت بیشتری پیدا می کند. منحنی بهنجار شکل محدودی از دو جمله ای متقارن است. معادله منحنی بهنجار عبارت است از :

$$\mathbf{Y} = \frac{N}{d\sqrt{2\Pi}}\,\mathbf{e}^{-(\mathbf{x}-\mathbf{m})2}\,/d^{\,2}$$

اجزای این فرمول

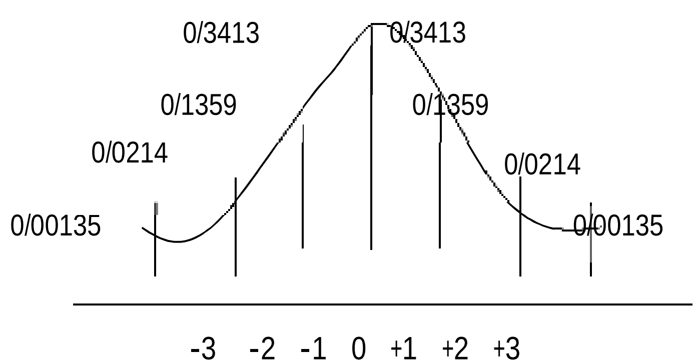
Y ارتفاع منحنی برای ارزش های **x** ، p عدد ثابت 3/1416 ، **e** پایه لگاریتم پنارین 2/7183، **N** تعداد موارد ، m میانگین و d انحراف معیار.

شکل توزیع بهنجار به میانگین و انحراف استاندارد بستگی دارد.

سطوح زیر منحنی بهنجار

به دلیل دستیابی به اهداف پژوهشی، محاسبه سطح زیر منحنی بهنجار در فاصله بین محورهای عمودی نقاط مختلف بر روی محور افقی از اهمیت زیادی برخوردار است. در منحنی بهنجار محور **x** بر اساس نمره های **z** بـه 1± و 2±و 3± انحراف معیار تقسیم می گردد. نسبت سطح بین **z** =0 و **z** =1 برابر 0/3413 است. از آنجایی که این منحنی متقارن است، نسبت سطح بین **z**=0 و **z**=-1 نیز برابر 0/3413 خواهد بود. بنابراین نسبت سطحی که بین حدود **z**=±1 قرار می گیرد عبارت است از: 0/6826= 0/3413 + 0/3413 یا تقریباً 68 درصد. نسبت سطحی که بین **z** =±2 قرار مـی گیـرد، 0/9544= 0/4772 + 0/4772 یا

تقریباً 95 درصد است. نسبت بین **z**=±3 برابر است با 0/99730=0/49865+ 0/49865 یا 99/73 درصد اسـت. در زیـر شکل منحنی بهنجار و سطوح زیر آن نشان داده شده است.



برای پیدا کردن سطح بین هر دو نقطه روی خط پایه، مثال زیر را در نظر می گیریم:

فرض کنید قصد داریم که سطح بین z =0/5 و z =1/5 را تعیین کنیم. بر اساس جدول نمره های معیار، نسبت سطح بین میانگین و z =0/50 برابر با 0/1915 است. همچنین سطح بین میانگین و z =1/5 برابر است با 0/4332 سطح بین z=0/5 و z=1/5 با تفریق این دو سطح از یکدیگر حاصل می شود. 0/4332 -0/1915 =0/2417

*مثال : توزیع نمره های هوشی حاصل از اجرای یک آزمون هوش با میانگین 100 و انحراف معیار 15 تقریباً بهنجار است . فرض کنید که از ما خواسته شده است تعیین کنیم چند درصد از افراد جامعه دارای هوش بهری مساوی یا بزرگتر از 120 هستند. بدین ترتیب داریم:

$$Z=\frac{120-100}{15}=1/33$$

بنابراین ، بهره هوشی 120 ، 1/33 واحد انحراف معیار بالاتر از میانگین است . مراجعه به جدول نمره معیار نشان می دهد که نسبت این نمره معیار برابر با 0/092 (یا 9/2٪) است.

ویژگی منحنی بهنجار

1- منحنی بهنجار متقارن است و میانگین، میانه و نما در آن بر هم منطبق هستند.

2- حداکثر عرض این منحنی در نقطه میانگین قرار دارد ، یعنی نقطه ای که در آن z=0 است. عرض این نقطه در منحنی بهنجار واحد برابر 0/3989 است.

3- این منحنی مجانب است، یعنی دو دنباله منحنی به محور افقی نزدیک می شوند ولی هرگز آن را قطع نمی کنند.

4- نقاط عطف این منحنی در نقاط مثبت یا منهای یک واحد انحراف معیار بالا یا پایین میانگین قرار دارند.

5- تقریباً 68% سطح این منحنی بین حدود مثبت و منفی یک واحد انحراف معیار نسبت به میانگین قرار دارد.

6- در منحنی بهنجار واحد حدود z = ± 1/96 ، 95% و حدود z =2/58 ، 99% کل سطح منحنی را در بر می گیرد.

کاربرد عملی منحنی بهنجار

1- محاسبه رتبه درصدی نمره ها در یک توزیع بهنجار .

2- بهنجار کردن یک توزیع فراوانی که روش مهم در استاندارد کردن یک آزمون روانی یا یک سیاهه رفتار است .

3- آزمودن معنی دار بودن اندازه های مشاهده شده در یک آزمایش و مربوط ساختن آنها به نوسان های تصادفی یا خطاهایی که همواره در فرایند نمونه برداری وجود دارد و تعمیم آن به جامعه ای که نمونه از آن جامعه انتخاب شده است .

تقریب توزیع بهنجار به دو جمله ای

مشاهده های به عمل آمده نشان می دهد که با افزایش(n)، توزیع دو جمله ای متقارن به توزیع بهنجار نزدیک می شود . این موضوع به معنای آن است که میتوان توزیع بهنجار را بـرای بـرآورد احتمـال هـای دو جملـه ای بکـار بـرد. هرگـاه موقعیتی وجود داشته باشد که در آن 10 سکه به دفعات زیاد پرتاب شده باشد، احتمال بدست آوردن 7 شـیر یـا بیشـتر چقدر است؟ در اینجا میانگین توزیع دو جمله ای $m = 10\left(\frac{1}{2}\right)$ و انحـراف معیـار آن $d = \sqrt{\frac{10}{4}} = 1/58$ اسـت . چـون توزیع بهنجار یک توزیع پیوسته است و نه گسسته، میتوان ارزش 7 را در دامنه حدود واقعی 6/5 تـا 7/5 درنظـر گرفـت. بنابراین باید نسبت سطح بالای محور عرض هـا را در نقطـه 6/5 باتوجـه بـه میـانگین 5 و انحـراف معیـار 1/58 بـرآورد کرد.نمره معیار ارزش 6/5 برابر است با : $Z = \frac{6/5 - 5}{1/58} = 0/949$. نسبت سطح منحنی بهنجار در بالای محور عرض هـا در نقطه $z = 0/949$ را به آسانی میتوان از جدول استخراج کرد. این مقدار برابر با 0/171 است .

تست های فصل دوم

۱- معلمی تصمیم گرفته است که ۲۵٪ از دانشجویان کلاس خود را مردود کند. چنانچه توزیع نمره های این کلاس تقریباً بهنجار و دارای میانگین ۷۲ و انحراف معیار ۶ باشد. نمره قبولی چقدر است؟

الف) 70 ب) 68 ج) 72 د) 66

۲- مساحت بین $1d$ و $2d$ در توزیع نرمال چگونه است؟

الف) برابر با مساحت بین $2d$ و $3d$ ب) بیشتر از مساحت بین $2d$ و $3d$

ج) کمتر از مساحت بین $2d$ و $3d$ د) نصف مساحت بین $2d$ و $3d$

۳- نمره z=+۱ در صورتی که توزیع نرمال باشد، از حدود چند درصد نمرات توزیع بزرگتر است؟

الف)34 ب)50 ج) 68 د) 84

۴- چه نسبتی از سطح زیر منحنی بین ۱z± تا ۲z ± قرار دارد؟

الف) 0/1718 ب)0/3718 ج) 0/2718 د) 0/0718

۵- در یک توزیع نرمال، میانگین برابر ۱۰ و انحراف معیار ۲ است، میانه کدام است؟

الف) 4 ب) 5 ج)8 د) 10

۶- در یک توزیع نرمال، اگر میانگین ۱۲و انحراف معیار ۳ باشد، ۶۸٪ افراد نمره هایشان در چه فاصله ای قرار دارد ؟

الف) 8-6 ب) 15-9 ج) 13-10 د) 15- 10

۷- در آزمونی که میانگین آن ۱۰ و انحراف معیار آن ۳۰ باشد، طبق منحنی بهنجار ، ۵۰ درصد افراد از چه نمره هایی پایین ترند؟

الف) 16 ب) 13 ج) 7 د) 10

۸- در یک توزیع نرمال سطح زیر منحنی در فاصله ای یک واحد از میانگین (m±۱) برابر است با:

الف) 0/68 ب)0/34

ج) بستگی به انحراف معیار دارد. د) بستگی به میانگین دارد.

۹- اگر متغیر تصادفی x دارای توزیع نرمال با میانگین ۲۵ باشد، $P(x<25)$ کدام است؟

الف) صفر   ب) 1   ج) 0/5   د) 0/75

۱۰- در منحنی طبیعی نمرات هوش ($\bar{x}=100$ و $s=15$) نمره چند درصد افراد بین $x_1=85$ و $x_2=115$ قرار دارد.

می دانیم فاصله یک انحراف معیار با میانگین برابر ۳۴/۱۳ درصد زیر منحنی است؟

الف) 48   ب) 68   ج) 95   د) 99

پاسخنامه تست های فصل دوم

1- گزینه ب

$0/25 \neq 0/2486 \longrightarrow \mathbf{z}=0/67$ (جدول)

$$z = \frac{x_i - \bar{X}}{s} \rightarrow 0/67 = \frac{-x_i - 72}{6} \Rightarrow x_i = 68$$

2- گزینه ب

- نکته: بین $1d$ و $2d$ حدود 13٪ نمرات قرار دارند و بین $2d$ و $3d$ حدود 2٪ نمرات قرار دارند. پس مساحت بـین $1d$ و $2d$ بیشتر از مساحت بین $2d$ و $3d$ است.

3- گزینه د

- نکته: در توزیع نرمال 50٪ زیر میانگین یعنی نقطه صفر قرار دارد و بین میانگین و یـک هـم 34٪ قـرار دارنـد . پـس 84٪ زیر نمره $\mathbf{z}=+1$ قرار می گیرد.

4- گزینه ج

5- گزینه د

6- گزینه ب

7- گزینه د

8- گزینه الف

9- گزینه ج

- نکته: در توزیع نرمال 50٪ نمرات از میانگین پایین ترند. حال چون میانگین برابر با 25 است. احتمـال آنکـه نمـره $\mathbf{x}$ کوچکتر از میانگین یعنی عدد 25 باشد، برابر با 0/5 است.

10- گزینه ب

فصل سوم: متغیرهای تصادفی و توزیعهای احتمال

ما اغلب می‌خواهیم از طریق مشاهدهٔ حوادث غیرقابل کنترل در طبیعت، یا به وسیلهٔ آزمایشهایی که تا حدی در کنترل ما هستند اطلاعاتی به دست آوریم **آزمایش تصادفی** اصطلاحا برای عملی که منجربه جمع‌آوری اطلاعات می‌شود به کار می‌رود. تعریف دقیقتر زیر را برای آزمایش بیان می‌کنیم.

3-1 **آزمایش تصادفی**

آزمایش تصادفی حادثه یا عملی است که نتایج آن قابل مشاهده هستند ولی نتیجهٔ آن را از قبل نمی‌دانیم.

مثالهایی از آزمایش تصادفی عبارتند از:

1- پرتاب یک سکه، نتایجی که ممکن است در این آزمایش مشاهده شوند آمدن شیر یا خط است.

2- پرتاب کردن یک تاس که نتایج ممکن آن عبارتند از یکی از اعداد 1 تا 6.

3- تعیین درآمد کارگری از یک کارخانهٔ معین که امکان دارد بین 80000 تا 180000 ریال باشد.

4- تعیین میزان بارندگی در یک روز ابری که ممکن است بزرگتر یا مساوی صفر میلیمتر باشد.

3-2 **فضای نمونه‌ای**

به مجموعهٔ تمام برآمدهای ممکن برای یک آزمایش تصادفی **فضای نمونه‌ای** آن آزمایش گوییم.فضای نمونه‌ای را با $\mathbf{S}$ و نقاط فضای نمونه‌ای را با $\boldsymbol{e}_{\mathbf{i}}$ نشان می‌دهیم.

3-3 **پیشامد**

یک پیشامد زیر مجموعه‌ای از فضای نمونه است. معمولا پیشامدها را با حروف بزرگ نشان می‌دهند. پیشامدها را با استفاده از نماد مجموعه یا از طریق توصیف آنها با عبارتهای انشایی مشخص می‌کنند.

مثال–یک تاس را پرتاب کنید و عددی که روی آن است مشاهده کنید. آزمایش عبارت است از پرتاب یک تاس. قبل از انداختن تاس از نتیجهٔ آن اطلاعی نداریم ولی می‌دانیم که به هر حال یکی از اعداد 1 تا 6 روی تاس ظاهر خواهد شد. پس فضای نمونه‌ای آزمایش عبارت است از مجموعهٔ

$$S = \{1, 2, 3, 4, 5, 6\}$$

فضای نمونه‌ای 6 نقطه دارد که آنها را با علامتهای $\boldsymbol{e}_{\mathbf{i}}$ نشان می‌دهیم.

$$\boldsymbol{e_1 = 1 \ , e_2 = 2 \ , e_3 = 3 \ , e_4 = 4 \ , e_5 = 5 \ , e_6 = 6}$$

هر زیر مجموعهٔ S یک پیشامد است. بعضی از پیشامدها عبارتند از:

پیشامد A: مشاهدهٔ یک عدد فرد، $A = \{1, 3, 5\}$

پیشامد B: مشاهدهٔ یک عدد کمتر از 4، $B = \{1, 2, 3\}$

پیشامد E_1 : *مشاهدهٔ* عدد1، $E_1 = \{1\}$

پیشامدE_2 : *مشاهدهٔ* عدد2، $E_2 = \{2\}$

نتایج آزمایشها می‌توانند کمی یا کیفی باشند اگر نتیج آزمایش کیفی باشد می‌توان به هر نتیجه یک عدد نسبت داد و نتایج را به صورت کمی درآورد. مثلا اگر آزمایش انداختن یک سکه باشد، مجموعه برآمدهای آزمایش عبارت است از

$S = \{\text{خط، شیر}\}$

مشاهده می‌شود که این نتایج کیفی هستند ولی با نسبت دادن عدد «1» به شیر و عدد «0»به خط می‌توان نتایج را به صورت کمی تبدیل کرد. در حالت کلی می‌توانیم به هر عضو فضای نمونه‌ای یک مقدار عددی نسبت دهیم. اگر این مقدار را با $\mathbf{X}$ نشان دهیم، وقتی آزمایش انجام و برآمد آن معلوم شود مقدار $\mathbf{X}$ دقیقا مشخص می‌شود ولی قبل از انجام آزمایش نمی‌توان گفت $\mathbf{X}$، دقیقا چه مقداری را خواهد پذیرفت. به همین دلیل $\mathbf{X}$ یک **متغیر تصادفی** نامیده می‌شود.

3-4 **متغیر تصادفی**

متغیرتصادفی، متغیری است که مقدار عددی‌اش به وسیلهٔ برآمد یک آزمایش تصادفی تعیین می‌شود. متغیر تصادفی را با حروف بزرگ $\mathbf{X}$ ، $\mathbf{Y}$ ، $\mathbf{Z}$ و مقادیر خاصی را که این متغیر می‌تواند بپذیرد با حروف کوچک $\mathbf{x}$، $\mathbf{y}$، $\mathbf{z}$ نشان می‌دهند.

3-5 **طبقه‌بندی متغیرهای تصادفی**

متغیرهای تصادفی به یک از دو گروه **گسسته** یا **پیوسته** تعلق دارند.

الف) متغیر تصادفی گسسته: متغیر تصادفی که فقط مقادیر مجزا را اختیار می‌کند را متغیر تصادفی گسسته می نامند. تعداد شیرهایی که در پرتاب سه سکه ظاهر می‌شوند یک متغیر تصادفی گسسته است، که می‌تواند یکی از مقادیر 1، 2 یا 3 را اختیار کند، تعداد این مقادیر متناهی است. تعداد سالهایی که طول می‌کشد تا سرمایهٔ یک شرکت به یک میلیارد تومان برسد یک متغیر تصادفی گسسته است که می‌تواند یک از مقادیر 1، 2 ، 3، ... را اختیار کند. تعداد این مقادیر ممکن است نا متناهی باشد یعنی هیچ وقت سرمایهٔ شرکت به یک میلیارد تومان نرسد.

ب) متغیر تصادفی پیوسته: متغیر تصادفی را که ممکن است مقادیر موجود در یک بازه را اختیار کند، متغیر تصادفی پیوسته گویند. مقادیر متغیرهای تصادفی پیوسته حاصل اندازه‌گیری کمیتهای پیوسته‌ای مانند زمان، مساحت، حجم یا طول هستند. مثالهایی از متغیرهای تصادفی پیوسته عبارتند از: 1- مدت زمان لازم برای انجام کاری معین در یک کارخانه، 2- مقدار نفتی که بوسیلهٔ یک پمپ از چاهی در یک ساعت خارج می‌شود، 3- مقدار اکسید کربن موجود در یک متر مکعب از هوا.

تفاوت بین متغیرهای تصادفی پیوسته و گسسته از این نظر مهم است که برای هر کدام از آنها توزیعهای احتمال با شرایط متفاوتی مورد نیاز است که موضوعقسمتهایی از فصول بعدی می‌باشند.

3-6 تابع احتمال

به تابعی که بتوان با استفاده از آن هر یک از مقادیر ممکن متغیر تصادفی را مشخص کرد « تابع احتمال» یا « توزیع احتمال» گفته می‌شود.تابع احتمال را می‌توان چنین تعریف کرد: « تابع احتمال، تابعی است که دامنهٔ آن مقادیر ممکن متغیر تصادفی وحوزهٔ آن احتمالات مربوط به هر مقدار متغیر تصادفی است». تابع احتمال در حالت کلی دارای شرایط زیر می‌باشد:

اولشرط $\sum_{i=1}^{n} f_i = 1$

شرط دوم $0 \leq f_i \leq 1$

3-7 تابع توزیع (تابع احتمال تجمعی)

تعریف این تابع چنین است: « تابع توزیع، تابعی است که به ازاء جمیع مقادیر ممکن متغیر تصادفی **X**، احتمال وقوع مقداری کوچکتر یا مساوی با **x** را نشان می‌دهد». تابع توزیع را با $\mathbf{P\ (X \leq x)}$ یا $\mathbf{F(x)}$ نشان می‌دهند. بدیهی است که تابع توزیع بیشترین مقداری که متغیر تصادفی می‌تواند اختیار کند برابر 1 است.

3-8 امید ریاضی متغیر تصادفی

امید ریاضی مفهومی اساسی در مطالعل توابع احتمال است. امید ریاضی همان میانگین موزون است که احتمالات در آن، نقش وزنها (ضرایب) را بازی می‌کنند. امید ریاضی متغیر تصادفی **X** را با **E(X)** نشان می‌دهیم. امید ریاضی متغیر تصادفی گسستهٔ **X** به این صورت محاسبه می‌شود (($\mathbf{f(x)}$ همان $\mathbf{P\ (X=x)}$ است).

$$E(X) = \sum_{x} x.f(x)$$

با توجه به امید ریاضی متغیر تصادفی، اگر **a** و **b** مقادیر ثابتی فرض شوند، دارای این خواص خواهند بود:

$$E(b) = b$$
$$E(aX) = aE(X)$$
$$E(aX + b) = aE(X) + b$$

9-3 **واریانس متغیر تصادفی**

واریانس متغیر تصادفی گسستهٔ $\mathbf{X}$ که میزان پراکندگی حول میانگین (امید ریاضی) نشان می‌دهد و آن را با $\mathbf{V\ (X)}$ نشان می‌دهیم، به اینصورت محاسبه می‌شود:

$$V(X) = \sum_x (x - E(X))^2 f(x)$$

در نتیجه واریانس را می‌توان به صورت زیر محاسبه نمود:

$$V(X) = E(X^2) - E(X)^2$$

در هر صورت، انحراف معیار به اینصورت محاسبه می‌شود:

$$SD(X) = \sqrt{V\ (X)}$$

اگر **a** و **b** مقادیر ثابتی فرض شوند، این خواص برای واریانس متغیر تصادفی قابل اثبات است:

$$V(b) = 0$$
$$V(aX) = a^2V(X)$$
$$V(aX + b) = a^2V(X)$$

10-3 **تابع احتمال توأم**

دو متغیر تصادفی $\boldsymbol{X}$ و $\boldsymbol{Y}$ را در نظر بگیرید؛ بطوریکه $\boldsymbol{X}$ بتواند مقادیر $\boldsymbol{x_i}$ به ازای $\boldsymbol{i}$ مساوی 1، 2، ...، $\boldsymbol{N}$، و $\boldsymbol{Y}$ مقادیر $\boldsymbol{y_j}$ به ازای $\boldsymbol{j}$مساوی 1، 2، ...، $\boldsymbol{k}$ را اختیارکند؛ در این صورت برای زوجهای$\boldsymbol{(x_i,y_j)}$تابع احتمال توأم$\boldsymbol{X}$ و $\boldsymbol{Y}$ وجود دارد که با$\boldsymbol{f(x_i\ y_j)}$نشان داده می‌شود؛ بنابراین، می‌توان گفت تابع احتمال توأم عبارت است از فهرستی از زوجهای $\boldsymbol{(x_i\ y_j)}$ و احتمالهای متناظر با آنها، یعنی$\boldsymbol{f(x_i\ y_j)}$. تابع احتمال دو متغیر تصادفی $\boldsymbol{X}$ و $\boldsymbol{Y}$ را در جدولی مشابه این جدول نشان می‌دهیم:

	y_1	y_2	$\cdots$	y_k
x_1	$y_1)f(x_1$	$y_1)f(x_1$	$\cdots$	$y_1)f(x_1$
x_2	$y_1)f(x_2$	$y_2)f(x_2$	$\cdots$	$y_K)f(x_2$
.	.	.	$\cdots$	.
.	.	.	$\cdots$	.
.	.	.	$\cdots$	.
x_N	$y_1)f(x_N$	$y_2)f(x_N$	$\cdots$	$y_K)f(x_N$

با در د ست داشتن تابع احتمال توأم، می توان احتمال جداگانه هر یک از متغیرهای تصادفی را پیدا کرد. با توجه به جدول فوق، برای پیدا کردن تابع احتمال متغیر تصادفی X، می‌توان احتمالات هر سطر را جمع کرد و آن را در حاشیهٔ سمت راست جدول نوشت و یا برای پیدا کردن تابع احتمال متغیر تصادفی Y، می‌توان احتمالات هر ستون را جمع کرد و آن را در حاشیهٔ پایین جدول نوشت.

مثال– تابع احتمال توأم زیر را در نظر بگیرید و تابع احتمال متغیرهای تصادفی X و Yرا بطور مجزا نمایش دهید:

احتمالات حاشیه‌ای x	1	0	
0/23	0/18	0/05	0
0/57	0/35	0/22	1
0/20	0/05	0/15	2
1	0/58	0/42	احتمالات حاشیه‌ای y

بنابراین خواهیم داشت:

y	0	1
f(y)	0/42	0/58

x	0	1	2
f(x)	0/23	0/57	0/20

3-11 کوواریانس و استقلال دو متغیر تصادفی

کوواریانس را امید ریاضی تغییرات دو متغیر بر حسب میانگینشان تعریف می‌کنیم. در اینجا، کوواریانس معیار عددی است که نوع و شدت رابطهٔ **خطی** بین دو متغیر تصادفی را نشان می‌دهد. رابطهٔ دو متغیر می‌تواند به یکی از سه حالت باشد: 1. با افزایش یک متغیر، دیگری افزایش یابد و با کاهش آن دیگری کاهش یابد (رابطهٔ مستقیم)، 2. با افزایش یک متغیر، دیگری کاهش یابد و با کاهش آن افزایش یابد (رابطهٔ معکوس)، 3. افزایش و یا کاهش یک متغیر هیچ تاثیری در دیگری نداشته باشد (دو متغیر مستقل باشند). مقدار کوواریانس در حالت اول مثبت، در حالت دوم منفی و در حالت سوم صفر است.

کوواریانس بین دو متغیر تصادفی X و Y را با **Cov (X,Y)** نشان می‌دهیم و به این صورت تعریف می‌کنیم:

$$cov(X,Y) = E(X-\mu_x)(Y-\mu_y)$$

$$cov(X,Y) = \mathrm{E}(\mathrm{X\,Y}) - \mathrm{E}(\mathrm{X})\,\mathrm{E}(\mathrm{Y})$$

مثال– تابع احتمال توأم دو متغیر تصادفی X و Y بصورت زیر است. کوواریانس را حساب کرده، دربارهٔ ارتباط دو متغیر توضیح دهید:

x \ y	0	1	2
-5	0/30	0/02	0/08
10	0/10	0	0/50

برای محاسبهٔ $cov(X,Y)$ ابتدا باید $\mathbf{E(X)}$، $\mathbf{E(Y)}$، $\mathbf{E(X\,Y)}$ را حساب کنیم.

x	-5	10	جمع
f(x)	0/40	0/60	1
x f(x)	-2	6	4

y	0	1	2	جمع
f(y)	0/40	0/02	0/58	1
y f(y)	0	0/02	1/16	1/18

پس $E(X) = 4$ و $E(Y) = 1/18$ است.

$E(XY) = \sum_i \sum_j x_i\, y_j \mathrm{f}(x_i, y_j) =$ 9/1

بنابراین:

$\mathbf{Cov\ (X,\ Y) = E\ (XY) - E(X)\ E(Y) =}$ 9/1 -4 ×18 = 14/38

مقدار مثبت کوواریانس، نشان از وجود رابطهٔ خطی مستقیم بین دو متغیر دارد.

نکته: اگر کوواریانس دو متغیر تصادفی صفر شود، این امر نشان می‌دهد که دو متغیر دارای استقلال خطی می‌باشند، ولی بدین معنی نمی‌باشد که دو متغیر کاملا از هم مستقل باشند زیرا ممکن است دارای وابستگی‌های غیر خطی بـا یکـدیگر باشند. بنابراین:

$\mathbf{Cov\ (X,Y)} = 0 \neq$ مستقلند $\mathbf{X\ ,\ Y}$

همانطورکه می‌دانیم دو پیشامد $\mathbf{A}$ و $\mathbf{B}$ در صورتی مستقلند که رابطهٔ $\mathbf{P(A \cap B) = P(A)P(B)}$ بـین آنهـا برقـرار باشد. بطریق همانند می‌توان گفت دو متغیر تصـادفی $\boldsymbol{X}$ و $\boldsymbol{Y}$، فقـط در صـورتی مسـتقلند کـه بـه ازای تمـام زوجهـای $(\boldsymbol{x_i}, \boldsymbol{y_j})$ این رابطه:

$\mathrm{f}(x_i, y_j) = f(x_i).f(y_j)$

بین آنها برقرار باشد.

در اینجا چند قاعدهٔ مهم دربارهٔ امید ریاضی و واریانس مجموع دو متغیر تصادفی در حالت کلی و در حالتی که دو متغیـر مستقل باشند، آورده می‌شود:

$$E(X+Y) = E(X) + E(Y)$$
$$V(X+Y) = V(X) + V(Y) + 2\,COV(X,Y)$$
$$V(X-Y) = V(X) + V(Y) - 2\,COV(X,Y)$$

$$V(aX + bY) = V(aX) + V(bY) + 2\,COV(aX, bY) = a^2V(X) + b^2V(Y) + 2abCOV(X, Y)$$
$$V(aX - bY) = V(aX) + V(bY) - 2\,COV(aX, bY) = a^2V(X) + b^2V(Y) - 2abCOV(X, Y)$$

حال اگر $\mathbf{X}$ و $\mathbf{Y}$ مستقل باشند، مقدار کوواریانس را برابر صفر قرار می‌دهیم؛ برای مثال:

$$E(X + Y) = E(X) + E(Y)$$
$$V(X + Y) = V(X) + V(Y)$$
$$V(X - Y) = V(X) + V(Y)$$

تست‌های فصل سوم

۱- تحلیل یک طرفه واریانس چه موقع مناسب است؟

1)زمانی که میانگین معلوم است . 2)زمانی که متغیر مستقل مشخص است.

3) زمانی که متغیر مستقل مشخص نیست. 4) زمانی که متغیر وابسته جنسیت است.

۲- نسبت Fدر تحلیل واریانس زمانی که ۴ باشد و احتمال۶۰۰ F/. باشد را تفسیر کنید.

1) فرضیه صفر به احتمال 6 در هزار مورد تحلیل واریانس قرار می‌گیرد.

2) از 6 مورد در هزار می‌توان به توزیع داده‌ها اعتماد کرد.

3) اگر فرضیه صفر صحیح می‌بود به احتمال 6 در 1000 **f** را معادل 4 یا بیشتر می‌دیدیم.

4) متغیرهای مورد استفاده دارای ارزش‌های کیفی‌اند که چنین نسبتی از**f** به دست آمده است.

۳- بزرگتر شدن مقدار Fناشی از چیست؟

1) بزرگتر شدن واریانس کل

2) کوچکتر شدن واریانس کل

3) بزرگتر شدن واریانس بین گروه‌ها نسبت به واریانس داخل گروه‌ها

4) کوچکتر شدن واریانس بین گروه‌ها نسبت به واریانس داخل گروه‌ها

۴- محققی می‌خواهد فرضیه:« متوسط سن ابتلا به اضطراب کمتر از متوسط سنی ابتلا به افسردگی می‌باشد» را مورد آزمون قرار دهد، کدام‌یک از آزمون‌های زیر لازم است انجام شود؟

1) مقایسه دو نسبت مستقل 2) مقایسه دو میانگین مستقل

3)مقایسه میانگین و انحراف معیار 4) مقایسه و میانگین همبسته

۵- محققی می‌خواهد نسبت دانشجویان دختر و پسر را در دانشکده علوم اجتماعی با همین نسبت در کل دانشگاه مقایسه کند آزمون آماری مناسب چه می‌باشد؟

1) آزمون**T-Test** 2) ضریب همبستگی**pearson**

3) مجذور خی(کی دو**X**2) 4) آنالیز واریانسی(**Anova**)

۶- برای سنجش تفاوت بین میانگین‌ها، کدام آزمون مناسب است؟

1)آزمون گاما
2)آزمون لاندا
3)آزمون‌های **F**و**T**
4)آزمون‌های کای اسکور **X**2

۷- اگر میانگین واریانس بین گروه‌ها برابر ۴۰۰ و میانگین واریانس داخل گروه‌ها برابر ۱۰۰ باشد، مقدار آزمون F چقدر خواهد بود؟

3(1
4(2
0/25(3
0/33(4

۸- پژوهشگری می‌خواهد تأثیر دو روش ارائه پوسترهای آموزشی و ارائه برنامه‌های کوتاه نمایش در کاهش تخلفات رانندگی را مقایسه نماید، آزمون آماری مورد استفاده در ارزیابی یافته‌ها چه می‌باشد؟

1)مقایسه میانگین‌ها
2)آزمون **T** زوج(**Paired- test**)
3)تست**T**(**T-Test**)
4)آزمون کا اسکوئر **X**2

۹- برای آزمون فرضیه این که میانگین نمرات درس«مبانی تحلیل محتوا» دانشجویان دو دانشگاه مختلف با هم تفاوت دارند یا خیر، کدام‌یک از آزمون‌های آماری زیر مناسب می‌باشد؟

1)آزمون مقایسه زوج‌ها(**Paired-TTesty**)
2)آزمون**Z**
3) آزمون مجذور کا اسکوئر **X**2
T-Test(4

۱۰- اگر در آنالیز واریانس، واریانس بین گروه‌ها برابر ۴۵۰۰ و واریانس داخل گروه‌ها برابر ۱۰۰۰ باشد، مقدار f چقدر است؟

4/5(1
5/5(2
35(3
45(4

۱۱- چنانچه قصد داشته باشیم میانگین مشارکت در فعالیت‌های اجتماعی را در نزد دو گروه شهری و روستایی مورد مقایسه قرار دهیم. کدام‌یک از آزمون‌های زیر مناسب‌تر است؟

1) آزمون مقایسه دو میانگین همبسته
2)آزمون**z** برای یک گروه
3)آزمون تحلیل واریانس
4)آزمون **t**برای دو میانگین مستقل

۱۲- در صورتی که فرضیه تحقیقی در یک پژوهش جهت‌دار باشد، کدام‌یک از آزمون‌های زیر مناسب‌تر است؟

1) دو دامنه　　2)یک دامنه　　3)بدون دامنه　　4)هم دامنه

۱۳- چنانچه قصد داشته باشیم میانگین درآمد یک گروه روزنامه نگار را با میانگین درآمد یک گروه خبرنگار مقایسه کنیم، کدام‌یک از فرضیه‌های زیر مناسب‌ترین است؟

بین میانگین‌ها اختلاف معنی‌داری وجود دارد.

1) $\begin{cases} H_0: \bar{X}_1 - \bar{X}_2 = 0 \\ H_A: \bar{X}_1 - \bar{X}_2 \neq 0 \end{cases}$　　2) $\begin{cases} H_0: u_1 - u_2 = 0 \\ H_A: u_1 - u_2 \neq 0 \end{cases}$

3) $\begin{cases} H_0: \bar{X}_1 - \bar{X}_2 \leq 0 \\ H_A: \bar{X}_1 - \bar{X}_2 \neq 0 \end{cases}$　　4) $\begin{cases} H_0: u_1 - u_2 \leq 0 \\ H_A: u_1 - u_2 \neq 0 \end{cases}$

۱۴- اگر در نمونه‌ای به حجم ۲۰۰، واریانس بین گروهی۱۲ و واریانس دو گروهی ۲۴ باشد، مقدار Fچقدر خواهد بود؟

1) 0/10　　2)50　　3)10　　4)0/50

۱۵- درآمد سالانه نمونه‌ای احتمالی از خانواده‌های سه شهر در جدول توزیع درآمد ارائه شده‌است؟برای آزمون معنی داری تفاوت متوسط درآمدها در این سه شهر چه آزمونی مناسب است؟

شهر الف	شهر ب	شهر ج
4	1	2
8	3	6

1) F

2) Z

3) X^2

4) T

۱۶- برای بررسی تفاوت تحصیلی(نمرات) ۲۰ دانشجو در دو ترم متوالی کدام آزمون مناسب است؟

1) رگرسیون یک متغیره　　2)آزمون T با نمونه‌های وابسته

3) آزمون T با نمونه‌های مستقل　　4) رگرسیون چند متغیره

۱۷- روابط کدام‌یک از متغیرهای زیر در آزمون T تست می‌شود؟

1) اسمی – فاصله‌ای　　2) رتبه‌ای- فاصله‌ای　　3) فاصله‌ای – نسبی　　4) اسمی- اسمی

۱۸- پر کاربردترین آزمون و سطح سنجش در تحلیل محتوا کدام است؟

1) نسبت – نسبی　　2) آزمون X^2 - نسبی

3) آزمون X^2 - اسمی　　4)نسبت – اسمی

۱۹- در صورتی که یک محقق قصد برآورد واریانس جامعه را از طریق نمونه‌ای که از این جامعه انتخاب کرده است، داشته باشد، مجموع مجذورات انحراف از میانگین را بر کدام‌یک از مقادیر زیر تقسیم کند.

1)$\mathbf{n}$　　2)$n-2$　　3)$n-1$　　4)$N-1$

۲۰- تحلیل واریانس با انتخاب و جایگزینی تصادفی با کدامیک از مقیاس‌های زیر بکار برده می‌شود؟

1)ترتیبی　　2)اسمی　　3)لیکرت و نسبی　　4)حداقل مقیاس فاصله‌ای

۲۱- در تجزیه و تحلیل برای سه گروه که در هر گروه ده آزمودنی در نظر گرفته شده مجموع مجذورات بین گروه‌ها ۴۵ و مجموع مجذورات بین درون‌گروه‌ها ۱۸۹ محاسبه شده است. عدد F کدام است؟

1) $2/84$　　2) $2/97$　　3) $3/15$　　4) $3/21$

۲۲- به منظور آزمون معنی‌دار بودن ضریب همبستگی ، از کدام توزیع استفاده می‌شود؟

1)$\mathbf{T}$　　2)$\mathbf{F}$　　3)کای دو　　4)پواسون

۲۳- جدول مقابل نمرات دو گروه در آزمون اندازه‌گیری مهارت آنها است، آماره آزمون T در مقایسه میانگین‌های این دو گروه کدام است؟

گروه اول	7	6	8	9	5
گروه دوم	4	8	5	7	

1) $\frac{\sqrt{5}}{3}$

2) $\frac{\sqrt{7}}{3}$

3) $\frac{\sqrt{5}}{4}$

4) $\frac{\sqrt{7}}{4}$

۲۴- اگر Z_1 تا Z_K متغیرهای استاندارد صفر و یک باشند، آنگاه توزیع $\sum_{i=1}^{K} z_i^2$ کدام است؟

1) F 2) کای دو 2) نرمال 4) t استیودنت

۲۵- در صورتی که قصد مقایسه میانگین یک گروه ۶۰ نفره در یک طرح پیش آزمون و پس آزمون را داشته باشیم، درجات آزادی مساوی است با

1) 58 2) 59 3) 120 4) 118

۲۶- درجات آزادی میانگین مجموع مجذورات بین گروه‌های مساوی کدامیک از عبارت‌های زیر است؟

1) $N-J$ 2) $N-1$ 3) $J-1$ 4) $N-1$

۲۷- رعایت مفروضه همسانی واریانس در کدامیک از آزمون‌های زیر با اصلاح درجات آزادی صورت می‌گیرد؟

1) میانگین‌های همبسته 2) ضریب همبستگی 3) میانگین مستقل 4) تحلیل واریانس یک طرفه

۲۸- آزمون‌های آماری با توجه به کدامیک از فعالیت‌های زیر تعیین می‌شود؟

1) هدف تحقیق 2) فرضیه‌های آماری 3) فرضیه‌های قیاسی 4) مسأله پژوهشی

پاسخنامه سوالات فصل سوم

۱- گزینه۲ صحیح است.

تحلیل واریانس به دو دسته تقسیم می‌شود:

1. تحلیل واریانس یک طرفه(**one- way**)

2. تحلیل واریانس دو طرفه(**two-way**)

تحلیل واریانس یک طرفه روشی است به منظور آزمون معنا دار بودن تفاوت بین میانگین‌ها، هنگامی که فقط یک متغیر مستقل دستکاری می‌شود، و طی آن اثر یک متغیر مستقل روی متغیر وابسته مورد بررسی قرار می‌گیرد. به عبارت دیگر در این نوع تحلیل واریانس یک متغیر مستقل و یک متغیر وابسته وجود دارد.(دلاور،1387: 348)

۲- گزینه ۳ صحیح است.

(دلاور،1387: 352)

۳- گزینه ۲ صحیح است.

$$F = \frac{(Msb)}{(Msw)}$$

میانگین واریانس بین گروه‌ها $\rightarrow$ Msb

میانگین واریانس درون گروه‌ها $\rightarrow$ Msw

با توجه به فورمول مذکور، بزرگتر شدن مقدار F ناشی از بزرگتر شدن واریانس بین گروه‌ها، نسبت به واریانس درون گروه‌هاست.(دلاور،1387: 354)

۴- گزینه ۴ صحیح است.

نمونه‌های همبسته در طرح‌های پژوهش زیر به کار برده می‌شوند:

1. اندازه‌گیری مکرر: در این نوع طرح برای هر یک از آزمودنی‌ها در نمونه مورد مطالعه دو بار اندازه‌گیری می‌شود.

2. طرح جفت‌های همتراز شده: در این نوع پژوهش آزمودنی‌های هر دو نمونه براساس یک یا چند متغیر که با متغیر وابسته یا متغیر ملاک رابطه دارند، همتراز می‌شوند.(دلاور،1387: 378)

۵- گزینه ۳ صحیح است.

آزمون خی دو، برای محاسبه رابطه معنی‌دار نسبت‌ها و درصدها در سطح سنجش اسمی بکار می‌رود، در این سوال جنسیت در سطح اسمی است (ساعی:108)

۶- گزینه ۳ صحیح است.

آزمون **F**(تحلیل واریانس)، یک روش آماری است که به منظور تحلیل تفاوت بین میانگین‌های دو یا چند نمونه آماری به کار می‌رود، در ضمن تحلیل واریانس زمانی که برای مقایسه میانگین‌های دو نمونه بکار برده شود، معادل آزمون **T**است. آزمون لاندا وآزمون گاما از ضرایب همبستگی می‌باشد. آزمون خی دو هم به منظور آزمون فرضیه درباره استقلال فراوانی‌هایی که در طبقه‌های مختلف قرار گرفته‌اند، بکار برده می‌شود.(دلاور، 1387: 346)

۷- گزینه ۲ صحیح است.

$$F = \frac{Msb}{Msw} = \frac{400}{100} = 4$$

میانگین واریانس بین گروه‌ها $\rightarrow$ Msb

میانگین واریانس درون گروه‌ها $\rightarrow$ Msw (دلاور،1387: 354)

۸- گزینه۳ صحیح است.

برای بدست آوردن آزمون معنی داری در این سوال از مقایسه میانگین‌ها استفاده می‌شود، مقایسه میانگین‌ها یا برای دو گروه بکار می‌رود یا برای بیش از دو گروه . در این سوال چون قصد داریم دو روش ارائه را با هم مقایسه کنیم، دو گروه هستند بنابراین از آزمون **t** بهره می‌گیریم. ضمناً مورد استفاده آزمون **T** مبتنی بر گروه‌های وابسته(زوج) زمانی است که از یک گروه آزمودنی دو بار آزمون گرفته شود ما در اینجا دو گروه آزمودنی داریم پس از آزمون**T** مبتنی بر گروه‌های مستقل استفاده می‌شود.(دلاور، 1387: 304)

۹- گزینه ۴ صحیح است.

آزمون **T(T-Test)** برای مقایسه میانگین بین دو نمونه یا دو گروه مناسب است.(دلاور، 1387: 304)

۱۰- گزینه ۱ صحیح است.

$$\Rightarrow \frac{4500}{1000} = 4/5 \ F = \frac{Msb}{Msw}$$

$\rightarrow$ Msb میانگین واریانس بین گروه‌ها

$\rightarrow$ Msw میانگین واریانس درون گروه‌ها

۱۱- گزینه ۴ صحیح است.

میانگین نمرات دو گروه مستقل (شهری – روستایی) توسط آزمون **t** مستقل انجام می‌شود. (دلاور، 1387: 312)

۱۲- گزینه ۲ صحیح است.

زمانی که محقق جهت تأثیر متغیر مستقل بر متغیر وابسته را قبل از تحقیق معلوم کند، فرضیه ما جهت دار است و دراین موارد باید ار آزمون یک دامنه استفاده شود. (دلاور، 1387: 297)

۱۳- گزینه ۲ صحیح است.

در آمار استنباطی فرضیه‌های صفر و خلاف به صورت پارامتر صورت‌بندی می‌شود، یعنی باید با حروف یونانی نوشته شوند. فرضیه صفر را با H_0 نشان می‌دهند. این فرض، اصل را بر این قرار می‌دهد که بین پارامترهای مورد مطالعه اختلاف یا ارتباط معناداری وجود ندارد. $\{H_0 : u_1 - u_2 = 0)$فرضیه خلاف را با H_1 یا H_A نشان می‌دهند. این فرضیه مخالف فرضیه صفر است و در اکثر موارد با فرضیه پژوهشی مطابقت دارد. به این معنی که فرضیه خلاف انتظار پژوهشگر را پیرامون نتایج آتی پژوهش بیان می‌کند. $\{H_a : u_1 - u_2 \neq 0)$

(دلاور، 1387: 292)

۱۴- گزینه ۴ صحیح است

$$F = \frac{Msb}{Msw} = \frac{12}{24} = 0/5$$

میانگین واریانس بین گروه‌ها $\rightarrow$ Msb

میانگین واریانس درون گروه‌ها $\rightarrow$ Msw

۱۵- گزینه ۱ صحیح است.

آزمون **F** برای بررسی تفاوت میانگین‌های سه گروه و بیشتر مورد استفاده قرار می‌گیرد.

۱۶- گزینه ۲ صحیح است.

چون در این سوال دو متغیر مستقل بر روی یک گروه ثابت اندازه‌گیری می‌شود، پس باید از **T** همبسته استفاده کرد.:

۱۷- گزینه ۱ صحیح است.

آنجا که متغیر وابسته در سطح فاصله‌ای باشد و تعداد طبقات متغیر مستقل کم باشد (اسمی) بهتر است در تحلیل داده‌ها به جای جدول توافقی از مقایسه میانگین‌های پاره‌گروه‌ها(آزمون **T**) استفاده کنیم.

۱۸- گزینه ۳ صحیح است.

پرکاربرد ترین آزمون تحلیل محتوا، آزمون خی دو است. که این آزمون بر معنی‌داری بین دو متغیر در سطح اسمی دلالت دارد.

۱۹- گزینه۴ صحیح است.

درجات آزادی مجموع مجذورات کل مطابق زیر، از جمع کردن درجات آزادی بین گروه‌ها و درون گروه‌ها بدست می‌آید.

$$d.ft = dfw + dfb = (N-k) + (k-1) = N-1$$

۲۰- گزینه ۴ صحیح است.

از مفروضه‌های تحلیل واریانس این است که مقیاس اندازه‌گیری حداقل فاصله‌ای باشد.

۲۱-گزینه ۴ صحیح است.

در این سوال ابتدا باید مجموع مجذورات بین گروهی و درون گروهی را به واریانس بین گروهی و درون گروهی تبدیل کنیم و سپس هر کدام را بر درجه آزادی خاص خود تقسیم کنیم.

$$F = \frac{Msb}{Msw} = \frac{\frac{ssb}{k-1}}{\frac{ssw}{N-1}}$$

$$SS_W = 189$$
$$df_w = N-k$$
$$SS_b = 45$$
$$df_b = k-1$$

$$F=\frac{\frac{45}{3-1}}{\frac{189}{30-3}}=\frac{22/5}{7}=3/21$$
$$F=3/21$$

قابل ذکر است که df_w از تفاضل تعداد کل آزمودنی‌ها(30 نفر) به تعداد گروه‌ها(یعنی 3) به دست می‌آید و df_b از تفاضل تعداد گروه‌ها از یک بدست می‌آید.

۲۲- گزینه ۱ صحیح است.

رد یا اثبات فرضیه (معنی‌داری همبستگی) از فرمول $t=r_s\sqrt{\frac{n-2}{1-r_s^2}}$ انجام می‌گیرد. (ساعی، 131)

۲۳- گزینه صحیح وجود ندارد.

x_1	x_2	$(x_1-\bar{x}_1)$	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)$	$(x_2-\bar{x}_2)^2$
7	4	0	0	-2	4
6	8	-1	1	2	4
8	5	1	1	-1	1
9	7	2	4	1	1
5		-2	4		
35	24		10		10

گروه اول $=x_1$ گروه دوم $=x_2$

$$s^2x_1=\frac{\sum(x_1-\bar{x}_1)^2}{N}=\frac{10}{5}=2$$
$$s^2x_2=\frac{\sum(x_2-\bar{x}_2)^2}{N}=\frac{10}{4}=2/5$$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}}}\Rightarrow\frac{35-24}{\sqrt{\frac{2}{5}+\frac{2/5}{4}}}=\frac{11}{\sqrt{0/4+0/62}}=\frac{11}{\sqrt{1/02}}=\frac{11}{1/009}=10/9$$
$$t=10/9$$

۲۴- گزینه ۳ صحیح است.

۲۵- گزینه ۲ صحیح است.

زیرا در این سوال برای مقایسه میانگین این گروه باید از آزمون **t** همبسته استفاده نمائیم که درجه آزادی در **t**همبسته **n-1** می‌باشد، یعنی $60-1=59$

۲۶- گزینه ۳ صحیح است.

درجه آزادی میانگین مجموع مجذورات $k-1$ یا $j-1$ می‌باشد.

۲۷- گزینه ۴ صحیح است.

به تجربه ثابت شده است که رعایت نکردن مفروضه‌های طبیعی بودن توزیع متغیر و همسانی واریانس‌ها بر اعتبار آزمـون $\mathbf{F}$ تأثیر زیادی ندارد. صحت این مسأله زمانی که اندازه نمونه‌ها مساوی باشد بیشتر است.

۲۸- گزینه ۲ صحیح است.

فصل چهارم:توزیع احتمال گسسته

4-1 متغیر تصادفی گسسته

همانطورکه بیان شد، متغیر تصادفی را معمولا بر حسب تعداد مقادیری که می‌تواند اختیار کند به دو دستهٔ « متغیر تصادفی گسسته» و « متغیر تصادفی پیوسته» تقسیم می‌کنند. عامل **x** را متغیر تصادفی گسسته می‌گویند که این عامل **x** مقادیری را که اختیار می‌کند، متغیر تصادفی مقدار خود را به طور تصادفی می‌گیرد و این مقدار تحت کنترل مشاهده کننده نمی‌باشد. که آن مقادیر دارای شرایط زیر باشند:اولا- قابل شمارش باشند ، ثانیا- جدا از هم باشند ، ثالثا – هر کدام احتمال مشخصی داشته باشند، به چنین عامل **x**، متغیر تصادفی گسسته می‌گویند.

4-2 چند توزیع احتمال گسسته

این متغیرها توزیع احتمالی دارند که مثالهایی از آنها در فصل قبل ارائه شد. سه توزیع احتمال گسسته وجود دارند که مدلهای خوبی برای توزیع احتمال متغیرهای تصادفی گسسته در مسائل عملی هستند. این سه توزیع عبارتنداز: توزیعهای دوجمله‌ای، پواسون، و فوق هندسی. در ادامه این توزیعها را بررسی کرده و موارد استفادهٔ آنها را بیان خواهیم کرد و فرمولهایی برای این سه توزیع به دست آورده، میانگین و واریانس آنها را معرفی خواهیم کرد.

4-3 متغیر تصادفی دو جمله‌ای

یکی از مهمترین انواع متغیرهای تصادفی گسسته، **متغیر تصادفی دو جمله‌ای** است که کاربردهای زیادی در شاخه‌های مختلف علوم دارد. این متغیر تصادفی وابسته به آزمایشی است که به **آزمایش دوجمله‌ای** معروف است که قبل از تعریف آن به تعریف آزمایش ساده‌تر **برنولی** می پردازیم.

۴-۳-۱ آزمایش برنولی

آزمایشی را که تنها دارای دو نتیجهٔ متفاوت باشد، آزمایش برنولی گویند. این نتایج موفقیت و شکست می‌نامند و به ترتیب با **S** و **F**، احتمال موفقیت را با **p** واحتمال شکست را با **q** نشان می‌دهند. چون این آزمایش تنها دو نتیجه دارد باید رابطهٔ زیر بین **p** و**q** برقرار باشد.

$$\mathbf{p}+\mathbf{q}=1$$

و در نتیجه: $\mathbf{q} = 1-\mathbf{p}$

یک مثال ساده از آزمایش برنولی، آزمایش انداختن یک سکهٔ سالم است. این آزمایش دارای دو نتیجه است، آمدن شیر یا آمدن خط. اگر آمدن شیر را موفقیت بنامیم آمدن خط شکست خواهد بود و چون سکه سالم است داریم:

$$p = \text{احتمال موفقیت} = p\left(\text{شیر}\right) = \frac{1}{2}$$

$$q = \text{احتمال شکست} = 1 - p = \frac{1}{2}$$

۴-۳-۲ آزمایش دو جمله‌ای

آزمایش دو جمله‌ای تعمیمی از آزمایش برنولی است و دارای شرایط زیر است.

1- آزمایش دو جمله‌ای شامل n بار تکرار یک آزمایش برنولی است.

2- ازمایشهای برنولی مستقل از هم انجام می‌شوند.

3- احتمال موفقیت در تمام آزمایشها ثابت است.

4- در این آزمایش، تعداد موفقیتهای به دست آمده در n آزمایش، مورد نظر است.

۴-۳-۳ متغیر تصادفی دو جمله‌ای

متغیر تصادفی **x** را که تعداد موفقیتهای به دست آمده در یک آزمایش دو جمله‌ای است، متغیر تصادفی **دو جمله‌ای** می‌نامند.

مثال- فرض کنید که **n** سکه با هم انداخته شوند و تعداد شیرهای مشاهده شده ثبت شود. دیدیم که پرتاب یک سکه یک آزمایش برنولی است و چون نتیجهٔ پرتاب سکه‌ها مستقل از هم و تعداد موفقیتها مورد نظر است این آزمایش یک آزمایش دو جمله‌ای است و متغیر تصادفی **x** که به صورت زیر تعریف می‌شود، یک متغیر تصادفی دو جمله‌ای خواهد بود.

$$x: \quad \text{تعداد شیرها در پرتاب } n \text{ سکه}$$

تذکر- برای تشخیص آزمایشهای دو جمله‌ای از سایر آزمایشها به موارد زیر توجه کنید.

1- اگر آزمایش شامل انتخاب n شیٔ از میان N شیٔ **با جایگذاری** باشد و تعداد اشیایی که دارای خاصیت یکسانی هستند مورد نظر باشد با آزمایش دو جمله‌ای سروکار داریم.

2- اگر انتخاب n شیء از میان N شیء **بدون جایگذاری** انجام شود ولی نسبت $\frac{n}{N}$ کوچک (کوچکتر از 0/05) باشد، بطوریکه احتمال موفقیت در تمام آزمایشهای برنولی تقریبا ثابت بماند می‌توان فرض کرد که آزمایشها بطور مستقل انجام شده‌اند و آزمایش انجام شده تقریبا دو جمله‌ای خواهد بود در غیر اینصورت آزمایش دو جمله‌ای نیست.

4-4 **توزیع احتمال دو جمله‌ای**

پس از تعریف آزمایش دو جمله‌ای و متغیر تصادفی دو جمله‌ای، توزیع احتمال متغیر تصادفی دو جمله‌ای را که به اختصار توزیع احتمال دوجمله‌ای نامیده می‌شود، معرفی می کنیم. در قسمتهای قبل، متغیر تصادفی دو جمله‌ای را « تعداد موفقیتها در یک آزمایش دو جمله‌ای که متشکل از **n** آزمایش برنولی مستقل است» تعریف کردیم و در این بخش توزیع احتمال دو جمله‌ای را برای **n=** 1,2,3 به دست آورده و فرمول کلی آن را که به ازای تمام **n** ها برقرار است بیان می کنیم.

برای **n=1** آزمایش، فضای نمونه‌ای شامل دو نقطه است که آنها را با **S** و **F** نشان می‌دهیم (**S** برای موفقیت و **f** برای شکست). طبق تعریف آزمایش برنولی احتمال موفقیت و شکست عبارتند از:

$$p(s) = p \ , \ p(F) = 1 - p$$

چون **x** تعداد موفقیتها در **n** ازمایش است می‌توان **x**=1 را به نقطهٔ **S** و **x**=0 را به نقطهٔ **F** نسبت داده و جدول توزیع احتمال **x** را تشکیل داد. جداول 4-1و 4-2 را ببینید.

نتایج آزمایش دو جمله‌ای وقتی **n**=2 ازمایش برنولی داریم در جدول 4-3 آمده است و توزیع احتمال **x** برای ازمایشی که از **n**=2 آزمایش برنولی تشکیل شده است مطابق جدول 4-4 است. فضای نمونه‌ای شامل چهار نقطه است و مثلا نماد **SF** نشان می دهد که نتیجهٔ آزمایش اول « موفقیت» و نتیجهٔ آزمایش دوم «شکست» است.

جدول 4-1 نتایج آزمایش دو جمله‌ای و احتمال آنها وقتی n=1

x	p(E_i)	نتایج آزمایش	پیشامدهای ساده
1	p	S	E_1
0	q	F	E_2

x	$p(E_i)$	نتایج آزمایش	پیشامدهای ساده
2	p^2	SS	E_1
1	pq	SF	E_2
1	qp	FS	E_3
0	q^2	FF	E_4

جدول 4-2 **p(x)** برای متغیر تصادفی دو جمله‌ای وقتی **n=1**

x	p(x)
0	q
1	p

$$\sum_{x=0}^{1} p(x) = p + q = 1$$

جدول 4-3 *نتایج ازمایش دو جمله‌ای و احتمال آنها وقتی* $n = 2$

جدول 4-4 *p(x) برای متغیر تصادفی دو جمله‌ای وقتی* $n=2$

x	p(x)
0	q^2
1	2pq
2	p2

$$\sum_{x=0}^{2} p(x) = (p+q)^2 = (1)^2 = 1$$

در جدول 4-3 مقادیر $\mathbf{p(E_i)}$ با استفاده از قانون ضرب احتمال به صورت زیر به دست می‌آید:

$$\mathbf{p(E_1)= p(SS)= p(S)\, p(S) = p^2 \quad p(E_3)= p(FS)= p(F)\, p(S) = qp}$$

$$\mathbf{p(E_2)= p(SF)= p(S)\, p(F) = pq p(E_4)= p(FF)= p(F)\, p(F) = q^2}$$

و مقادیر p(x) از روابط زیر محاسبه شده‌اند.

$$\mathbf{P(0)= P\,(x=0) = p(E_4)= q^2 \quad , P(1)= P\,(x=1) = p(E_2)+ p(E_3) = 2pq \quad , P(2)= P\,(x=2) =}$$

$\mathbf{p(E_1)= p^2}$مشاهده می‌شود که

$$\sum_{x=0}^{2} p(x) = q2 + 2pq + p2 = (p+q)^2 = (1)^2 = 1$$

برای **n=3** در جدول 4-5 نتایج آزمایش دو جمله‌ای برای **n=3**، نشان داده شده و با استاده از آن جدول توزیع احتمال **x**، که جدول 4-6 است تشکیل شده است.

جدول 4-5 *نتایج ازمایش دو جمله‌ای و احتمال آنها وقتی* $n = 3$

پیشامدهای ساده	نتایج آزمایش	$p(E_i)$	x
E_1	SSS	p3	3
E_2	SSF	$p^2 q$	2
E_3	SFS	$p^2 q$	2
E_4	FSS	$p^2 q$	2
E_5	SFF	pq^2	1
E_6	FSF	pq^2	1
E_7	FFS	pq^2	1
E_8	FFF	q^3	0

جدول 4-6 $p(x)$ *برای متغیر تصادفی دو جمله‌ای وقتی* $n=3$

x	p(x)
0	q^3
1	$3pq^2$
2	$3p^2q$
3	p^3

$\sum_{x=0}^{3} p(x) = (p+q)^3 = (1)^3=1$

بنابراین توزیع احتمال دوجمله‌ایی در حالت کلی عبارتست از:

$$p(x) = C_x^n p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

که x مقادیر 0 و1 و2 و 000 و n را اختیار می‌کند و بـا نمادهـای ! و C_x^n آشـنا مـی باشـیم. در فرمـول $p(x)$ مقـدار $p^x q^{n-x}$ نشان دهندهٔ احتمال پیشامد ساده‌ای است که شامل x موفقیت و $(n\text{-}x)$ شکست در ترتیـب معینـی اسـت و C_x^n تعداد این نوع پیشامدها را نشان می‌دهد.

4-5 میانگین و واریانس توزیع دو جمله‌ای

می‌دانیم که توزیع دو جمله‌ایی شامل n آزمایش مستقل برنولی است. اگر فرض کنیم که $n=1$ است، توزیع دو جمله‌ای به توزیع برنولی‌ای تبدیل می شود که میانگین و واریانس آن به اینصورت است:

y	0	1
p(Y=y)	q	p

$$E(Y) = 0 \times q + 1 \times p = p$$
$$E(Y^2) = (0)^2 \times q + (1)^2 \times p = p$$
$$V(Y) = E(Y^2) - [E(Y)]^2 = P - P^2 = P(1-P) = pq$$

اگر X را تعداد موفقیتها در توزیع دو جمله‌ای بدانیم، خواهیم داشت:

$$X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$$

در این رابطه Y_1 تعاد موفقیتها در آزمایش اول (0 یا 1)، Y_2 تعداد موفقیتها در آزمـایش دوم (0 یـا 1) و 000 اسـت و چون آزمایشها از هم مستقل هستند، خواهیم داشت:

$$E(X) = E(Y_1) + \ldots + E(Y_n) = p + \cdots + p = np$$
$$V(X) = V(Y_1) + \ldots + V(Y_1) = pq + \cdots + pq = npq$$

4-6 توزیع پواسون

درتوزیع دو جمله‌ای، وقتی n بزرگ شود محاسبات کار ساده‌ای نخواهد بود؛ مثلا اگر بخواهیم بدانیم از بین 2000 واحـد کالای موجود که هر یک با احتمال 0/0015 معیوبند، با چه احتمالی 5 عدد آنها معیوبند، ناچـاریم $\binom{2000}{5}$ و همینطـور

$(0.0015)(0.9985)^{1995}$ را محاسبه و در هم ضرب کنیم که این محاسبات کار دشواری است. در اینگونه مواقع یعنی هنگامیکه **n** به سمت بی نهایت و **p** به سمت صفر یا یک میل کند و در عین حال مقدار **np** ثابت بماند، استفاده از توزیع پواسون با این احتمال تقریب مناسبی برای توزیع دو جمله‌ای خواهد بود.

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \text{ و} x = 0, 1, 2, \ldots$$

که در این رابطه پارامتر توزیع و برابر **np**، $(\boldsymbol{\lambda} = \boldsymbol{np})$ و $\boldsymbol{e}$= 2/718 است.

به طورکلی وقتی $\boldsymbol{n}$≥20 و $\boldsymbol{p}$≤0/05 باشد، توزیع پواسون تقریب خوبی برای توزیع دو جمله‌ای و وقتی $\boldsymbol{n}$≥100 و $\boldsymbol{np} \leq$ 10 باشد، تقریبی بسیار عالی برای آن محسوب می‌شود.

امید ریاضی و واریانس توزیع پواسون $\boldsymbol{\lambda}$ است. به عبارت دیگر $\boldsymbol{E(X) = \lambda}$ و $\boldsymbol{V(X) = \lambda}$. از بین توزیعهای رایج، چه پیوسته و چه گسسته، توزیعی که میانگین و واریانس آن با هم برابر اس توزیع پواسون است و این یکی از خصوصیات جالب توجه این توزیع است.

۴-۶-۱ توزیع پواسون برای تعداد مراجعات

معمولا تعداد مراجعاتی که به سیستمی می‌شود، از توزیع پواسون برخوردار است که در این صورت متوسط تعداد مراجعات در واحد زمان است. کاربرد توزیع پواسون در این زمینه خیلی بیشتر از تقریب آن برای توزیع دو جمله ای است؛ مثلا تعداد مشتریانی که در هر ساعت به یک سیستم صف مراجعه می‌کنند، تعداد اتوموبیل‌هایی که در هر دقیقه برای زدن بنزین به پمپ بنزینی مراجعه می‌کنند، تعداد مشتریانی که در هر ساعت به رستورانی مراجعه می‌کنند،... می‌تواند دارای توزیع پواسون باشد. در این حالت فرمول توزیع پواسون به صورت زیر است:

$$P(X = x) = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!} \text{ و} x = 0, 1, 2, \ldots$$

در این رابطه، **t** واحد زمانی است که متوسط مراجعات $(\boldsymbol{\lambda})$ با توجه به آن تعریف می‌شود.

مثال- در یک بانک به طور متوسط 7 نفر حساب خود را در یک سال می‌بندند. فرض کنید تعدا کسانیکه در یک دورهٔ زمانی خاص حساب خود را می بندند توزیع پواسون داشته باشد، مقادیر زیر را محاسبه کنید.

الف) احتمال اینکه در یک دورهٔ 4 ماهه هیچکس حساب خود را نبندد.

ب) احتمال اینکه در یک سال حداقل سه نفر حساب خود را ببندند.

حل با دانستن مقدار متوسط یک متغیر تصادفی پواسون توزیع آن کاملا معلوم می‌شود. در هـر کـدام از مـوارد الـف و ب ابتدا مقدار μt را به دست می‌آوریم.

الف) می‌دانیم در یک سال بطور متوسط 7 نفر حساب خود را می‌بندند پـس در مـدت 4 مـاه بـه طـور متوسـط 2/33 = $\lambda t = 7 \times \frac{4}{12}$ نفر حساب خود را می‌بندند. اگر متغیر تصادفی x را به صورت زیر تعریف کنیم.

x: تعداد افرادی که در یک دورهٔ چهار ماهه حساب خود را می‌بندند.

x متغیر تصادفی با پارامتر μ= 2/3 است و توزیع احتمال عبارتست از:

$$P(X) = \frac{e^{-2/3}(2/3)^x}{x!} \text{و} x = 0, 1, 2, \dots$$

احتمال پیشامد (x=0) مورد نظر است، پس

$$p(x = 0) = p(0) = e^{-2.3} = 0/1003$$

ب) در این قسمت متغیر تصادفی x، توزیع پواسون با پارامتر $\lambda t = 7$ داردف زیرا مدت دورهٔ زمـانی یـک سـال اسـت. بنابراین توزیع احتمال x عبارتست از

$$P(X) = \frac{e^{-7}(7)^x}{x!} \text{و} x = 0, 1, 2, \dots$$

احتمال پیشامد (x≥3) مورد نظر است. با استفاده از رابطهٔ پیشامدهای مکمل داریم:

$$p(x \geq 3) = 1 - p(x < 3) = 1 - p(x \leq 2)$$
$$p(x \leq 2) = p(0) + p(1) + p(2)$$

که در آن

$$p(0) = e^{-7} = 0.0009$$
$$p(1) = \frac{e^{-7} \times 7}{1!} = 0/0063$$
$$p(2) = \frac{e^{-7} \times 7^2}{2!} = 0/0221$$
$$p(x \leq 2) = 0/0009 + 0/0063 + 0/0221 = 0/0293$$
$$p(x \geq 3) = 1 - 0/0293 = 0/9707$$

4-7 **توزیع احتمال فوق هندسی**

فرض کنید که می‌خواهیم یک نمونه از جامعه‌ای انتخاب کنیم و ببینیم که چه تعدادی از اعضای نمونـه دارای مشخصـهٔ معینی هستند وجود مشخصه در یک عضو نمونه « موفقیت» و عدم وجود مشخصه را « شکست» تلقی می‌کنیم.

در بخش توزیع احتمال دو جمله‌ای گفتیم که اگر انتخاب نمونه بدون جایگذاری انجام شود ولی تعداد عضوهای نمونه نسبت به تعداد عضوهای جامعه خیلی کم باشد، $n \leq 0/05\ N$ ، فرض می‌کنیم این انتخاب یک آزمایش دو جمله‌ای است و **x**، تعداد موفقیتها در نمونه، یک متغیر تصادفی با توزیع تقریبی دو جمله‌ای خواهد بود. بنابراین فقط در مواردی که انتخاب بدون جایگذاری انجام شود و اندازهٔ نمونه نسبت به اندازهٔ جامعه بزرگ باشد، $n > 0/05\ N$، از توزیع فوق هندسی استفاده می‌کنیم.

۴-۷-۱ متغیر تصادفی فوق هندسیاگر انتخابنمونه **بدون جایگذاری** انجام شود، احتمال موفقیت در یک آزمایش، به نتیجهٔ آزمایشهای قبل بستگی دارد و **x**، تعداد موفقیتها در نمونه، یک متغیر تصادفی **فوق هندسی** است.

در توزیع احتمال فوق هندسی از نمادهای زیر استفاده می‌کنیم:

N= تعداد اعضای جامعه،

K= تعداد اعضای جامعه که دارای مشخصهٔ مورد نظر هستند،

n= تعداد اعضای نمونه،

X= تعداد موفقیتها در نمونه.

۴-۷-۲ توزیع احتمال فوق هندسیتوزیع احتمال فوق هندسی برایمتغیر تصادفی **x** به شرح زیر است:

$$p(x) = \frac{C_x^k C_{n-x}^{N-k}}{C_n^N} \quad , x = 0, 1, 2, \dots, n n \leq k$$

و

$$\mu = E(x) = n\frac{k}{N} \quad , \sigma^2 = V(X) = n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1}$$

مثال- از یک گروه 20 نفری، 10 نفر به طور تصادفی برای انجام کاری انتخاب می‌شوند. احتمال اینکه 5 شخص مورد نظر در بین این 10 نفر باشند، چقدر است؟

حل انتخاب 10 نفر بدون جایگذاری انجام می‌شود و تعداد عضوهای نمونه نسبت به تعداد عضوهای جامعه زیاد است (به نسبت $\frac{10}{20}$) پس در این مثال با توزیع فوق هندسی سروکار داریم در اینجا **n**=10 و **N**=20 ، و چون 5 نفر دارای مشخصه مورد نظر هستند، پس **K**=5 و چون هر 5 نفر انتخاب شوند پس **x**= 5.

ابتدا فرمول کلی توزیع احتمال **x** را می‌نویسیم

$$p(x) = \frac{C_x^k C_{n-x}^{N-k}}{C_n^N} = \frac{C_x^5 C_{10-x}^{15}}{C_{10}^{20}} = \frac{\frac{5!}{5!0!}\frac{15!}{5!10!}}{\frac{20!}{10!10!}} = \frac{10!10!}{20!} = \frac{21}{1292} = 0/0163$$

فصل پنجم: توزیع احتمال پیوسته

1-5 متغیر تصادفی پیوسته

عامل **x** را متغیر تصادفی پیوسته می‌گویند که:

اولا- مقادیری را که اختیار می‌کند آن مقادیر، قابل شمارش نباشند (مانند کلیهٔ اعداد بین صفر و یک که، قابل شمارش نیستند.

ثانیا - جدا از هم نباشند.

ثالثا- هر کدام احتمال مشخصی نداشته باشند.

همچنین به عامل **x** متغیر تصادفی پیوسته گویند وقتی دارای شرایط زیر باشد:

اولشرط : $\int_{-\infty}^{+\infty} f(x)dx = 1$

دومشرط : $f(x) \geq 0$

به شرایط متغیر تصادفی پیوسته، شرایط تابع چگالی نیز می گویند.

تابع فراوانی = تابع احتمال = تابع چگالی = $f(x)$

2-5 تابع فراوانی یا چگالی

رابطهٔ بین **x** و **f(x)** با در نظر گرفتن ضابطهٔ **y = f(x)** را تابع چگالی می‌گویند.

نکته: اگر از تابع چگالی انتگرال بگیریم تابع توزیع تجمعی بدست می‌آید:

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

نکته: اگر از تابع چگالی مشتق اول بگیریم و مساوی صفر قرار دهیم مقدار x حاصله، همان مد توزیع خواهد بود:

$$\acute{f}(x) = 0 \quad \rightarrow \quad x = M_o$$

3-5 تابع توزیع تجمعی

اگر سه شرط زیر برقرار باشد تابع مورد نظر، تابع توزیع تجمعی خواهد شد:

شرط اول : $F(-\infty) = 0$

شرط دوم : $F(+\infty) = 1$

شرط سوم : $\acute{F}(x) \geq 0$

نکته: همانطورکه بیان شد، اگر از تابع توزیع تجمعی مشتق اول بگیریم، تابع چگالی حاصل می‌شود؛ و اگر از تابع توزیع تجمعی مشتق دوم بگیریم و مساوی صفر قرار دهیم مقدار **x** حاصله، مد خواهد بود؛ به عبارتی:

$$\acute{F}(x) = f(x)$$
$$\acute{\acute{F}}(x) = 0 \rightarrow x = M_o$$

نکته: اگر تابع توزیع تجمعی را برابر 0/5، 0/25 و 0/75 قرار دهیم و معادله را حل کنیم مقدار **x** حاصل شده به ترتیب میانه یا چارک دوم، چارک اول و چارک سوم خواهد بود:

$$F(x) = \frac{1}{2} \rightarrow x = Q_2 = M_d = \text{چارک دوم}$$
$$F(x) = \frac{1}{4} \rightarrow x = Q_1 = \text{اولچارک}$$
$$F(x) = \frac{3}{4} \rightarrow x = Q_3 = \text{سومچارک}$$

4-5 امید ریاضی x (میانگین توزیع)و واریانس و انحراف معیار توزیع

$$E(X) = \bar{X} = \int_{-\infty}^{+\infty} x.f(x)dx$$
$$E(X^2) = \bar{X} = \int_{-\infty}^{+\infty} x^2.f(x)dx$$
$$\sigma^2 = E(X^2) - E^2(X) \quad \rightarrow \quad \sigma = \sqrt{E(X^2) - E^2(X)}$$

مثال– تابع **f(x)** با ضابطهٔ $\mathbf{f(x) = k\,(-x^2 + bx)}$ در دامنهٔ (10 و 0) تعریف شده است و تابع فراوانی یک متغیر تصادفی است. مطلوبست:

الف) مقدار **k** را تعیین کنید؟

ب) چند درصد افراد اندازهٔشان بین 3 و 5 است؟

ج) میانگین توزیع را حساب کنید؟

د) واریانس توزیع را حساب کنید؟

و) ثابت کنید که مقدار مد با میانه مساوی می‌باشد؟

الف) برای اینکه **k** را حساب کنیم شرایط تابع فراوانی را می‌نویسیم:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \rightarrow \int_0^{10} k\,(-x^2 + 10x)dx = 1 : \text{اولشرط}$$

$$\boldsymbol{f(x) \geq 0} \qquad : \text{دومشرط}$$

$$\rightarrow k\left|\frac{-1}{2+1}x^{2+1} + \frac{10}{1+1}x^{1+1}\right|_0^{10} = 1 \rightarrow k\left|\frac{-1}{3} \times 10^3 + \frac{10}{2} \times 10^2\right| - k|0+0| = 1 \rightarrow k = \frac{6}{1000}$$

ب) در این قسمت ابتدا معادلهٔ تابع توزیع را محاسبه کرده و سپس به محاسبه می‌پردازیم:

$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{6}{1000}(-x^2 + 10x)dx \rightarrow F(x) = \frac{6}{1000}\left|-\frac{1}{3}x^3 + \frac{10}{2}x^2\right|_0^x \rightarrow F(x)$$
$$= \frac{6}{1000}\left(-\frac{1}{3}x^3 + \frac{10}{2}x^2\right)$$

$0/284 = 0/216 - 0/5\boldsymbol{F(5) - F(3)} =$

ج) میانگین توزیع:

$$E(X) = \bar{X} = \int_{-\infty}^{+\infty} x.f(x)dx = \int_0^{10} x.\frac{6}{1000}(-x^2 + 10x)\mathrm{dx} = \frac{6}{1000}\left|-\frac{1}{4}x^4 + \frac{10}{3}x^3\right|_0^{10} = 10^4 \times \frac{6}{1000}\left(-\frac{1}{4} + \frac{1}{3}\right) = 5$$

د) واریانس توزیع: برای پیدا کردن واریانس از میاگین توزیع استفاده می‌کنیم.

$$E(X^2) = \bar{X} = \int_{-\infty}^{+\infty} x^2.f(x)dx \to \int_0^{10} x^2.\frac{6}{1000}(-x^2 + 10x)\mathrm{dx}$$

$$\to \frac{6}{1000}\int_0^{10}(-x^4 + 10\,x^3) = \frac{6}{1000}\left|-\frac{1}{5}x^5 + \frac{10}{4}x^4\right|_0^{10} = 30$$

$$\sigma^2 = E(X^2) - E^2(X) = 30 - 5^2 = 5 \ \to \ \sigma = \sqrt{5}$$

و)برای پیداکردن مد از تابع چگالی مشتق می‌گیریم و آنرا مساوی صفر قرار می‌دهیم. مقدار $\mathbf{x}$ حاصل شده مد نامیده می‌شود.

$$\acute{f}(x) = \frac{6}{1000}(-2x + 10) = 0 \to -2x + 10 = 0 \to M_0 = x = 5$$

در تابع توزیع محاسبه شده، به جای $\mathbf{x}$ مقدار مد (یعنی 5) را قرار می‌دهیم، اگر حاصل برابر 0/5 شود لذا می‌گوییم مد همان میانه است (برابر 5).

$$F(5) = \frac{6}{1000} \times 5^3\left(-\frac{1}{3} + 1\right) = \frac{1}{2} \to M_{\mathrm{d}} = 5 = M_{\mathrm{o}}$$

مثال– تابع $\mathbf{F(x)}$ با ضابطهٔ $F(x) = \frac{7}{4}\left(\frac{X-1}{X+2}\right)$ در دامنهٔ (5 و 1) تعریف شده است.

اولا تحقیق کنید که این تابع، تابع توزیع یک متغیر تصادفی است؟

ثانیا تابع فراوانی و میانه و چارک اول و چارک سوم توزیع را پیدا کنید؟

اولا)

$$F(-\infty) = 0 \to F(1) = \frac{7 \times 1 - 7}{4 \times 1 + 8} = 0$$ شرط اول

$$F(+\infty) = 0 \to F(15) = \frac{7 \times 5 - 7}{4 \times 5 + 8} = 1$$ شرط دوم

$$\acute{F}(x) \geq 0 \to \acute{F}(x) = \frac{84}{(4x+8)^2} > 0$$ شرط سوم

چون هر سه شرط فوق برقرار است پس نتیجه میگیریم که تابع فوق یک تابع توزیع متغیر تصادفی است.

ثانیا) اگر از تابع توزیع مشتق بگیریم تابع فراوانی بدست می‌آید:

$$f(x) = \frac{84}{(4x + 8)^2} = \frac{21}{4}.\frac{1}{(x + 2)^2}$$

$$\acute{f}(x) = \frac{-21}{2(x + 2)^3} < 0$$

همانطورکه مشاهده می‌کنیم مشتق تابع احتمال در بازهٔ (5 و 1)، همواره منفی است و نقطهٔ اکسترمم نسبی (ماکزیمم نسبی) وجود ندارد و تابع احتمال نزولی می‌باشد.

برای پیدا کردن چارک اول، دوم و سوم، تابع توزیع را به ترتیب $\boldsymbol{F(x) = \frac{1}{4}, \frac{3}{4}, \frac{1}{2}}$ قرار دهیم و مسأله را حل کنیم:

$$F(x) = \frac{1}{4} \to 28x - 28 = 4x + 8 \to Q_1 = \mathrm{x} = \frac{36}{24} = \frac{3}{2}$$ اولچارک

$$F(x) = \frac{1}{2} \to 14x - 14 = 4x + 8 \to Q_2 = \mathrm{M_d} = \mathrm{x} = \frac{36}{24} = \frac{22}{10}$$ چارک دوم

$$F(x) = \frac{3}{4} \to 28x - 28 = 12x + 24 \to Q_3 = \mathrm{x} = \frac{13}{4}$$ سومچارک

اگر بخواهیم امید ریاضی توزیع فوق را محاسبه کنیم بطریق زیر عمل می‌کنیم:

$$E(X) = \bar{X} = \int_1^5 x.f(x)dx = \frac{21}{4}\int_1^5 x.\frac{1}{(\mathrm{x}+2)^2}dx = \frac{21}{8}\int_1^5 \frac{2x+4-4}{(\mathrm{x}+2)^2}dx = \frac{21}{8}\left[\int_1^5 \frac{2x+4}{(\mathrm{x}+2)^2}dx - \int_1^5 \frac{4}{(\mathrm{x}+2)^2}dx\right]$$

$$\to$$

$$3/746 \to \frac{21}{8}\left[\ln(\mathrm{x}+2)^2 + \frac{8}{(\mathrm{x}+2)^3}\right]_1^5 = \frac{21}{8}\left[(\ln 49 + \frac{8}{343}) - (\ln 9 + \frac{8}{27})\right] = \frac{21}{8}\left[\ln\frac{49}{9} - \frac{273}{1000}\right] = \frac{21}{8}\left[\frac{17}{10} - \frac{273}{1000}\right] =$$

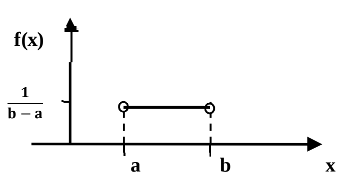
5-5توزیع یکنواخت

متغیر پیوستهٔ $\mathbf{X}$ را در نظر بگیرید که مقادیر خود را بین دو نقطه $\mathbf{a}$ و $\mathbf{b}$ انتخاب می‌کند و$\mathbf{a} < \mathbf{b}$است. اگر احتمال وقوع $\mathbf{X}$ در فاصله‌های هم‌اندازه، در فاصلهٔ $\mathbf{a}$ و $\mathbf{b}$ برابر باشد، چگالی احتمال مربوط به آن یکنواخت خواهد بود.

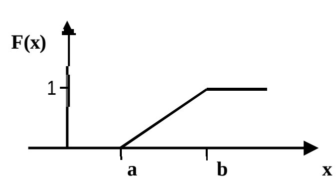
متغیر تصادفی پیوستهٔ $\mathbf{x}$ در صورتی دارای چگالی یکنواخت است که چگالی احتمال آن به این صورت باشد:

$$\mathrm{f(x)} = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0 & \text{در غیر این صورت} \end{cases} \to F(X) = \begin{cases} \dfrac{X-a}{b-a}, & a < x < b \\ 1, & x > b \\ 0, & x < b \end{cases}$$

و نمودارهای تابع چگالی و تابع توزیع از قرار زیر است:



و



میانگین و واریانس چگالی احتمال یکنواخت طبق تعریف برابر است با :

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx = \frac{1}{2}(\mathrm{a}+\mathrm{b}) \qquad , \mathrm{V(X)} = \int_{-\infty}^{+\infty}[X - E(X)]^2.f(x)dx = \frac{1}{12}(b-a)^2$$

6-5 **توزیع نمایی**

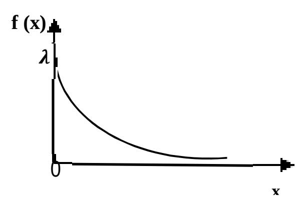
اگر تعداد موفقیتها یا ورودیهادارای توزیع پواسون باشد، زمان بین موفقیتها یا ورودیهای متوالی دارای توزیع « نمایی منفی» است. چون زمان پیوسته است، توزیع نمایی منفی نیز توزیعی پیوسته است.

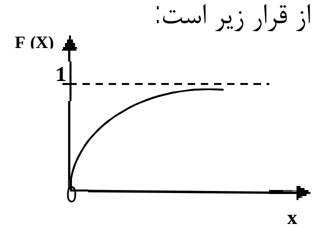
تابع چگالی و توزیع، توزیع نمایی منفی به صورت زیر است:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x \\ 0 & \text{, در غیر این صورت} \end{cases}$$

$$F(x) = P(x \le x) = \int_0^x \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^x = 1 - e^{-\lambda x} \rightarrow F(x) = \begin{cases} 1 - e^{-\lambda x}, & 0 < x \\ 0 & \text{, در غیر این صورت} \end{cases}$$

و نمودارهای تابع چگالی و تابع توزیع آن از قرار زیر است:





میانگین و واریانس چگالی احتمال توزیع نمایی منفی طبق تعریف برابر است با :

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx = \int_0^{+\infty} x.\lambda e^{-\lambda x}dx = \frac{1}{\lambda}$$

$$V(X) = \int_{-\infty}^{+\infty} [X - E(X)]^2 . f(x)dx = \int_0^{+\infty} \left(x - \frac{1}{\lambda}\right)^2 . \lambda e^{-\lambda x}dx = \frac{1}{\lambda^2}$$

7-5 **توزیع نرمال**

مهمترین توزیع پیوسته، توزیع نرما ل است. این توزیع نقش اساسی در آمار دارد و دارای کاربردهای وسیعی اس؛ چراکه اولا خیلی از پدیده‌های طبیعی دارای این توزیع هستند و ثانیا شکل حدی بسیاری از توزیعهای دیگر نیز نرمال است.

توزیع نرمال را می‌توان چنین تعریف کرد: « متغیر تصادفی پیوستهٔ Xبا میانگینμ و انحراف معیار $\boldsymbol{\sigma}$ دارای توزیع نرمال است اگر تابع چگالی آن به این صورت باشد:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2}$$

متغیر تصادفی $\mathbf{X}$ با میانگین μ و انحراف معیار $\boldsymbol{\sigma}$ را به صورت : $\boldsymbol{X}\sim(\mu,\boldsymbol{\sigma})$ نشان می‌دهیم و به این صورت می‌خوانیم: $\mathbf{X}$ دارای توزیع نرمال با میانگین μ و انحراف معیار $\boldsymbol{\sigma}$ است. μ و $\boldsymbol{\sigma}$ دو پارامتر توزیع نرمال می‌باشند که با مشخص بودن آنها، توزیع دقیقا مشخص و منحنی آن قابل ترسیم می‌شود.

۵-۷-۱توزیع نرمال استاندارد: تابع چگالی احتمال نرمال به گونه‌ای است که محاسبهٔ احتمال مورد نظر آن، به این دلیل که نمی‌توان از این تابع انتگرال گرفت، کار مشکلی است؛ بنابراین جدولی تهیه شده که فقط برای توزیع نرمال با میانگین صفر و واریانس یک قابل استفاده است. حال اگر بتوانیم به جای متغیر $\boldsymbol{X}\sim(\mu,\boldsymbol{\sigma})$ از متغیر نرمالی استفاده می‌کنیم که میانگین آن صفر و واریانس آن یک باشد، می‌توانیم برای پیدا کردن احتمال مورد نظر به **جدول** انتهای کتاب مراجعه کنیم.

گفتیم که تابع چگالی منحنی نرمال چنین است:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2} \quad \text{برای} \quad -\infty < x < +\infty$$

در این تابع، متغیر تصادفی $\boldsymbol{X}$ دارای میانگین μ و واریانس $\boldsymbol{\sigma}^{\mathbf{2}}$ است. حال اگر متغیری همچون $\boldsymbol{Z}$ را به اینصورت تعریف کنیم: $\mathbf{Z} = \frac{\boldsymbol{X-\mu}}{\boldsymbol{\sigma}}$؛ آنگاه تابع چگالی به اینصورت خلاصه می‌شود:

$$\varphi(z) == \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}z^2}$$

این متغیر دارای توزیع نرمال با میانگین صفر و واریانس یک به شرح زیر است:

$$\mathrm{E(Z)} = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\mathrm{E}\,(X-\mu) = \frac{1}{\sigma}[E(X) - E(\mu)] = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}[V(X) + V(\mu)] = \frac{1}{\sigma^2}(\sigma^2+0) = 1$$

به این توزیع، توزیع نرمال استاندارد می‌گوییم؛ بنابراین می‌توان گفت هر توزیع متغیر تصادفی نرمالی همانند $\mathbf{Z}$ که دارای میانگین صفر و واریانس یک باشد، توزیع نرمال استاندارد است که به زبان ریاضی چنین نوشته می‌شود: $\boldsymbol{Z}\sim(\mathbf{0},\mathbf{1})$.

با داشتن میانگین و واریانس هر متغیری- در صورتیکه توزیع آن نرمال باشد- می‌توان ابتدا آن را به توزیع نرمال استاندارد تبدیل کرد و سپس با مراجعه به جداول مربوطه (پیوست) احتمال آن را پیدا کرد.

نکته :در توزیع نرمال، احتمال فاصله‌ای به اندازهٔ « 1 انحراف معیار در هر طرف میانگین»، برابر 0/68 ، «2 انحراف معیار در هر طرف میانگین» برابر 0/954؛ و « 3 انحراف معیار در هر طرف میانگین»، برابر 0/997 است؛ به بیان ریاضی:

$0/683 P(\mu - \sigma \leq X \leq \mu + \sigma) \sim P(-1 \leq Z \leq 1) =$

$0/954 P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \sim P(-2 \leq Z \leq 2) =$

$0/997 P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \sim P(-3 \leq Z \leq 3) =$

مثال– دستگاه پر کننده شیشه‌های آبلیمو طوری تنظیم شده است که فقط 330 گرم آبلیمو را در داخل هر شیشه بریزد، با وجود این میزان آبلیمویی که وارد هر شیشه می‌شود دارای توزیع نرمال میانگین 330 گرم و انحراف معیار 5 گرم است. می‌خواهیم این احتمالات را محاسبه کنیم:

الف) شیشه‌ای بین 322 تا 328 گرم آبلیمو داشته باشد.

ب) شیشه‌ای بیش از 335 گرم آبلیمو داشته باشد.

ج) دایرهٔ کنترل کیفیت، میزان آبلیموی 70 شیشه را به صورت تصادفی وزن می‌کند. انتظار می رود چند شیشه بیش از335 گرم آبلیمو داشته باشد.

الف)

$$P(322 \leq X \leq 328) = P\left(\frac{24-30}{9} \leq Z \leq \frac{43-30}{9}\right) = P(-1/6 \leq Z \leq -0/4)$$

$$= P(Z \leq -0/4) - P(Z \leq -1/6) = 0/3446 - 0/0548 = 0/2898$$

ب)

$$P(X \geq 335) = P\left(Z \geq \frac{335-330}{5}\right) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0/8413 = 0/1587$$

د) تعداد شیشه‌هایی که انتظار می‌رود بیش از 335 گرم آبلیمو داشته باشند:

$$0/1587 \times 70 = 11/109 \simeq 11$$

5-8تقریب توزیع دو جمله‌ای به وسیلهٔ توزیع نرمال

اگر متغیر تصادفی $\mathbf{X}$ دارای توزیع دو جمله‌ای با پارامترهای $\mathbf{n}$ و $\mathbf{p}$ باشد، در صورتیکه $\mathbf{n}$ بزرگ باشد و $\mathbf{p}$ به صفر یا یک زیاد نزدیک نباشد، تقریب نرمال با پارامترهای $\mu = np$ و $\sqrt{npq}$ = تقریب خوبی برای توزیع دو جمله‌ای خواهد بود؛ به عبارت ریاضی: $X = N(np, \sqrt{npq})$. بنا به تجربه، اگر $\mathbf{np}$ و $\mathbf{nq}$ هر دو بزرگتر از 5 باشند، تقریب نرمال تقریب خوبی خواهد بود. در مواردی هم که $\mathbf{p}$ به 0/5 نزدیک باشد، تقریب نرمال برای $\mathbf{n}$ های کوچک نیز خوب است.

تقریب نرمال به ما این امکان را می‌دهد که احتمالهای مربوط به متغیر تصادفی **X** با توزیع دو جمله‌ای را چنان محاسبه کنیم که گویی **X** متغیر تصادفی با توزیع نرمال است. همانطورکه می‌دانیم، توزیع دوجمله‌ای توزیعی گسسته و توزیع نرمال توزیعی پیوسته است؛ مثلا در توزیع دو جمله‌ای با $n=10$ و $p=0/4$، $P(X=5)$ مقداری مثبت است، ولی در توزیع نرمال مقدار $P(X=5)$ برابر صفر است؛ بنابراین وقتی تقریبنرمال را برای دو جمله‌ای بکار می‌بریم، باید از تصحیح پیوستگی استفاده کنیم؛ یعنی به جای محاسبهٔ $P(X=5)$ باید $P(4/5 \le X \le 5/5)$ محاسبه شود. تصحیحهای پیوستگی مختلف در **جدول** زیر نشان داده شده است.

احتمال موردنظر از توزیع دو جمله‌ای	احتمال موردنظر از توزیع نرمال
$P(X = x)$	$P(x - 0/5 \le X \le x + 0/5)$
$P(X \le x)$	$P(X \le x + 0/5)$
$P(X < x)$	$P(X \le x - 1 + 0/5) = P(X \le x - 0/5)$
$P(X \ge x)$	$P(X \ge x - 0/5)$
$P(X > x) = P(X \ge x + 1)$	$P(X \ge x + 1 - 0/5) = P(X \ge x + 0/5)$
$P(x_1 \le X \le x_2)$	$P(x_1 - 0/5 \le X \le x_2 + 0/5)$

مثال– استادی به تجربه دریافته است که 45 درصد از دانشجویانش در درس خاصی نمرهٔ «ب» میگیرند. اگر در ترم جاری 25 نفر این درس را با او گرفته باشند، احتمال گرفتن نمرهٔ «ب»را برای این تعداد محاسبه کنید:

الف) حداقل 10 نفر ، ب) بین 8 تا 20 نفر، ج) دقیقا 10 نفر

$np = 11/25 , nq = 13/75$

بنابراین می‌توان از تقریب نرمال استفاده نمود.

$\mu = np = 11/25 \qquad , \sigma = \sqrt{npq} = \sqrt{6/1875} = 2/487$

الف)

$$P_{\text{دو جمله‌ای}}(X \ge 10) = P_{\text{نرمال}}(X \ge 9/5) = P(Z \ge -0/7) = 1 - P(Z \le -0/7) = 1 - 0/242 = 0/758$$

ب)

$$P_{\text{دو جمله‌ای}}(8 < X < 20) = P_{\text{نرمال}}(8/5 \le X \le 19/5) = P(-1/1 \le Z \le 3/317)$$
$$= 0/9995 - 0/1357 = 0/8638$$

ج)

$$P_{\text{دو جمله‌ای}}(X = 10) = P_{\text{نرمال}}(9/5 \le X \le 10/5) = P(-0/7 \le Z \le -0/3)$$
$$= 0/3821 - 0/242 = 0/14$$

9-5تقریب توزیع پواسون به وسیلهٔ توزیع نرمال

وقتی میانگین توزیع پواسون($\boldsymbol{\lambda}$) نسبتا بزرگ شود می‌توان تقریب نرمال را برای توزیع پواسون بکار برد. هر چقدر $\boldsymbol{\lambda}$ افزایش می‌یابد توزیع پواسون به توزیع نرمال نزدیکتر می‌شود. بطورکلی اگر $\boldsymbol{\lambda \geq 10}$ باشد، تقریب نرمال تقریب خوبی برای پواسون خواهد بود. در اینصورت میانگین و انحراف معیار توزیع نرمال برابر $= \boldsymbol{\lambda}$ و $= \sqrt{\boldsymbol{\lambda}}$ خواهد بود.

در هنگام حل مسائل، در استفاده از توزیع نرمال به جای پواسون نیز، به دلیل گسسته بودن توزیع پواسون و پیوسته بودن توزیع نرمال، باید از تصحیح پیوستگی- همانند تقریب دو جمله‌ای به وسیلهٔ نرمال- استفاده کرد.

مثال– در خلال ساعت 4 تا 7 بعد از ظهر، دورهٔ اوج کاری، به طور متوسط در هر 5 دقیقه یک ماشین جهت تعمیر به تعمیرگاهی مراجعه می‌کند. احتمال اینکه بین ساعت 5 تا 6/25 بیش از 12 ماشین به این تعمیرگاه مراجعه کند، چقدر است؟

بر اساس داده‌های مسئله، در فاصلهٔ زمانی5 تا 6/25 دقیقه، بطور متوسط 17 ماشین وارد تعمیرگاه خواهند شد؛پس $\boldsymbol{\lambda t = 17}$ بوده و می‌توان از تقریب نرمال با پارامترهای $= \mathbf{17}$ و $= \sqrt{\mathbf{17}} = 4/1231$ استفاده کرد.

$$P_{\text{پواسون}}(X > 12) = P_{\text{نرمال}}(X \geq 12/5) = P(Z \geq -1/09) = 1 - P(Z < -1/09) = 1 - 0/1379 = 0/8621$$

10-5قضیهٔ حد مرکزی

دو موضوع در قضیهٔ حد مرکزی مطرح می‌شود؛ موضوع اول جمع $\boldsymbol{n}$ متغیر تصادفی مستقل و موضوع دوم توزیع‌های نمونه‌گیری است.توزیع نرمال تقریب خوبی برای توزیع مجموع متغیرهای تصادفی است. اگر $\mathbf{X_1}$، $\boldsymbol{X_2}$، ...، $\boldsymbol{X_n}$، $\mathbf{n}$ متغیر تصادفی مستقل باشند، مشروط بر آنکه $\mathbf{n}$ به اندازهٔ کافی بزرگ باشد، آنگاه متغیر تصادفی $\mathbf{Y=X_1+\ X_2+...+X_n}$دارای توزیع نرمال با $\boldsymbol{\mu_Y} = \sum_{i=1}^{n} \boldsymbol{\mu_{X_i}}$و $\boldsymbol{\sigma_Y^2} = \sum_{i=1}^{n} \boldsymbol{\sigma_{X_i}^2}$ خواهد بودکه $\boldsymbol{\mu_{X_i}}$ و $\boldsymbol{\sigma_{X_i}^2}$ به ترتیب میانگین و واریانس متغیر تصادفی $\boldsymbol{X_i}$ هستند. ممکن است این سوال مطرح شود که اندازهٔ $\boldsymbol{n}$ چقدر باید باشد تا بتوان از توزیع نرمال برای جمع $\boldsymbol{n}$ متغیر تصادفی استفاده کرد. اگر توزیع هر جملهٔ $\mathbf{X_i}$ چولگی زیادی نداشته باشد و اگر محتمل نباشد که یک جمله سهم عمده و تعیین‌کننده‌ای از مجموع را داشته باشد، توزیع مجموع حداقل 10 متغیر تصادفی ($\boldsymbol{n}$≥10) تقریبا توزیع نرمال خواهد بود.

این نکتهکه $\mathbf{X_i}$ها می‌توانند هر توزیعی داشته باشند، اهمیت توزیع نرمال را نشان می‌دهد. مثالهایی عملی این قضیه ممکن است زمانبندی پروژه‌ها که از فعالیتهای مختلفی تشکیل شده‌اند و یا وزن محموله‌ایی که وزن هر جزء از آنها دارای توزیع خاصی است، باشد.

تست ها ی فصل پنجم

۱- به چند طریق می توان کمیته ای شامل ۳ نفر از یک گروه ۵ نفری انتخاب کرد؟

الف) 10   ب) 11   ج) 12   د) 13

۲- به یک حذف دو نفر تیراندازی می کنند. احتمال اینکه نفر اول به هدف بزند، ۶٪ و احتمال اینکه نفـر دوم بزند، ۷٪ می باشد، احتمال اینکه لااقل یکی از آنها نیز به هدف بزند، کدام است؟

الف) $P(B \cup A) = \%18$ ب) $P(B \cup A) = \%28$

ج) $P(B \cup A) = \%42$ د) $P(B \cup A) = \%88$

۳- می خواهیم از بین صد هزار نفر دانشجوی یک دانشگاه که تقریباً نیمی از آن ها پسر و نیمـی دختـر هسـتند شش نفر را به طور تصادفی انتخاب کنیم. احتمال اینکه چهار نفر از این شش نفر دختر باشد چقدر است؟

الف) $\frac{5}{64}$ ب) $\frac{10}{64}$ ج) $\frac{15}{64}$ د) $\frac{20}{64}$

۴- میزان احتمال نمره صفر در ۴ سؤال صحیح و غلط برای یک فرد چقدر است؟

الف) $\frac{1}{16}$ ب) $\frac{1}{8}$ ج) $\frac{1}{4}$ د) $\frac{1}{2}$

۵- اگر A و B دو پیشامد، $p(A) = \frac{1}{14}$ و $P(B \cup A) = \frac{1}{3}$ و A زیر مجموعه B باشـد، در ایـن صـورت $P(B)$ چقدر است؟

الف) $\frac{1}{12}$ ب) $\frac{1}{4}$ ج) $\frac{1}{3}$ د) $\frac{7}{12}$

۶- ۷۰٪ تولیدات کارخانه ای سالم است. ۲۰ نمونه از این تولیدات را بطور تصادفی انتخاب می کنـیم احتمـال آن که حداقل یکی سالم باشد کدام است؟

الف) 7٪   ب) 13٪   ج) 39٪   د) 1/92

۷- با برداشتن تصادفی ۲ لامپ از بین ۳ لامپ عیوب و ۶ لامپ سالم، احتمال سالم بودن هر دو لامپ چقدر است؟

الف) $\frac{1}{16}$ ب) $\frac{5}{12}$ ج) $\frac{7}{12}$d د) $\frac{5}{6}$

۸- اگر A و B دو حادثه باشند و احتمال متناظر آن ۲٪ = P(A) و ۶٪ = P(B) و همچنین ۷٪ (A یا B)P باشد(A و B)P کدام است؟

الف) 1٪ ب) 3٪ ج) 4٪ د) 5٪

۹- سکه سالمی را ۶۴ بار پرتاب می کنیم، انحراف معیار آمدن یک روی سکه کدام است؟

الف) 4 ب) 8 ج) 16 د) 32

۱۰- احتمال موفقیت یک عمل جراحی $\frac{4}{5}$ است. احتمال آنکه در ۵ مورد از این عمل، ۴ مورد آن موفقیت آمیز باشد چقدر است؟

الف) $\frac{4}{5}$ ب) $\left\{\frac{4}{5}\right\}^4$ ج) $\frac{16}{25}$ د) $\frac{64}{125}$

۱۱- در یک آزمون تستی چهار گزینه ای ، داوطلبی چهار سؤال را بطور تصادفی پاسخ میدهد. احتمال آنکه هر چهار سؤال را غلط پاسخ دهد کدام است ؟

الف) $\frac{1}{256}$ ب) $\frac{9}{256}$

ج) $\frac{27}{256}$ د) $\frac{81}{256}$

۱۲- دانشجویی سه سؤال ۴ جوابی را بطور تصادفی علامت می زند . احتمال اینکه حداقل یک جواب درست باشد برابر است با :

الف) $\frac{1}{64}$ ب) $\frac{63}{64}$ ج) $\frac{27}{64}$ د) $\frac{37}{64}$

۱۳- اگر دانش آموزی از روش شانس و حدس کورکورانه ، به ۳ سؤال صحیح – غلط یک آزمون پاسخ دهـد ، احتمال اینکه به حداقل ۲ سوال پاسخ صحیح بدهد، چند درصد است؟

الف) 12/5 ب) 37/5

ج) 50 د) 66

۱۴- دو پیشامد B,A مستقل اند . اگر $P(A \mathbf{I} B)=\frac{1}{8}$ ، $P(A)=\frac{1}{2}$ باشد ، P(B) کدام است؟

الف) $\frac{1}{4}$ ب) $\frac{1}{16}$

ج) $\frac{3}{8}$ د) $\frac{5}{8}$

۱۵- در دو جمله ای $(P+l)^5$ ضریب دو جمله ای $P^3 l^2$ (جمله سوم) را محاسبه کنید.

الف) 5 ب) 6

ج) 15 د) 10

۱۶- از کیسه ای محتوای ۵ مهره سفید و چهار مهره سیاه ، ۳ مهره به تصادف خارج میکنیم، احتمـال آنکـه مهره ها همرنگ باشد، چقدر است ؟

الف) $\dfrac{\begin{bmatrix}5\\3\end{bmatrix}\mathbf{m}\begin{bmatrix}4\\3\end{bmatrix}}{\begin{bmatrix}9\\3\end{bmatrix}}$ ب) $\dfrac{\begin{bmatrix}5\\3\end{bmatrix}}{\begin{bmatrix}9\\3\end{bmatrix}}$

ج) $\dfrac{\begin{bmatrix}4\\3\end{bmatrix}}{\begin{bmatrix}9\\3\end{bmatrix}}$ د) $\dfrac{\begin{bmatrix}5\\3\end{bmatrix}\begin{bmatrix}4\\3\end{bmatrix}}{\begin{bmatrix}9\\3\end{bmatrix}}$

۱۷- در مقایسه توزیع دو جمله ای با توزیع طبیعی ، اگر $p=\frac{1}{2}$ باشد، کدامیک از گزینه های زیر درست مــی باشد ؟

الف) هر دو توزیع هر مقداری را در فاصله $\pm\infty$ می پذیرند .

ب) هر دو توزیع متقارن هستند .

ج) هر دو توزیع کشیدگی مثبت دارند .

د) هر دو توزیع کشیدگی منفی دارند.

۱۸- خانواده ای دارای ۴ پسر و ۱ دختر است ، احتمال اینکه فرزند بعدی این خانواده پسر باشد چقدر است ؟

الف) $\frac{4}{5}$　　ب) $\frac{1}{5}$

ج) $\frac{1}{2}$　　د) $\frac{1}{4}$

۱۹- در نشستن ۸ نفر روی ۸ صندلی ، تعداد ترتیب های ممکن نشستن چقدر است؟

الف) 40230　　ب) 32040

ج) 40320　　د) 23004

۲۰- به چند طریق می توان کمیته ای شامل ۳ نفر از یک گروه ۵ نفری انتخاب کرد؟

الف) 10　　ب) 8

ج) 12　　د) 9

پاسخنامه فصل پنجم

1- گزینه الف

$$C^n_r=\frac{n!}{r!(n-r)!}\ C_3^5=\frac{5!}{3!(5-3)!}=\frac{5\times4\times3!}{3!2!}=10$$

2- گزینه د

3- گزینه ج

$$\text{احتمال}:\frac{ck}{n}\times(p^k\times1-p)=\frac{15}{64}$$

4- گزینه الف

5- گزینه الف

$$P(B\grave{E}A)\ P\sim(A)=P(B)$$

1/3-1/4=1/12=1/12

6- گزینه د

$$\left[c_2^1\left[\frac{7}{10}\right]^1\left[\frac{3}{10}\right]^1\right]+\left[c_2^2\left[\frac{7}{10}\right]^2\left[\frac{3}{10}\right]\right]=\frac{42}{100}+\frac{49}{100}=0/91$$

7- گزینه د

8- گزینه الف

B) **I A= P(A)+P(B) – P(I)B U P(A**

B) = 0/1 $\longrightarrow$

0/7 = 0/2 + 0/6- **(APB)** **I (AP I**

9- گزینه الف

$$d=\sqrt{\frac{n}{4}}=\sqrt{\frac{64}{4}}=4$$

10- گزینه ب

$$n=5\qquad x=4\qquad p=\frac{4}{5}\qquad 1-p=1-\frac{4}{5}=\frac{1}{5}$$

$$c^4_5\left[\frac{4}{5}\right]^4\left[\frac{1}{5}\right]^1=5\left[\frac{4}{5}\right]^4\left[\frac{1}{5}\right]=\left[\frac{4}{5}\right]^4$$

11- گزینه الف

$$P=\frac{1}{4}\qquad l=\frac{3}{4}\qquad n=4\qquad X=4$$

$$P(X)=C_n^2P^Xg^{n-x}=\frac{n!}{r!(n-r)}.P^X l^{n-x}=\frac{4!}{4!(0)!}\times\frac{14}{3}\times\frac{3}{4}=\frac{1}{256}$$

12- گزینه د

$$\left(\frac{1}{4}+\frac{3}{4}\right)^3 = c_3^0\frac{1}{4}^0\left(\frac{3}{4}\right)^3 + c_3^1\frac{1}{4}\left(\frac{3}{4}\right)^2 + c_3^2\frac{1}{4}^2\left(\frac{3}{4}\right) + c_3^3\frac{1}{4}^3\left(\frac{3}{4}\right)^0$$

$$\Rightarrow \frac{3!}{1!2!}\frac{1}{4}\left(\frac{3}{4}\right)^2$$

$$\frac{3!}{2!(3-2)!}\frac{1}{4}^2\left(\frac{3}{4}\right)+\frac{13}{4} \Rightarrow \frac{27}{64}+\frac{9}{64}+\frac{1}{64}=\frac{37}{64}$$

13- گزینه ب

$$C_3^2\frac{1}{2}^3 = \frac{3!}{2!1!}\times\frac{1}{8} = 0/375 \Rightarrow 37/5\%$$

14- گزینه الف

$$P(A \cap B) = P(A).P(B)$$

$$\frac{1}{8}=\frac{1}{2}.P(B) \Rightarrow P(B)=\frac{1}{4}$$

15- گزینه د

$$C_n^n = \frac{n!}{r!(n-r!)} = \frac{5!}{3!(5-3)!} = 10$$

16- گزینه الف

17- گزینه ب

18- گزینه ج

19- گزینه ج

$$8! = 8\times7\times6\times5\times4\times3\times2\times1 = 40320$$

20- گزینه الف

$$C_r^n = \frac{n!}{rI(n-r)!} \Rightarrow C_3^5 = \frac{5!}{3!(5-3)!} = \frac{5\times4\times3!}{3!2!} = 10$$

فصل ششم : رگرسیون[1] و همبستگی[2]

همبستگی به روابط بین دو یا چند متغیر که قابل تبدیل به یکدیگر هستند، اطلاق می شود. هدف از مطالعه رابطـه بـین دو متغیر نه تنها مستلزم روش هایی برای تعریف و اندازه گیری متغیرهای تحت مطالعه اسـت، بلکـه تعیـین روش هـای آماری برای توصیف ماهیت روابط بین متغیرها را نیز در بر می گیرد. هدف از مطالعه ی رابطه ی بین دو متغیر $\mathbf{X}$ و $\mathbf{Y}$ نه تنها مستلزم روش هایی برای تعریف و اندازه گیری متغیرهای تحت مطالعه اسـت، بلکـه تعیـین روشـهای آمـاری بـرای توصیف ماهیت روابط بین متغیرها را نیز دربر می گیرد. در بعضی موقعیتها برآورد یـا پـیش بینـی یـک متغیـر براسـاس اطلاعات موجود از یک متغیر دیگر مورد توجه است .

از طرف دیگر امکان دارد که میزان یا مقدار رابطه بین دو متغیر یعنی میزان تغییر همایند مـورد توجـه باشـد. شـاخص آماری که میزان و حدود رابطه بین متغیرها را نشان می دهد، ضریب همبسـتگی نامیـده مـی شـود. ضـریب همبسـتگی تعیین کننده شدت و جهت همبستگی بین دو متغیر است. جهت همبستگی توسط علامت ضریب همبسـتگی[3] و شـدت همبستگی به وسیلهٔ قدر مطلق ضریب همبستگی مشخص می شود.

هنگامی که افزایش در یک متغیر با افزایش در متغیر دیگر یا کاهش یک متغیر با کاهش یک متغیر دیگـر همـراه باشـد، همبستگی بین دو متغیر مستقیم و مثبت است و چنانچه افزایش در یک متغیر با کـاهش درمتغیـر دیگـر همـراه باشـد، همبستگی بین دو متغیر منفی و معکوس است.

چنانچه رابطه بین متغیرها به صورت یک خط مستقیم باشد این همبستگی را خطی می نامند.

اما اگر رابطه به صورت خطی نباشد، بلکه شبیه منحنی باشد، این همبستگی را غیر خطی می نامند.

- نکته: شدت همبستگی مستقل از علامت ضریب همبستگی است.

ضریب همبستگی با علامت $\mathbf{r}$ نشان داده می شود و دامنه ارزش های آن از 1- تا 1+ می باشد. ضریب 1+ نشـان دهنـدهٔ همبستگی مثبت و کامل و ضریب همبستگی 1- نشان دهندهٔ همبستگی منفی و کامل است، یعنی بین دو متغیر رابطـه معکوس وجود دارد. چنانچه بین متغیرها همبستگی وجود نداشته باشد، ضریب همبستگی مساوی صفر خواهد بود.

---

[1] Regression

[2] Correlati

[3] Correlation Coefficient

- نکته: برای بررسی رابطهٔ بین دو متغیر از همبستگی و برای مطالعه بیش از دو متغیر از روش های آماری چند متغیری استفاده می شود.

نمودارهای پراکندگی

یکی از روش هایی که به وسیلهٔ آنها می توان همبستگی بین دو متغیر را نشان داد، نمودار پراکندگی است. این نمودار نوعی نمایش ترسیمی است که از طریق آن ارزش های دو متغیر نشان داده می شود.

محاسبه ضریب همبستگی با استفاده از نمره های انحرافی

برای محاسبه ضریب همبستگی با استفاده از نمره های انحرافی از فرمول زیر استفاده می شود:

$$\mathbf{r}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x^2-\bar{x})(y^2-\bar{y})}}=\frac{\sum xy}{\sqrt{\sum x^2\sum y^2}}$$

ضریب همبستگی گشتاوری پیرسون

متداولترین شاخص مورد استفاده در محاسبه ضریب همبستگی، ضریب همبستگی گشتاوری پیرسون[4] است. هنگامی که داده ها دارای مقیاس فاصله ای و یا نسبی هستند، ضریب همبستگی بین دو متغیر را می توان از طریق محاسبه ضریب همبستگی گشتاوری پیرسون محاسبه کرد. برای محاسبه ضریب همبستگی پیرسون فرمول های مختلفی وجود دارد . ساده ترین راه محاسبه **r**، هنگامی که داده ها کم هستند و نیاز به طبقه بندی نیست از فرمول زیر استفاده می شود :

$$\mathbf{r_{xy}}=\frac{n\sum XY-(\sum X)(\sum Y)}{\sqrt{\left[n\sum X^2-(\sum x)^2\right]\left[n\sum Y^2-(\sum Y)\right]}}$$

فرض هایی درباره ضریب همبستگی پیرسون

ضریب همبستگی پیرسون که یک شاخص معتبر برای تعیین رابطه بین متغیرهاست، دارای مفروضه های زیر است :

1- رابطه بین دو متغیر خطی باشد : منظور از رابطه خطی ، رابطه ای است که نمودار پراکندگی آن بصورت خط باشد. ضریب همبستگی پیرسون فقط برای توصیف همبستگی خطی مناسب است. مشاهده نمودار پراکندگی ساده ترین روشی است که میتوان از طریق آن خطی یا غیر خطی بودن رابطه بین متغیرها را تعیین کرد.

2- توزیع ها دارای شکل های مشابه باشند : چنانچه بین شکل متغیرهای مورد پژوهش تفاوت زیادی وجود داشته باشد، نمی توان از ضریب همبستگی پیرسون استفاده کرد.

---

[4] . Pearson product moment correlation coefficient

3- نمودار پراکندگی یکسان باشد : این فرض به فرض نقاط پراکندگی ارتباط دارد. به این معنا که باید عرض نقاط در سراسر نمودار یکسان باشد. این فرض را یکسانی نقاط پراکندگی میگویند.

مثال: فرض کنید دو آزمون متفاوت، یکی آزمون هوش (**x**) و دیگری آزمون استعداد ریاضی (**Y**) روی گروهی از دانش آموزان اجرا شده و نمره های زیر بدست آمده است. می خواهیم بدانیم آیا بین هوش و استعداد ریاضی رابطه وجود دارد و این رابطه یا همبستگی چه مقدار است؟

	X	Y	XY	X^2	Y^2
1	120	31	3720	14400	961
2	112	25	2855	12544	625
3	110	20	2200	12100	400
4	123	24	2952	15129	576
5	105	18	1890	11025	324
N=5	507	118	13617	65198	2886

$$\mathbf{r_{xy}} = \frac{5\times 13617-(507)(118)}{\sqrt{\left[5\times 65198-(507)^2\right]\left[5\times 2886-(118)^2\right]}} = \frac{8259}{5906/28} = 0/39$$

(بین دو متغیر هوش و استعداد ریاضی همبستگی مثبت و کامل وجود دارد .)

ضریب تعیین

پس از محاسبه ضریب همبستگی می توان ضریب تعیین ر ابا فرمول $\mathbf{cd}=\mathbf{r}^2\times 100$ برآورد کرد.

ضریب تعیین نشان می دهد که چند درصد واریانس متغیر **Y** (وابسته) را می توان به وسیله متغیر **x**(مستقل) تبیین کرد. برای محاسبه ضریب همبستگی از فرمول های دیگری نیز استفاده می شود که به شرح زیر می باشد:

1- محاسبه **r** از روی نمره های انحرافی با استفاده از فرمول :

$$r = \frac{\sum xy}{ns_x s_y}$$

در این فرمول **x** و**y** نمره های انحرافی و $\mathbf{s_x}$ و $\mathbf{s_y}$انحراف معیار مربوط به متغیرهای **x** و **y** هستند.

2- محاسبه **r** از روی تفاوت بین زوج انحراف ها با استفاده از فرمول:

$$r = \frac{s^2x + s^2y - s^2d}{2s_x s_y}$$

در این فرمول s^2_y و s^2_x واریانس نمره هـای x و y، s_y و s_x انحـراف اسـتانداردهای متغیرهـای مـذکور هسـتند و s^2d واریانس تفاوت بین نمره های انحرافی x و y است.

3- r را روی نمره های z (تراز شده) و با استفاده از فرمول:

$$r = \frac{\sum z_x z_y}{n} \Rightarrow z_x = \frac{x}{s_x}, z_y = \frac{Y}{s_y}$$

ضریب همبستگی اسپیرمن (sparman)

در مواردی که داده ها دارای مقیاس رتبه ای باشند، برای محاسبه ضریب همبستگی از ضریب رتبه ای اسپیرمن استفاده می شود. ضریب همبستگی اسپیرمن را با r_s یا با حرف لاتین r نشان می دهند و فرمول محاسبـهٔ آن بـه صـورت زیـر است:

$$r_s \text{ یا } r = 1 - \frac{6\sum D}{n^2(n-1)}$$

در این فرمول D^2 مجذور تفاوت رتبه های دو متغیر X و Y است.

* مثال: فرض کنید می خواهیم ضریب همبستگی دو دسته نمـره (هـوش و اسـتعداد ریاضـی) را کـه در تمـرین پـیش داشتیم از روش اسپیرمن محاسبه کنیم.

	X	Y	رتبه X	رتبه Y	D	D^2
1	120	31	2	1	1	1
2	112	25	3	2	1	1
3	110	20	4	4	0	0
4	123	24	1	3	-2	4
5	105	18	5	5	0	0
$\sum n = 5$					0	6

$$r_s = 1 - \frac{6 \times 6}{5(25-1)} = 1 - \frac{36}{120} = 0/7$$

ضریب همبستگی فای یا فی (j) (phi)

برای محاسبه همبستگی دو متغیر که هر دو دو ارزش یا دو وجهی هسـتند از ضـریب همبسـتگی اسـتفاده مـی شـود. همچنین زمانی که داده ها دارای مقیاس اسمی باشند. رابطه بین متغیرهای دو وجهی به صورت یک جدول 2×2 نشـان داده می شود و فراوانی پاسخ های درست آزمودنی ها به هریک از متغیرها در خانه های جدول نوشته می شود.

ضریب فای از فرمول زیر محاسبه می شود:

$$j = \frac{AB - BC}{\sqrt{[A+B][C+D][A+C][B+D]}}$$

A	**B**
C	**D**

ضریب همبستگی دو رشته ای نقطه ای (rpbis)

وقتی داده ها به نحوی باشند که یک متغیر از کمیت های پیوسته و متغیر دیگر از کمیت هـای ناپیوسـته (دو وجهـی) برخوردار باشد از ضریب همبستگی دو رشته ای نقطه ای استفاده می کنیم . برای مثـال اگـر بخـواهیم همبسـتگی یـک سؤال را با کل آزمون بسنجیم از این ضریب استفاده می کنیم . (سؤال یک متغیر دو ارزشی ساختگی و کل آزمون یـک متغیر پیوسته است.) فرمول محاسبه به صورت زیر است:

$$\mathbf{r_{pb}} = \frac{M_1 - M_0}{S_X}\sqrt{pq}$$

در این فرمول $\mathbf{M}_1$ میانگین **x** های گروهی است که در متغیر دو وجهی نمره 1 دارند. $\mathbf{M}$. میانگین **x** های گروهی اسـت که نمره آنها دو متغیر دو وجهی است ، $\mathbf{s_x}$ انحراف معیار نمره های **x** است و **p** نسـبت افـرادی اسـت کـه در متغیـر دو وجهی 1 گرفته و **q** نسبت افرادی است که نمره آنها صفر است.

ضریب همبستگی دو رشته ای (bis)

شاخصی است برای محاسبه رابطه بین دو متغیر که توزیع پیوسته دارند ولی یکی از متغیرها دو ارزشی سـاختگی اسـت. فرمول محاسبه چنین است:

$$\mathbf{b_r} = \frac{MP - Mx}{SX}\left(\frac{px}{Y}\right)$$

در این فرمول $\mathbf{M_P}$ میانگین نمره های خام کسانی است که در متغیر دو وجهی نمره قبولی گرفته اند،$\mathbf{M_q}$میـانگین نمـره های خام افرادی است که در متغیر دو وجهی مردود شده اند، **p** نسبت به آزمودنی های اسـت کـه در متغیـر دو وجهـی

قبول شده اند و **q** نسبت آزمودنیهای است که در متغیر دو وجهی مردود شده اند، S_x انحراف معیار متغیر **X** و **Y** عرض نقطه تکفیک در منحنی نرمال است که از جدول بدست می آید.

- تفاوت ضریب همبستگی دو رشته ای با ضریب همبستگی دو رشته ای نقطه ای

شرط استفاده از ضریب همبستگی دو رشته ای قبول فرضیه بهنجار بودن توزیع نمره ها هم در متغیر پیوسته و هم در متغیر دو وجهی است، در صورتی که ضریب همبستگی دو رشته ای نقطه ای مستلزم قبول فرضیه بهنجار بودن توزیع نمره های متغیرهای موردنظر نیست.

ضریب همبستگی تتراکوریک

از این ضریب همبستگی زمانی استفاده می شود که هر دو داده یا متغیر اسمی باشند و به صورت دو ارزشی ساختگی عنوان گردند.

ضریب همبستگی ایتا

ضریب ایتا یک ضریب همبستگی است که از آن برای تعیین خطی یا غیرخطی بودن همبستگی استفاده می شود.

ضریب همبستگی توافقی پیرسون (C)

ضریب همبستگی توافقی پیرسون **(C)** در جداول دو بعدی (2×2) بکار می رود. این ضریب متقارن است و در سطح سنجش اسمی- اسمی کاربرد دارد. فرمول محاسبه ان به صورت زیر است:

$$C_{\max} = \sqrt{\frac{c^2}{1+c^2}}$$

ضریب همبستگی یول (Q)

این ضریب نیز در جداول دو بعدی بکار می رود و متقارن بوده و در سطح سنجش اسمی- اسمی کاربرد دارد.

ضریب همبستگی لاندا

ضریب همبستگی لاندا پیش بینی کننده است و در سطح سنجش اسمی- اسمی کاربرد دارد. دامنه این ضریب بین صفر تا 1 بوده و از نوع کاهش نسبی خطاست. این شاخص بیانگر آن است که با آگاهی از مقادیر مستقل میزان دقت، پیشگویی متغیر وابسته چقدر است.

- نکته: اگر لاندا صفر باشد متغیر وابسته قابل پیشگویی نیست.
- نکته: لاندا تفسیر روشنی از رابطه بین دو متغیر را نشان می دهد.

ضریب همبستگی تای وای گودمن وکراسکال

ضریب همبستگی تای وای گودمن وکراسکال که برای متغیرهای اسمی بکار می رود و نامتقارن است شدت گرایش ها را می سنجد و دامنه آن بین صفر تا 1 است.

ضریب توافق چوپر یا آزمون نسبت ها

ضریب توافق چوپر و یا آزمون نسبت ها در سطح سنجش اسمی- ترتیبی کاربرد دارد و متقارن می باشد. دامنه تغییرات آن بین صفر تا 1+ است.

ضریب همبستگی کرامرز (V)

ضریب کرامرز در سطح سنجش اسمی- اسمی کاربرد دارد و متقارن است و برای جداول مستطیلی استفاده می شود. دامنه تغییرات این ضریب بین صفر تا 1+ بوده و در رگرسیون دو متغیر همبستگی، دو متغیر مستقل بر روی هم هیچ ارتباطی با جمع مجذور همبستگی های ساده متغیرهای مستقل ندارد.

عوامل مؤثر بر ضریب همبستگی

1- پراکندگی متغیرها در جوامع مختلف متفاوت است.
2- همبستگی بین دو متغیر میتواند تحت تاثیر یک متغیر سوم قرار گیرد.
3- رابطه اصلی بین متغیرها از یک جامعه به جامعه ای دیگر فرق می کند.
4- همبستگی بین دو متغیر در یک جامعه ناهمگن بیشتر از همان همبستگی در یک جامعه همگن است.

تفسیر ضریب همبستگی

ضریب همبستگی جهت و شدت رابطه بین دو متغیر را تعیین می کند . اگر چه این ضریب به صورت اعشار بیان می شود، ولی تفسیر آن نباید بر حسب درصد صورت گیرد. ضریب همبستگی 0/70 به معنای آن نیست که هفتاد درصد از رابطه بین متغیرهای Y,X تبیین می شود. ضریب همبستگی را نمی توان به صورت نسبت مورد مقایسه و تفسیر قرار داد.

کوواریانس

کوواریانس تغییرات مشترک دو متغیر X و Y را نشان می دهد و به عبارت دیگر این موضوع را بیان می کند که چه میزان از تغییرات X و Y در اشتراک با یکدیگر می باشند و تعریف آن عبارت است" از میانگین حاصل ضربهای نمره ها منهای حاصلضرب میانگین ها." مقدار کوواریانس تابع واحد اندازه گیری است و مقدار آن با کوچک یا بزرگترشدن واحد اندازه گیری تغییر می کند. اگر کوواریانس را بر حاصلضرب S_y و S_x تقسیم کنیم، ضریب همبستگی پیرسون بدست می آید.

فرمول محاسبه کوواریانس بصورت زیر است:

$$\textbf{Cov} = \overline{X}.\overline{Y} - (\overline{x.y})$$

$$r = \frac{\overline{x}.\overline{y} - \overline{x.y}}{S_x \ S_y}$$

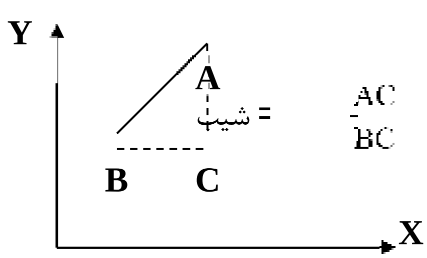
رگرسیون

همبستگی و رگرسیون (پیش بینی) در آمار به اندازه ای به یکدیگر نزدیکند که گاهی یکی را به جای دیگری بکار می برند. همبستگی به معنای درجه ارتباط دو متغیر با یکدیگر است. هرگاه یکی از دو متغیری را که با یکدیگر همبستگی دارند درست داشته باشیم، می توانیم دیگری را از روی آن پیش بینی کنیم. با استفاده از خط رگرسیون و یا معادله آن می توان یک متغیر را از روی متغیر دیگر پیش بینی کرد. اگر همبستگی بین متغیرها بالا باشد رگرسیون کمتر اتفاق می افتد و اگر همبستگی بین متغیرها صفر باشد، رگرسیون در اطراف میانگین به صورت کامل اتفاق خواهد افتاد. به طورکلی رگرسیون به طرف میانگین با همبستگی بین متغیرها رابطه معکوس دارد.

رگرسیون خطی

پژوهشگر برای اینکه بتواند از روی متغیر مستقل **X** متغیر وابسته **Y** را پیش بینی کند، باید ضریب همبستگی بین آن ها را در مورد یک گروه نمونه آماری محاسبه کرده باشد. یکی از فرضیه های پذیرفته شده در فرایند پیش بینی این است که رابطه میان دو متغیر یک رابطه خطی است و لذا این رابطه را میتوان با یک نمودار خطی نشان داد. خط رگرسیون خطی است که مجذور خطاهای پیش بینی را به حداقل برساند. هر نمودار دارای یک معادله است که با فرمول کلی $\hat{Y} = bx + a$ مشخص می شود.

در این فرمول **a** یک مقدار ثابت است و فاصله محور **y** از مبدأ تا نقطه ای را نشان می دهد که محور **y** را قطع می کند و کمیت **b** شیب این خط را نشان می دهد. شیب خط نسبت فاصله محور عمودی به فاصله محور افقی است. این شیب میزان افزایش در **Y** را با زیاد شدن **X** نشان می دهد.



در فرمول معادله خط رگرسیون برای محاسبه **a** و **b** از فرمول های زیر استفاده می شود:

$$\mathbf{b_{xy}} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{\sum xy - n\overline{xy}}{\sum x^2 - n\bar{x}^2}$$

$$\mathbf{a_{yx}} = \frac{\sum Y - byx \times \sum x}{n} = \bar{Y} - \mathbf{b_{yx}} \cdot \bar{X}$$

خطای پیش بینی

اختلاف بین نمره مشاهده شده (**y**) و نمره پیش بینی شده ($\bar{y}$) را خطای پیش بینی می نامند. خطای پیش بینی با همبستگی رابطه معکوس دارد. این خطا برای هر آزمودنی از فرمول زیر به دست می آید:

$$\mathbf{e} = \mathbf{Y} - \bar{Y}$$

خطای معیار برآورد

اگر محقق بدون هیچ گونه اطلاعاتی درباره متغیر مستقل بخواهد متغیر وابسته ای را پیش بینی کند، بهترین پیش بینی برای **Y** میانگین نمره های **Y** است. زیرا اگر توزیع نمره ها یک توزیع طبیعی باشد، نمره های 68 درصد افراد مورد مطالعه بین میانگین و± 1 انحراف معیار قرار می گیرد. برای تعیین خطای استاندارد برآورد می توان از فرمول زیر استفاده کرد:

$$\mathbf{SE_{est}} = \sqrt{\frac{\Sigma(\mathrm{Y}-\bar{\mathrm{Y}})^2}{\mathrm{N}-2}} = \sqrt{\frac{\sum \mathrm{e}^2}{\mathrm{N}-2}}$$

معمولاً برای سرعت سهولت محاسبه، خطای استاندارد برآورد را از فرمول زیر که از نقطه نظر ریاضی با فرمول بالا معادل است بدست می آورند :

$$\mathbf{SE_{set}} = \mathbf{Sy}\ \sqrt{1-\mathrm{r}\times\mathrm{y}^2}$$

در فرمول بالا مقدار **Sy** از این فرمول محاسبه می شود:

$$\frac{\sqrt{\sum \mathrm{Y}^2 - \frac{(\sum \mathrm{Y})^2}{\mathrm{N}}}}{\mathrm{N}-1}$$

* مثال: نمره های زیر نتایج یک آزمون هوش (متغیر مستقل **X**) و نمره های آزمون پیشرفت تحصیلی (متغیر وابسته **Y**) را در مورد یک گروه نمونه 10 نفری نشان می دهد. معادله خط رگرسیون را برای پیش بینی اندازه های **Y** از روی اندازه های **X** بنویسید و خطای معیار برآورد را نیز محاسبه کنید.

										Σ	
X	1	2	2	3	4	6	8	9	10	5	50
Y	2	2	3	5	5	5	6	7	8	7	50
X^2	1	4	4	9	16	36	64	81	100	25	340
Y^2	4	4	9	25	25	25	36	49	64	49	290
XY	2	4	6	15	20	30	48	63	80	35	303

$$\bar{x} = \frac{\sum x}{N} \longrightarrow \frac{5}{10} = 5$$

$$s_x = \sqrt{\frac{N\sum x^2 - (\sum x)}{N(N-1)}} - \longrightarrow \sqrt{\frac{10(340) - (50)^2}{10(10-1)}} = 3/16$$

$$\bar{Y} = \frac{\sum Y}{N} \longrightarrow \frac{50}{10} = 5$$

$$s_Y = \sqrt{\frac{N\sum Y^2 - (\sum Y)^2}{N(N-1)}} \longrightarrow \sqrt{\frac{10(290) - (50)^2}{10(10-1)}} = 2/11$$

$$r_x y = \frac{N\sum xY - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}} \rightarrow \frac{10(303) - 50(50)}{\sqrt{\left[10(340) - (50)^2\right]\left[10(290) - (50)^2\right]}} = 0/88$$

$$\mathbf{b} = \frac{N\sum xY - (\sum x)(\sum Y)}{N\sum x^2 - (\sum x^2)} \longrightarrow \frac{10(303) - 50(50)}{10(340) - (50)^2} = 0/59$$

$$\mathbf{a} = \bar{Y} \ \mathbf{-b}\,\bar{x} \longrightarrow 50 - (0/59)(50) = 2/05$$

حال اگر مقادیر محاسبه شده برای **a,b** را در فرمول $\bar{Y}$ **=bx+a** قرار دهیم معادله خط رگرسیون به صورت زیر نوشته می شود:

$$\bar{Y} = 0/59x + 2/05$$

*برای محاسبه خطای معیار بر آورد :

$$s_Y = \sqrt{\frac{\sum Y^2 - (\sum Y)^2/N}{N-1}} = \sqrt{\frac{290 - (50)^2/10}{10-1}}$$

$$SE_{set} = s_y = \sqrt{1-r}^2 \longrightarrow 2/108\sqrt{1-(0/88)}^2 = 1$$

- نکته: انواع تقسیم بندی ضرایب همبستگی بر اساس سطوح سنجش مقیاس ها:

- مقیاس اسمی: خی دو، کای اسکوائر، ضریب فی، ضریب لاندا، ضریب توافق چوپر، ضریب کرامرز، ضریب توافق پیرسون ، ضریب یول و ضریب کراسکال و گودمن
- مقیاس ترتیبی: گاما، سامرز (**d**)، کندال، و همبستگی اسپیرمن
- مقیاس فاصله ای و نسبی: ضریب همبستگی پیرسون، ضریب تعیین، رگرسیون خطی ساده، رگرسیون خطی استاندارد.
- نکته: ضرایب کاهش نسبی خطا عبارتند از :

ضریب تعیین(r^2)، گاما، همبستگی درون طبقه ای ، سامرز، کندال

- نکته: ضرایب همبس
- تگی متقارن عبارتند از :

کرامرز، کندال، گاما و پیرسون

- نکته: و ضرایب همبستگی نامتقارن عبارتند از :

لاندا، گودمن و کراسکال، سامرز

تست ها ی فصل ششم

۱- شاخص پیوستگی T_b گودمن – کراسکال برای اندازه گیری شدت پیوستگی چه نوع متغیرهایی مناسب است؟

الف) رتبه ای- رتبه ای　　ب) اسمی- رتبه ای

ج) رتبه ای- فاصله ای　　د) فاصله ای- فاصله ای

۲- تای کندال b برای همبستگی بین چه نوع متغیرهایی بکار می رود؟

الف) هر دو رتبه ای برای جداول مربع

ب) هر دو اسمی برای هر اندازه جدول

ج) هر دو فاصله ای برای هر اندازه جدول

د) هر دو رتبه ای برای جداول مستطیلی

۳- برای سنجش همبستگی میان دو متغیر جنسیت (مرد- زن) و میزان موفقیت (هیچ ، کم ، متوسط، زیاد، خیلی زیاد) از کدام ضریب همبستگی باید استفاده کرد؟

الف) تای کندالc　　ب)تای کندال b　　ج)v کرامرز　　د)r پیرسون

۴- ضریب همبستگی رتبه ای اسپیرمن در توزیع آماری زیر چقدر است ؟

الف) 1-　　ب)0/5-　　ج)0/5+　　د)1+

X	5	6	8	4	3
Y	4	2	1	6	8

۵- در مطالعه ای نمرات میان ترم و پایان ترم دو گروه از دانشجویان مقایسه شده است و ارزش های زیر بدست آمده است. (s=۰/۰۰۱ و r=۰/۱۵) کدامیک از جملات زیر توصیف کننده این ارزش ها است؟

الف)همبستگی بسیار قوی بین نمرات وجود دارد ولی از نظر آماری معنی دار نیست.

ب) همبستگی قوی بین نمرات وجود دارد و از نظر آماری معنی دار است.

ج) همبستگی ضعیفی بین نمرات وجود دارد ولی از نظر آماری هم معنی دار است.

د) همبستگی ضعیفی بین نمرات وجود دارد ولی از نظر آماری معنی دار نیست.

۶- ضریب تعیین ($R^۲$) نشان دهنده چیست؟

الف) درصدی از واریانس مقادیری که دارای همبستگی خیلی ضعیف هستند.

ب) درصدی از واریانس مقادیری که دارای همبستگی خیلی قوی هستند.

ج) درصدی از واریانس متغیر وابسته که توسط متغیر تبیین نمی شود.

د) درصدی از واریانس متغیر وابسته که توسط متغیر مستقل تبیین می شود.

داده های جدول روبرو نشان می دهد که:

علاقه به تحصیل بر حسب طبقه اجتماعی				
علاقه به تحصیل	بالا	متوسط	پایین	کل
بالا	8	14	8	30
پایین	2	16	12	20
کل	10	30	20	50

الف) بین علاقه به تحصیل و طبقه اجتماعی رابطه کامل وجود دارد.

ب) بین علاقه به تحصیل و طبقه اجتماعی رابطه مثبت وجود دارد .

ج) بین علاقه به تحصیل و طبقه اجتماعی رابطه مثبت وجود دارد .

د) بین علاقه به تحصیل و طبقه اجتماعی رابطه ای وجود ندارد.

۸- اگر مقدار r پیرسون ۰/۴۵- باشد، چگونه آن را تفسیر می کنید؟

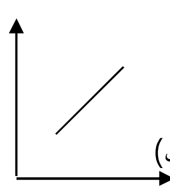
الف) رابطه همبستگی بین دو متغیر ضعیف و مستقیم است .

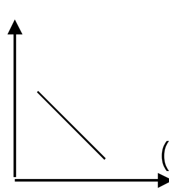
ب) رابطه بین دو متغیر متوسط و مستقیم است.

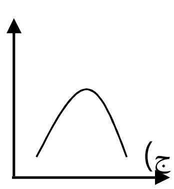
ج)رابطه بین دو متغیردرحد متوسط قوی وجهت آن معکوس است.

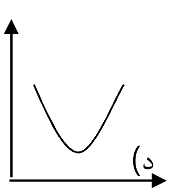
د) رابطه این دو متغیر ضعیف و معکوس است.

۹- منحنی خطی همبستگی کامل و منفی چگونه است؟

الف)


ب)


ج)


د)


۱۰- مقدار ضریب همبستگی پیرسون و ضریب همبستگی اسپیرمن حداقل چقدر و حداکثر چقدر است؟

الف) پیرسون صفر تا 1 و اسپیرمن بین 1- و 1+

ب) پیرسون بین 1- و 1+ و اسپیرمن بین صفر تا 1

ج) پیرسون بین 1 تا 10 و اسپیرمن بین صفر تا 10

د) هر دو بین 1- و 1+ تغییر می کنند.

۱۱- ضریب همبستگی بین دوسری اعداد A,B جدول مقابل چقدر است؟

الف) 1-　　ب) 0

ج) 0/30　　د) 1+

A	2	5	8	11	14
B	14	11	8	5	2

۱۲- معنی دار بودن ضریب همبستگی بین دو متغیر r_{xy} به کدام عامل بستگی بیشتری دارد؟

الف) واریانس نمرات　　ب) میانگین نمرات

ج) مقدار آزمودنی ها　　د) عدم تجانس آزمودنی ها

۱۳- چنانچه دو متغیر را که در سطح مقیاس رتبه ای مورد اندازه گیری قرار گرفته اند در اختیار داریم و بخواهیم رابطه همبستگی بین این دو متغیر را محاسبه کنیم، کدام روش آماری مناسب تر است؟

الف)اسپیرمن　　) پیرسون　　) دورشته ای　　د) دورشته ای نقطه ای

۱۴- اگر ضریب همبستگی بین دو متغیر برابر ۰/۹۰ باشد، دو متغیر در چند در صد از واریانس کل مشترک می باشند؟

الف) 9　　ب) 18　　ج) 81　　د)90

۱۵- خط رگرسیون با ضریب زاویه (a) و ضریب پایه (۵) را بنویسید و بازاء x=۵ مقدار Y را محاسبه کنید. کدام گزینه پاسخ صحیح است؟

الف) 15　　ب) 5　　ج) 23　　د) 27

۱۶- شرط نرمال بودن در یکی از متغیرهای x یا Y وجود ندارد. به جای ضریب همبستگی پیرسون کدام ضریب همبستگی را حساب کنیم؟

الف) r_{tet} (ضریب تتراکوریک) ب) r_{bis} (دورشته ای)

ج) f (ضریب همبستگی فی) د) r_{pbis} (دو رشته ای نقطه ای)

۱۷- در صورتی که همبستگی بین دو متغیر مساوی ۰/۸۰ و میزان پراکندگی متغیر ملاک مساوی ۶/۷ باشد. مقدار خطای استاندارد پیش بینی کدام است؟

الف) 1/54 ب) 3/90 ج) 4/16 د) 5/20

۱۸- اگر $\sum[x-\bar{x}][Y-\bar{Y}]$ برابر با ۲۵۰۰ باشد، با توجه به n=۲۰ و s_x=۵۰ و s_y=۱۰ ضریب همبستگی پیرسون چقدر است؟

الف) 0/25 ب) 0/5 ج) 0/7 د) 1

۱۹- با افزایش مقدار ضریب همبستگی بین دو متغیر x وy خطای معیار برآورد y از روی x :

الف) افزایش می یابد ب) به صورت غیر سیستماتیک تغییر می کند.

ج) ثابت باقی می ماند. د) کاهش می یابد.

۲۰- کدامیک از ضرایب همبستگی زیر همبستگی قویتری را نشان می دهد؟

الف) r_{xy} =0 ب) r_{xy} =0/80 ج) r_{xy} =0/70 د) r_{xy} =0/50

۲۱- در کدام یک از وضعیت های زیر اثر رگرسیون وجود ندارد ؟

الف) همبستگی صفر ب) همبستگی 0/5- ج) همبستگی 0/5+ د) همبستگی کامل

۲۲- اگر sp_{xy} =۳۰۰ و ss_x =۲۰۰ و $\bar{x}$ =۲ و $\bar{Y}$ =۳ باشد، معادله خط رگرسیون کدام است؟

الف) $Y = \frac{2}{3}x + 3$ ب) $Y = 1/5x + 2$ ج) $Y = x\frac{2}{3}$ د) $Y = 1/5x$

۲۳- رعایت کدام یک از مفروضات زیر در محاسبه ضریب همبستگی پیرسون الزامی است؟

الف) وجود مقیاس رتبه ای یا فاصله ای ب) توزیع طبیعی

ج) یکسانی پراکنش بین متغیرها د) عدم وجود رابطه خطی

۲۴- تحقیق همبستگی قادر به پاسخگویی به کدام موضوع زیر نیست؟

الف) تعیین جهت رابطه علّی میان دو دسته از نمره ها

ب) تعیین رابطه میان دو دسته از نمره ها

ج) تعیین جهت همبستگی میان دو دسته از نمره ها

د) تعیین میزان وابستگی میان دو دسته از نمره ها

۲۵- ضریب تبیین (مربع ضریب همبستگی) میان دو متغیر فاصله ای عبارت است از . . . بین دو متغیر.

الف) واریانس مشترک　　ب)رابطه علت و معلولی

ج) تعیین نوع رابطه موجود　　د) میزان اشتراک تغییرات

۲۶- وقتی دو متغیر در جهت عکس هم تغییر کنند، مقدار ضریب همبستگی . . . است.

الف) منفی　　ب) مثبت　　ج) صفر　　د) یک

۲۷- کدام عامل بر ضریب همبستگی تأثیر بیشتری دارد؟

الف) توزیع طبیعی　　ب) شکل نمره ها

ج)محدودیت در نمره ها　　د) همگنی واریانس ها

۲۸- ضریب همبستگی دو متغیر را محاسبه کرده ایم مقدار ۰/۵۰ بدست آمده است، کدام تفسیر زیر از ضرایب حاصل درست است؟

الف) اگر یک عامل را در عامل دیگر پیش بینی می کنیم، پیش بینی ما در 50 درصد موارد درست است.

ب) دامنه ارزش ها در دو توزیع وسعت کافی نداشته است و در صورتی که می داشت ضریب کامل حاصل می شد.

ج) میان دو عامل به واسطه کمی میزان همبستگی نمی توان حکم به رابطه علت و معلولی کرد.

د) 25 درصد از تغییرات یک متغیر به علت تغییرات متغیر دیگر است.

پاسخنامه تست های فصل ششم

1- گزینه ب

2- گزینه الف

3- گزینه ج

4- گزینه الف

5- گزینه د

6- گزینه د

7- گزینه ج

8- گزینه ج

9- گزینه ب

10- گزینه د

11- گزینه الف

- نکته: چون افزایش متغیر **A** با کاهش متغیر **B** همراه است، ضریب همبستگی منفی یک خواهد بود.

12- گزینه ج

- نکته: عاملی که بیشترین تأثیر را بر معنی داری ضریب همبستگی دارد حجم واحد نمونه است.

13- گزینه الف

14- گزینه ج

15- گزینه الف

$\mathbf{b}=2 \qquad \mathbf{a}=5 \Rightarrow Y=2\mathbf{x}+5 \qquad \mathbf{x}=5 \Rightarrow Y=2(5)+5=15$

16- گزینه ب

17- گزینه الف

18- گزینه الف

$$\mathbf{f}_{xy}=\frac{\frac{2500}{20}}{(50)(10)}=0/25$$

19- گزینه د

20- گزینه ج

21- گزینه د

22- گزینه د

$$\mathbf{a} = \frac{sp}{ss} = \frac{300}{200} = 1/5 \qquad \mathbf{b} = \overline{Y} - \mathbf{a}\,\overline{x} = 3-1/5\,(2) = 0 \qquad \mathbf{y} = 1/5\mathbf{x}$$

23- گزینه ج

24- گزینه الف

25- گزینه الف

26- گزینه الف

27- گزینه د

28- گزینه د

آزمون خودسنجی فصل ششم

۱- اگر ضریب همبستگی در معادله رگرسیون برابر با ۰/۶ باشد، مقدار ضریب تعیین یا تشخیص برابر با چه عددی است؟

1) 0/12   2) 0/18   3) 0/24   4) 0/36

۲- اگر در مدل تحلیل ضریب همبستگی خطی چند متغیره نتیجه Multiple-R برابر با ۰/۳۶ باشد، جایگاه تحلیل ۰/۶۴ باقیمانده در کجاست؟

1)باید احتمال تأثیر متغیرهای درگیر را مردود کرد.

2)باید از تحلیل جزئی و **Residual** استفاده کرد.

3) باید از تحلیل مبتنی بر انحراف معیار و لامبدا استفاده کرد.

4)باید تحلیل را متوجه پاسخ‌های نادرست و عدم سنجش‌پذیریِ مفاهیم کرد.

۳- برای تحلیل داده‌ها و کشف روابط ابتدا از کدام دو آماره موجود زیر استفاده می‌کنید؟

1) آزمون کرامر**v** و گاما   2)آزمون کاپا و آزمون لامبدا

3)آزمون همبستگی پیرسون و اسپیرمن   4)آزمون تحلیل واریانس پیشرفته و آزمون کاپا

۴- نتیجه لامبدا مساوی ۰/۲۸ است، چگونه آن را تفسیر می‌کنید؟

1)قوت همبستگی میان دو متغیر مستقل و وابسته 0/28 از یک است.

2)با دانستن متغیر مستقل می‌توان در 28 درصد موارد تغییرات متغیر وابسته را پیشگویی کرد.

3)ارزش 0/28 آن چنان بالا نیست که بتوان با آن قوت همبستگی میان دو متغیر را تفسیر کرد.

4)با دانستن ارزش‌های متغیر وابسته می‌توان در 28 درصد موارد تغییرات متغیر مستقل را پیش‌بینی کرد.

۵- از میان سه آماره زیر کدام‌یک از سنجش قوت همبستگی دقیق‌تر عمل می‌کند:

۱- ضریب phi ۲- ضریب contingency و ۳- کرامرز v

1)کرامرز**v**   2)ضریب **phi**

3)ضریب **contingency**   4) ضریب **phi** و ضریب **contingency**

۶- تفاوت میان آزمون گاما و سامرز D چیست؟

1)گاما همان سامرز D است و در گاما متغیر مستقل مشخص است.

2) سامرز D همان گاما است با این تفاوت که در آن متغیر وابسته مشخص است.

3) سامرز D برای سنجش متغیرهای اسمی بکار می‌رود و گاما برای سنجش در متغیرهای ترتیبی استفاده می‌شود.

4) گاما برای سنجش قوت همبستگی متغیرهای اسمی است ولی سامرز D برای سنجش قوت همبستگی در متغیر ترتیبی است.

۷- مقدار B در تحلیل همبستگی خطی برای متغیر نرخ باروری مساوی -۰/۷۰ است. اگر مقدار ثابت B برابر با ۸۹(امید به زندگی) باشد، چه تفسیری باید کرد؟

1)مقدار B تأثیری به روند امید به زندگی ندارد چرا که مقدار 7/.- کم است.

2)هر چه سن بالا می‌رود به ازای 70/.-از نرخ باروری کم می‌شود.

3) به ازای 89 واحد تغییر در یک متغیر به میزان 0/70- از نرخ باروری بر روی منحنی مختصات کاسته می‌شود.

4) به ازای یک واحد تغییر در نرخ باروری، منحنی خطی 70/.- به سمت پایین تغییر می‌کند، یعنی از امید به زندگی کم می‌شود.

۸- هدف از تحلیل همبستگی چند متغیره چیست؟

1) نشان دادن رابطه میان مربع R و خی دو همه متغیرها

2) بی‌تأثیر کردن آزمون ضریب همبستگی و موثر کردن آزمون لامبدا

3) نشان دادن رابطه چند متغیر باهم و تعیین ضریب میانگین تأثیرها

4) نشان دادن نسبت تأثیر متغیرهای متفاوت و تعیین ضریب باقیمانده

۹- همبستگی بین یک سوال و نمره کل آزمون به صورت کدام‌یک از ضرایب زیر محاسبه می‌شود؟

1) اسپیرمن 2) دو رشته‌ای 3)خی دو 4)تتراکوریک

۱۰- رگرسیون بیانگر چیست؟

1)درجه ارتباط دو متغیر با یکدیگر 2)ارتباط یک متغیر یا چند متغیر

3) میزان تأثیر گذاری چند متغیر روی یک متغیر 4)پیش‌بینی یک متغیر از روی متغیر دیگر

۱۱- در کدام‌یک از وضعیت‌های زیر رگرسیون در اطراف میانگین به صورت کامل خواهد بود؟

1) همبستگی صفر 2) همبستگی 1+ 3) همبستگی 0/5+ 4)همبستگی 1-

۱۲- اگر ضریب تعیین یا تشخیص $R2$ در یک معادله رگرسیون برابر با ۰/۲۵ باشد، ضریب همبستگی مربوط به معادله کدام است؟

0/8(1 0/32(2 0/45(3 0/50(4

۱۳- برتری آزمون $X2$ (کی دو) نسبت به آزمون همبستگی خطی کدام است؟

1) نشان دادن رابطه‌های خطی معنی‌دار

2)نشان دادن رابطه‌های غیر خطی معنی‌دار

3) نشان دادن میزان تأثیر یک متغیر بر متغیر دیگر

4)مناسبت بیشتری برای متغیرهای فاصله‌ای

۱۴- لاندای قرینه نمایانگر کدام رابطه بین متغیرهاست؟

1) اثر متغیر الف بر متغیر ب 2) اثر متغیر ب بر متغیر الف

3)رابطه متقابل متغیرهای الف و ب 4) اثر متغیر ثالث بر متغیرهای الف و ب

۱۵- نمونه‌ای از تحلیل چند متغیره عبارتست از:

1)تحلیل همبستگی بین سن، جنس و نوع جرایم تکرار شده در یک شهر معین

2) تحلیل همبستگی بین انواع رشته‌های کارشناسی دانشگاه‌ها و مقامی که در اشغال افراد است.

3) تحلیل همبستگی بین جرم و طول مدت زندان برای کسانی که محکمه قضایی شده‌اند.

4) آزمون سنین همه زنانی که مشاغل صنفی بالا دارند.

۱۶- در یک تحقیق اثرات اعتماد به نفس، هوش و تحصیل والـدین در موفقیـت تحصـیلی دانـش-آموزان مورد بررسی قرار گرفته، آزمون آماری تحلیل نتایج چه آزمونی باشد؟

1)ضریب گشتاوری 2)ضریب همبستگی

3)رگرسیون تک متغیری 4)رگرسیون چند متغیری

۱۷- شخصی از شما می‌خواهد که مقدار گامای ۱/۳ را تفسیر کنید، شما چه خواهید گفت؟

1)نشانه یک همبستگی مثبت است.

2)این مقدار نشانه یک همبستگی قوی است.

3)این مقدار نشانه یک همبستگی معتدل است.

4) در محاسبات اشتباه شده است.

۱۸- آماره صحیح برای پیشگویی رابطه همبستگی میان جنس و مذهب کدام است؟

1)خی دو 2)گاما 3)لامبدا 4)غیر متغیرها

۱۹- در جدول زیر چه نوع رابطه‌ای بین تماشای تلویزیون و تحصیلات وجود دارد؟

1)رابطه خطی نزولی

2)رابطه غیر خطی نزولی

3)رابطه خطی صعودی

4) رابطه غیر خطی صعودی

میزان تماشای تلویزیون (بر حسب ساعت)	تحصیلات
0	8
1	6
2	4
3	2
4	0

۲۰- با توجه به جدول سوال ۲۷ ضریب زاویه(b) رگرسیون میزان تماشای تلویزیون بـه تحصـیلات چقدر است؟

1)-2 2)-1 3) $-\frac{1}{2}$ 4) $-\frac{1}{4}$

۲۱- با توجه به جدول سوال ۲۷ عرض از مبدأ(a)، چقدر است؟

1) صفر 2) $\frac{1}{2}$ 3)2 4)4

۲۲- مقدار ضریب تعیین(r^2) سوال ۲۷، چقدر است؟

1(1   0/5(2   0/8(3   0/3(4

۲۳- توزیع کتابخوانی مردم یک محله در جدول زیر ارائه شده است. برای سنجش رابطه کتابخوانی با وضع اشتغال، کدام ضریب مناسب است؟

آیا کتاب می-خوانید؟	شاغل	غیر شاغل	کل
بله	1	4	5
خیر	4	1	5
کل	5	5	10

1)ضریب تعیین(r^2)

2)ضریب گاما(λ)

3) تا او گودمن و کراسکال(T_r)

4) ضریب رتبه‌ای اسپیرمن(p_s^2)

۲۴- برای آزمون رابطه دو متغیر رتبه‌ای، کدام ضریب مناسب است؟

1)لاندا   2)گاما   3)کی دو   4)همبستگی پیرسون

۲۵- در صورتی که همبستگی بین دو. متغیر مساوی ۸۰/. باشد کدام‌یک از جملات زیر صحیح است؟

1)متغیر ملاک 80 درصد از تغییرات متغیر پیش‌بینی کننده را تبیین می‌کند.

2)64 درصد از تغییرات متغیر ملاک توسط متغیر پسش‌بینی کننده تبیین می‌شود.

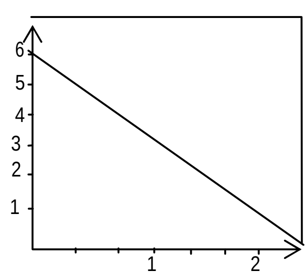
3) مجذور همبستگی بین این دو متغیر 80 درصد است.

4) متغیر پیش‌بینی کننده 64 درصد از تغییرات کلیه متغیری مورد پژوهش را تبیین می‌کند.

۲۶- بین همبستگی و کدام‌یک از شاخص‌های زیر رابطه معکوس وجود دارد؟

s_y (1 s_x (2 s_{xy} (3 s^* (4

۲۷- نمودار پراکنش X, Yمبین چه نوع رابطه‌ای بین این دو متغیر است؟



1)خطی مستقیم

2)خطی معکوس

3)غیر خطی مستقیم

4) غیر خطی معکوس

۲۸- با توجه به نمودار سوال ۳۸، ضریب تعیین(r^2) رگرسیون yبهx چقدر است؟

0/5(1 0/8(2 0/9(3 1(4

۲۹- در جدول توزیع زیر فراوانی مورد انتظار برای خانه‌های «ج» و «ب»، به ترتیب از راست بـه چـپ چقـدر است؟|

1) 10 و 10

2)12 و 42

3)35 و 65

4)65 و 35

	A	B	جمع
C	الف	ب	60
D	ج	د	40
جمع	30	70	100

۳۰- در یک رابطه خطی ساده بین سن و درآمد، اگر ضریب عرض از مبدأ برابر ۳۰۰۰ تومـان و ضـریب زاویـه برابر ۲۰ باشد، درآمد فرد ۴۰ ساله چقدر است؟

2200(1 2940(2 3060(3 3800(4

۳۱- با توجه به جدول زیر، مقدار ضریب گاما چقدر است؟

1) 0/8

2) 0/6

-0/8(3

-0/5(4

گرایش به مساوات	تحصیلات بالا	تحصیلات پایین
بالا	20	150
پایین	40	100

۳۲- اگر فردی بخواهد همبستگی بین طبقه اجتماعی و بزهکاری جوانان را به شرط آن‌که تعداد بـرادران یـا خواهران را ثابت نگه دارد، بررسی نماید، مناسب‌ترین تکنیک کدام خواهد بود؟

1)همبستگی جزئی 2)همبستگی چند متغیره

3)تحلیل سری زمانی 4)تحلیل عامل

۳۳- در یک توزیع دو در دو، مقدار آزمون کی‌دو ۳۵ و تعداد موارد ۱۴۰ باشد، ضریب فی چقدر است؟

0/25(1 0/5(2 4(3 16(4

۳۴- چنانچه همبستگی بین دو متغیر مساوی ۰/۹۰- باشد، چند درصد از واریانس بین این دو متغیـر در ایـن پژوهش تبیین شده‌است؟

1) 81 درصد 2) 90 درصد 3) 19 درصد 4) منفی 81 درصد

۳۵-برای تعیین معنی‌دار بودن یک ضریب همبستگی درجات آزادی مساوی کدام‌یک از مقادیر زیر است؟

1) $n-1$ 2) n_1+n_2-1 3) n_1+n_2-2 4) $n-2$

۳۶- مقدار ضریب لاندای رابطه گرایش به برابری با طبقه چقدر است؟

گرایش به برابری	تحصیلات بالا	تحصیلات پایین
بالا	10	90
پایین	90	50

1) -0/60

2) 0/60

3) 0/40

4) -0/40

۳۷- ضریب همبستگی $r(X, Y)$ چقدر است؟

X_i	Y_i
0	6
1	4
2	2
3	0

1) -1

2) 1

3) 0/50

4) -0/50

۳۸- آزمون«وی کرامر» از خانواده آزمون‌هایی است که ....

1) شدت رابطه را می‌سنجد.

2) امکان ارزیابی ماهیت رابطه میان دو یا چند متغیر را با یک متغیر وابسته فراهم می‌کند.

3) نسبت متغیرها را مورد بررسی قرار می‌دهد.

4) فقط معنی‌داری رابطه را می‌سنجد.

۳۹- در جدول زیر چه رابطه‌ای بین عزت نفس و موفقیت تحصیلی وجود دارد؟

1) مستقیم

2) رابطه‌ای وجود ندارد.

3) معنی‌دار غیر خطی

4) معکوس

موفقیت	عزت نفس			
	ضعیف	متوسط	قوی	جمع
خوب	10	15	55	80
متوسط	12	60	20	95
ضعیف	60	10	5	75
جمع	82	90	80	252

۴۰- با توجه به اطلاعات سوال ۱۲۰، رابطه بین عزت نفس و موفقیت توسط کدام شاخص اندازه‌گیری می‌شود؟

1)گاما 2) آنالیز واریانس 3) همبستگی پیرسون 4) لاندا

۴۱- ضرایب همبستگی در مدل تحلیل مسیر عبارت است از...............................

1) میزان همبستگی جزیی متغیرها 2) میزان همبستگی دو متغیر با متغیر سوم

3) میزان همبستگی همه متغیرها با هم 4) میزان همبستگی دو به دو متغیرها باهم

۴۲- رابطه هم بستگی پیرسون از چه نوع رابطه‌ای است؟

1) علّی 2) هم‌تغییری 3) کنترلی 4) غیر خطی

۴۳- جدول متقاطع (مرد و زن)، سن (بالا، متوسط و پایین) و وضع سواد (با سواد و بی سواد) جدول چند متغیره و چند در چند است؟

1) سه متغیره- 6×3 2) شش متغیره-$2\times2\times3$ 3) شش متغیره 4×3 4) سه متغیره- $2\times3\times3$

۴۴- چنانچه قصد داشته باشیم همبستگی بین متغیر گسسته دو ارزشی واقعی را تعیین کنیم، کدامیک از شاخص‌های زیر مناسب‌تر است؟

1) خی دو 2)فی 3)دو رشته‌ای 4)دو رشته‌ای نقطه‌ای

۴۵- مطابق داده‌های جدول روبرو، میزان همبستگی بین دو صفت X,Y کدام است؟

X	5	7	9	3
Y	4	8	6	2

0/25 (1

0/60 (2

0/75 (3

0/80 (4

۴۶- ضریب همبستگی بین «استعداد» و «معدل دانشگاهی» برابر ۰/۸ است. اگر واریانس متغیر استعداد $\frac{4}{9}$ باشد، خطای استاندارد پیش بینی کدام است؟

1/2(1 0/4 (2 0/6 (3 0/9 (4

۴۷- پژوهشگری ده آزمودنی را برای سه طبقه متمایز آزمون می‌کند، درجه آزادی در تنظیم نتایج این آزمون کدام است؟

9(1 10 (2 12(3 18(4

۴۸- از بین افراد یک جامعه ۶۰۰ نفر به تصادف انتخاب می‌شوند، اگر ۳۶۰ نفر از آنان علاقه به مطالعه داشته باشند، خطای برآورد نسبت کدام است؟

0/02 (1 0/025(2 0/04(3 0/075(4

۴۹- پژوهشگری از ۵۰ نفر زن و ۵۰ نفر مرد درباره پخش یک برنامه صدا و سیما پرسش کرد، نظر آنها در جدول زیر آمده است. شدت این همبستگی کدام است؟

	موافق	مخالف
زن	20	150
مرد	40	100

0/16(1

0/18(2

0/21 (3

0/24(4

پاسخنامه آزمون خودسنجی فصل ششم

۱- گزینه ۴ صحیح است.

ضریب تعیین شاخصی است که نسبتی از واریانس متغیر $\mathbf{y}$ را که توسط متغیرهای $x_1, x_2, \ldots x_i$ به حساب آمده است، را معین می‌کند. در واقع می‌توان گفت ضریب تعیین مجذور ضریب همبستگی است.

$$\text{ضریب تعیین} = \left(r_{xy}\right)^2 \Rightarrow (0/6)^2 = 0/36$$

۲- گزینه ۲ صحیح است.

0/64 معادل واریانس پسمانده یا مجموع مجذورات باقیمانده(**Residual sum of squares**) می‌باشد که بیانگر مقدار تغییرات متغیر(**y**) است مه توسط متغیرهای مستقل $x_1, x_2, \ldots x_i$ مورد تبیین و پیش بینی قرار نگرفته است.

۳- گزینه ۳ صحیح است.

آزمون همبستگی اسپیرمن و پیرسون هم برای تحلیل داده‌ها و هم برای کشف وجود یا عدم وجود رابطه معنی‌دار مورد استفاده قرار می‌گیرند. اما گزینه‌های دیگر فقط برای تحلیل داده ها هستند، نه برای کشف ارتباط معنی دار.

۴- گزینه ۲ صحیح است.

شاخص لاندا، از نوع شاخص کاهش نسبی خطاست. یعنی شاخصی است که بیانگر آن است که با آگاهی از مقادیر متغیر مستقل، میزانِ دقت پیش گویی متغیر وابسته را مشخص می‌کند. برای مثال در این سوال مقدار لامبدا برابر با 0/28 می‌باشد، تفسیر آن این است که با آگاهی از متغیر مستقل می‌توان 28 درصد موارد تغییرات متغیر وابسته را پیشگویی کرد.

۵- گزینه ۱ صحیح است.

هر سه ضریب همبستگی در سطح سنجش اسمی مورد استفاده قرار می‌گیرند. اما کرامر(**V**) بهترین و کامل-ترین شاخص همبستگی در این سطح سنجش می‌باشد.

۶- گزینه ۲ صحیح است.

ضریب همبستگی سامرز و ضریب همبستگی گاما در سطح سنجش ترتیبی مورد استفاده قرار می‌گیرند و همچنین هر دو جزء شاخص‌های کاهش نسبی خطا هستند. اما از آنجایی که ضریب همبستگی سامرز نامتقارن است، می‌بایست متغیر وابسته مشخص شده باشد.(یک شاخص نامتقارن بر حسب اینکه متغیر وابسته در سطر باشد یا در ستون، ارزش‌های متفاوتی به خود می‌گیرد، بنابراین در شاخص‌های نامتقارن می-بایست مشخص گردد کدام متغیر مستقل است و کدام متغیر وابسته است.)

۷- گزینه ۴ صحیح است.

در تحلیل رگرسیون خطی (تحلیل رگرسیون ساده)، $\mathbf{B}$ مبین تغییر حاصل در $\mathbf{Y}$(متغیر وابسته) به ازای یک انحراف تغییر در متغیر $\mathbf{X}$(متغیر مستقل) می‌باشد، با توجه به سوال $(B=-0/7)$ می‌باشد، یعنی به ازای یک واحد تغییر در متغیر $\mathbf{X}$(نرخ باروری)، متغیر $\mathbf{Y}$(امید به زندگی) $-0/7$ واحد تغییر می‌کند. به عبارت دیگر منحنی خطی 70%- به سمت پایین تغییر می‌کند، یعنی امید به زندگی کم می‌شود.

۸- گزینه ۴ صحیح است.

تکنیک تحلیل رگرسیون(تحلیل همبستگی) از لحاظ ساختاری به سه نوع قابل تقسیم است:

1. تحلیل رگرسیون ساده(یک متغیر وابسته و یک متغیر مستقل)
2. تحلیل رگرسیون چند متغیری(یک متغیر وابسته و چند متغیر مستقل)
3. تحلیل رگرسیون چندگانه(چند متغیر وابسته و چند متغیر مستقل)

در تحلیل رگرسیون چند متغیره واریانس متغیر $\mathbf{y}$ از طریق مشارکت نسبی و ترکیب خطی دو یا چند متغیر مستقل $\mathbf{x}$ مورد تبیین و پیش‌بینی قرار می‌گیرد، بنابراین، در تحلیل رگرسیون چند متغیره، یک متغیر وابسته $\mathbf{y}$ و مجموعه‌ای از متغیرهای مستقل$\mathbf{x}$ وجود دارد که هر کدام از این متغیرها، ضریب تأثیر متفاوتی دارند به همین دلیل می‌توان گفت هدف از تحلیل رگرسیون چند متغیره، نشان دادن نسبت تأثیر متغیرهای متفاوت و همچنین تعیین مقدار تغییراتی از $\mathbf{y}$ که مورد تبیین و پیش‌بینی قرار نگرفته است، یعنی تعیین ضرایب باقیمانده.

۹- گزینه ۱ صحیح است.

ضریب همبستگی رتبه‌ای اسپیرمن **r**، برای توصیف همبستگی میان دو متغیر که با استفاده از مقیاس فاصله-ای یا نسبی اندازه‌گیری شده باشند، به کار برده می‌شود

۱۰- گزینه ۴ صحیح است.

با استفاده از تحلیل رگرسیون می‌توان به تبیین و پیش‌بینی تغییرات متغیر وابسته از طریق متغیر مستقل پرداخت.

۱۱- گزینه ۱ صحیح است.

چنانچه همبستگی بین متغیرها صفر باشد، رگرسیون در اطراف میانگین به صورت کامل اتفاق خواهد افتاد.

۱۲- گزینه ۴ صحیح است.

ضریب تعیین $(R^2) = 0/25$

ضریب هم بستگی $R = \sqrt{0/25} = 0/5$

۱۳- گزینه ۲ صحیح است.

هدف این آزمون آن است که مشخص کند که آیا متغیرهای مورد مطالعه مستقل از یکدیگر هستند یا با هم پیوستگی دارند. آزمون کای اسکوئر یکی از آزمون‌های غیر پارامتری است که با استفاده از آن می‌توان احتمال شانسی بودن اختلاف بین فراوانی‌های مشاهده شده و فراوانی‌های مورد انتظار را برآورد کرد. بدین معنا که تفاوت موجود در بین فراوانی‌های مورد انتظار تفاوتی است معنادار یا تفاوت ناچیز و شانسی است.

آزمون $\mathbf{x}^2$ صرفاً برای به دست آوردن این نکته است که آیا رابطه معناداری بین متغیرها وجود دارد یا نه. این آزمون از شدت و ضعف رابطه بین متغیرها هیچ صحبتی نمی‌کند. هم چنین این آزمون رابطه غیر خطی معنادار را نشان می‌دهد.

۱۴- گزینه ۳ صحیح است.

ضریب لاندا، شاخصی است که بیانگر آن است که با آگاهی از مقادیر متغیر مستقل، میزانِ دقت پیش گویی متغیر وابسته را مشخص می‌شود. در واقع بیانگر رابطه متقابل بین دو متغیر مستقل و وابسته می‌باشد

۱۵- گزینه ۱ صحیح است.

در تحلیل رگرسیون چند متغیره واریانس متغیر **y** از طریق مشارکت نسبی و ترکیب خطی دو یـا چنـد متغیـر مستقل **x** مورد تبیین و پیش‌بینی قرار می‌گیرد، بنابراین، در تحلیل رگرسیون چند متغیره، یک متغیر وابسـته **y** و مجموعه‌ای از متغیرهای مستقل**x** وجود دارد. با این توصیف تنها گزینه الف دارای این ویژگی است، یعنـی از یک متغیر وابسته(نوع جرایم تکرار شده) و دو متغیر مستقل(سن و جنس) تشکیل شده است. .

۱۶- گزینه ۴ صحیح است.

در تحلیل رگرسیون چند متغیره واریانس متغیر **y** از طریق مشارکت نسبی و ترکیب خطی دو یـا چنـد متغیـر مستقل **x** مورد تبیین و پیش‌بینی قرار می‌گیرد، بنابراین، در تحلیل رگرسیون چند متغیره، یک متغیر وابسـته **y** و مجموعه‌ای از متغیرهای مستقل**x** وجود دارد. در این سوال موفقیت تحصیلی متغیر وابسته(**y**) می‌باشـد، و مجموعه متغیرهای مستقل(**x**) اعتماد به نفس، هوش و تحصیل والدین هستند، که میزان اثرات آنها بـر متغیـر وابسته قرار است سنجیده شوند.

۱۷- گزینه ۴ صحیح است.

ضریب همبستگی گاما در سطح سنجش ترتیبی مورد استفاده قرار می‌گیرد، و مقدار آن بـین 1+ تـا 1- متغیـر است. بنـابراین بـه هـیچ وجـه مقـدار آن 1/3 نخواهـد شـد، بـدین ترتیـب در محاسـبات اشـتباه شـده اسـت.

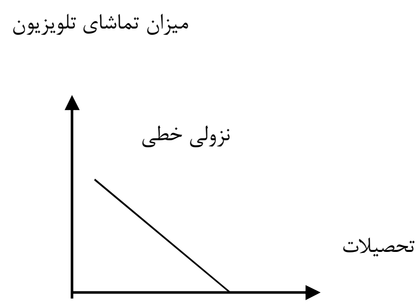
$$-1 < \gamma < +1$$

۱۸- گزینه ۳ صحیح است.

برای محاسبه شدت همبستگی دو متغیر جنس و مذهب که هر دو در سطح سنجش اسمی سنجیده می‌شـوند، می‌بایست از آماره‌ای استفاده نمود که در ایـن سـطح کـاربرد داشـته باشـد، بـا توجـه بـه سـوال تنهـا ضـریب همبستگی لامبدا در سطح سنجش اسمی مورد استفاده قرار می‌گیرد.

۱۹- گزینه ۱ صحیح است.

رابطه بین تماشای تلویزیون و تحصیلات یک رابطه خطی نزولی است. درواقع با افزایش تحصیلات، میزان تماشای تلویزیون کاهش می‌یابد.



۲۰- گزینه ۳ صحیح است.

تحصیلات(x)	تماشای تلویزیون(y)	xy	x^2
8	0	0	64
6	1	6	36
4	2	8	16
2	3	6	4
0	4	0	0
$\sum x_i = 20$	$\sum x_i = 10$	$\sum x_i y_i = 20$	$\sum x_i^2 = 120$

$$b_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$\Rightarrow b = \frac{5(20) - 20\times 10}{5(120) - (20)^2} = \frac{100-200}{600-400} = \frac{-100}{200} = -\frac{1}{2}$$

$$b = -\frac{1}{2}$$

در محاسبه **b** باید دقت کرد که ابتدا مقدار $\sum xy$، باید در **n** ضرب شود، و سپس از $(\sum x)(\sum y)$ کم شود.

(دلاور، 1387 : 206)

۲۱- گزینه ۴ صحیح است.

فرمول محاسبه عرض از مبدأ(**a**):

$$a_{xy} = \bar{y} - b_{xy}\bar{x}$$

$$b=-\frac{1}{2}$$
$$\bar{y}=?$$
$$\bar{x}=?$$

$$\bar{y}=\frac{\sum y}{n}\Rightarrow\frac{10}{5}=2$$
$$\bar{x}=\frac{\sum x}{n}\Rightarrow\frac{20}{5}=4$$

$$a=2-(-\frac{1}{2})(4)\Rightarrow 2-(-2)=4$$
$$a=4$$

۲۲- گزینه ۱ صحیح است.

با توجه به جدول سوال 20، به ازای دو واحد کاهش در متغیر تحصیلات(**x**)، میزان تماشای تلویزیون(**y**) یک واحد افزایش می‌یابد. به عبارت دیگر تغییرات در **x**و**y** به نسبت معینی می‌باشد، یعنی همبستگی کامل است و از طرفی تغییرات **x**و**y** بر خلاف یکدیگر هستند. بدین ترتیب رابطه معکوس خواهد بود.(**r = -1**)

حال اگر مقدار **r** این باشد، ضریب تعیین عبارت است از:

$$r=-1\Rightarrow r^2=1$$

۲۳- گزینه ۳ صحیح است.

تا او گودمن و کروسکال یک شاخص اسمی است که برای سنجش شدت و ضعف پیوستگی بین دو متغیر به کار می‌رود.

۲۴- گزینه ۲ صحیح است.

گاما شاخصی است که برای سنجش رابطه دو متغیر ترتیبی به کار می‌رود.

۲۵- گزینه ۲ صحیح است.

نام دیگر متغیر مستقل، متغیر پیش‌بینی کننده است و نام دیگر متغیر وابسته، متغیر ملاک است. ضریب تعیین(r^2) شاخصی است که از مجذور همبستگی به دست می‌آید و مبین مقدار واریانس تبیین شده متغیر وابسته(ملاک) توسط متغیر مستقل(پیش بینی کنند) است

۲۶- گزینه ۳ صحیح است.

خطای استانداردِ پیش بینی (s_{yx}) با همبستگی رابطه معکوس دارد. به این معنی که هر چه قدر همبستگی افزایش یابد خطای استاندارد پیش بینی کاهش می‌یابد. خطای استانداردِ پیش بینی به عنوان اختلاف بین نمره‌های مشاهده شده(**y**)

و نمره‌های پیش بینی شده (y') تعریف شده است. چنانچه بین نمره‌های فوق اختلافی وجود داشته باشد، خطای پیش بینی وجود نخواهد داشت، چنانچه همبستگی بین متغیرها زیاد باشد، خطای پیش بینی کم خواهد شد.

۲۷- گزینه ۲ صحیح است.

با کاهش متغیر **x**، متغیر **y** نیز به همان مقدار کاهش یافته است.

۲۸- گزینه ۴ صحیح است.

چون همبستگی بین **y** و **x** کامل است (1-)؛ بنابراین ضریب تعیینِ آن نیز 1 می‌شود.

$$v = (rxy)^2 \Rightarrow (-1)^2 = 1$$

۲۹- گزینه ۲ صحیح است.

تعداد کل ÷ (حاصل جمع سطر × حاصل جمع ستون) = فراوانی مورد انتظار

	A	B	جمع
C	الف	ب؟	60
D	ج؟	د	40
جمع	30	70	100

فراوانی مورد انتظار (ج) $= \frac{40 \times 30}{100} = \frac{1200}{100} = 12$

فراوانی مورد انتظار (ب) $= \frac{60 \times 70}{100} = \frac{4200}{100} = 42$

(دلاور، 1387 :490)

۳۰- گزینه ۴ صحیح است.

معادله ساده رگرسیون **x**و**y** عبارت است از:

$$y = a + b(x)$$

$a = 3000$

$b = 20$

$x = ?$

سن×ضریب زاویه(شیب رگرسیون) + عرض از مبدأ = درآمد

درآمد $= 3000 = 20(40) = 3800$

۳۱- گزینه ۲صحیح است.

$$G=\frac{NS-Nd}{NS+Nd}$$

برای محاسبه ضریب گاما ابتدا باید **NS** (مقدرا زوج‌های موافق)؛ را بدست آوریم. بدین منظور کوچکترین یا بزرگترین رتبه‌های هماهنگ **x** با **y** را انتخاب می‌کنیم. در این مثال گرایش به مساوات بالا و تحصیلات بالا را انتخاب می‌کنیم. این انتخاب می‌تواند تحصیلات پایین و گرایش به مساوات بالا هم باشد، در نتیجه فرقی نمی‌کند، فقط باید باهم هماهنگ باشند. حال فراوانی ستون و فراوانی سطر مربوط را ثابت می‌گیرند و عدد موجود در خانه زوج هماهنگ را با حاصل جمع فراوانی‌های باقیمانده در هم ضرب می‌کنیم.

برای محاسبه **Nd**(مقدار زوج‌های مخالف)، رتبه‌های ناهماهنگ را انتخاب می‌کنیم. در این مثال گرایش به مساوات بالا را با تحصیلات پایین در نظر می‌گیریم. سپس فراوانی ستون و فراوانی سطر زوج مخالف مورد نظر را ثابت گرفته، عدد موجود در خانه زوج مخالف را در حاصل جمع باقیمانده فراوانی‌ها یا داده‌ها ضرب می‌کنیم.

$$NS=20(100)+40(0)=2000$$
$$Nd=150(40)+40(0)=6000$$
$$G=\frac{2000-6000}{2000+6000}=\frac{-4000}{8000}=-0/5$$

۳۲- گزینه ۱ صحیح است.

ضریب همبستگی جزئی نشان‌دهنده شدت رابطه بین دو متغیر وابسته و مستقل است، منتها با از بین بردن اثر متغیر-های مستقل تأثیر دیگری که حذف شده‌است.

۳۳- گزینه ۲ صحیح است.

ضریب فی حاصل جذر مقدار خی دو از تعداد کل می‌باشد.

$$X^2=35$$
$$N=140$$
$$\phi=\sqrt{\frac{X^2}{N}}=\sqrt{\frac{35}{140}}=\sqrt{0/25}=0/5$$

۳۴- گزینه ۱ صحیح است.

تبیین مقدار واریانس (تغییرات) متغیر وابسته توسط متغیر مستقل به وسیله آماره‌ای به نام ضریب تعیین انجام می‌پذیرد، ضریب تعیین از مجذور ضریب همبستگی ضربدر 100 به دست می‌آید و چون ضریب همبستگی بـه تـوان دو مـی‌رسـد، پس ضریب تعیین هیچ وقت منفی نمی‌شود. در این مثال متغیر مستقل 81 درصد از تغییـرات متغیـر وابسـته را تبیـین می‌کند.

$$R^2 = (rxy)^2 \Rightarrow (-0/90)^2 = 81$$

۳۵- گزینه ۴ صحیح است.

برای رد یا اثبات فرضیه (تعیین معنادار بودن) در همبستگی از فرمول $t = r_s\sqrt{\frac{n-2}{1-r_s^2}}$ استفاده می‌شود که درجه آزادی آن در اینجا $(N-2)$ است. اگر **t** محاسبه شده با آلفای 0/05 بزرگتر از **t** استیودنت جدول باشد، فرض صفر رد می‌شود

۳۶- گزینه ۳ صحیح است.

فرمول ضریب لاندا عبارتست از:

$$\lambda = \frac{\sum m_y - M_y}{N - M_y}$$

در این فرمول $\sum m$ اندیس **y** عبارت از مجموع بزرگترین فراوانی توزیع‌های شرطی **y**ها مـی‌باشـد. بـه بیـان دیگر مجموع فراوانی‌های نمایی متغیر **y** در متغیـر **x** اسـت. و **M** انـدیس **y** عبارتسـت از بزرگتـرین فراوانـی توزیع حاشیه‌ای. به عبارت دیگر، فراوانی نمایی توزیع حاشیه‌ای متغیر **y** است. در این مثال

$$\sum m_y = 90+90=180 \qquad M_y = 900$$

$$N = 240$$

$$\lambda = \frac{180-140}{240-140} = 0/40$$

۳۷- گزینه ۱ صحیح است.

در این گزینه بین دو متغیر **x**و**y** رابطه خطی معکوس کامل وجود دارد. بدین معنی که با افزایش هـر واحـد از **x**ها دو واحد از **y**ها به طور ثابت کاسته می‌شود(افزایش **x** کاهش **y**) از آنجایی که مقدار کـاهش **y**هـا ثابـت است.(عدد 2). پس رابطه معکوس کامل مساوی 1- می‌باشد.

۳۸- گزینه ۱ صحیح است.

کرامر از جمله آماره‌هایی است که شدت رابطه میان دو متغیر اسمی را می‌سنجد.

۳۹- گزینه ۱ صحیح است.

در مقوله جدول خوانی، برای متغیرهای ترتیبی باید فقط اعداد چهار طرف جدول را مد نظر قرار دهیم. در اینجا اعداد 10، 55، 60 و5 مد نظر قرار می‌گیرند.

موفقیت	عزت نفس			جمع
	ضعیف	متوسط	قوی	
خوب	(a)۱۰	15	(b)۵۵	80
متوسط	12	60	20	95
ضعیف	(c)۶۰	10	(d)۵	75
جمع	82	90	80	252

سپس برای کشف اینکه آیا رابطه‌ای وجود دارد یا نه باید به گونه زیر عمل کنیم.

اگر $b < a$ $d > c,$ باشد، یا برعکس، $b > a$ $d < c,$ باشد، رابطه معنی‌داری بین دو متغیر وجود دارد. نکته‌ای که در اینجا وجود دارد این است که اگر در مواردی $b = a$ و **c** کوچکتر یا بزرگتر از **d** باشد (نامساوی) و یا $d = c$ باشد و **a** کوچکتر یا بزرگتر **b**، باز هم رابطه وجود دارد.

حالا برای تعیین جهت رابطه باید بزرگترین فراوانی‌های این چهارقطب را مشخص کنیم که در این سوال 60 و55 (c, b) می‌باشند. پس از این کار باید توجه کنیم که آیا بزرگترین فراوانی قطبی جدول در خانه‌ای قرار گرفته است که ارزش‌های ستون و ردیف متغیرها باهم به لحاظ ترتیبی همسان هستند یا ناهمسان. اگر همسان باشند، رابطه مستقیم است و اگر غیر همسان باشند رابطه معکوس است. در این سوال عدد 60 (خانه**c**) در سلول ضعیف – ضعیف ، قرار گرفته است وچون ارزش‌های دو متغیر همسان است، پس رابطه مستقیم است. اگر عدد 60 مثلاً در خانه ضعیف – قوی، قرار می‌گرفت، در نتیجه رابطه هم به دلیل ناهمسان بودنِ ارزش‌های ردیف و ستون معکوس می‌شد.

در این سوال همانگونه که ملاحظه می‌شود،

۴۰- گزینه ۱ صحیح است.

در این گزینه گاما تنها آماره‌ای است که برای سنجش شدت هم بستگیِ بین دو متغیـر ترتیبـیِ عـزت نفـس و موفقیت آورده شده است.

۴۱- گزینه ۴ صحیح است.

اصل اساسی در تحلیل مسیر پیدا کردن همبستگی دو به دو متغیرها باهم است.

۴۲- گزینه ۲ صحیح است.

پیرسون آماره‌ای است که هم‌تغییری دو متغیـر خطـی را نشـان مـی‌دهـد.(دلاور، 1387: 188) در واقـع ایـن ضریب همبستگی دارای مفروضاتی زیر است:1- رابطه بین دو متغیـر خطـی اسـت. 2- توزیـع‌هـا دارای شـکل مشابه هستند. 3- نمودار پراکندگی یکسان است.

۴۳- گزینه ۴ صحیح است.

در این مسأله جنس، سن و وضعیت سواد متغیرهای ما هسـتند. جـنس و وضـعیت سـواد متغیـر دو ارزشـی و اسمی هستند. و سن در سه طبقه دسته‌بندی شده است. این سه متغیر در ترکیب بـاهم تشـکیل یـک جـدول سه متغیره $3\times2\times2$را می‌دهند، در واقع بـا ضـرب ارزش‌هـای آن‌هـا در همـدیگر چنـد در چنـد بـودن جـدول مشخص می‌شود. این جدول دارای 12 سلول است.

وضعیت سواد	مرد			زن		
	سن بالا	سن متوسط	سن پایین	سن بالا	سن متوسط	سن پایین
باسواد						
بیسواد						

۴۴- گزینه ۲ صحیح است.

ضریب فی برای شدت همبستگی بین دو متغیر دو ارزشی می‌باشد. توجـه شـود در ایـن گزینـه خـی‌دو پاسـخ صحیح نمی‌باشد، به این دلیل که خی‌دو شدت همبستگی را نمی‌سنجد بلکه معنی‌داری اسـتقلال دو متغیـر را نشان می‌دهد.

۴۵- گزینه ۳ صحیح است.

x	y	x^2	y^2	xy
5	4	25	16	20
7	8	49	64	56
9	6	36	36	54
3	2	9	4	6
24	20	119	120	136

$$rxy=\frac{(4\times136)-(24\times20)}{\sqrt{\left[(4\times119-(24)^2\right]}\times\sqrt{\left[(4\times1120-(20)^2\right]}}=0/75$$

$$r=\frac{\sum xy-\frac{\sum x\sum y}{N}}{\sqrt{(\sum x^2-\frac{(\sum x)^2}{N})\times(\sum y^2-\frac{(\sum y)^2}{N})}}$$

۴۶- گزینه ۲ صحیح است.

$$s_{xy}=s_y\sqrt{1-(rxy)^2}$$

در اینجا واریانس ما $\frac{4}{9}$ ؛ (0/44) می‌باشد که برای تبدیل به انحراف استاندارد از آن جذر می‌گیریم.

$$s=\sqrt{0/44}=0/66$$

سپس اعداد را در فورمول خطای استاندارد پیش بینی قرار می‌دهیم.

$$s_{xy}=0/66\sqrt{1-(0/8)^2}=0/39\ ;\ 0/4$$

۴۷- گزینه ۴ صحیح است.

$$df=(c-1)(R-1)$$
$$df=(10-1)(3-1)=18$$

۴۸- گزینه ۱ صحیح است.

خطای برآورد نسبت$=\sqrt{\frac{pq}{n}}$

p: تعداد افراد با علاقه به مطالعه

$$p=\frac{360}{600}=0/6$$

1-**P**: q $\qquad q=1-0/6=0/4$

$$\Rightarrow\sqrt{\frac{0/6\times0/4}{600}}=0/02$$

n :تعداد کل افراد

۴۹- گزینه ۳ صحیح است.

Fo فراوانی مشاهده شده	Fe فراوانی مورد انتظار	Fo-Fe	$(Fo\text{-}Fe)^2$	$\sum\frac{(Fo-Fe)^2}{Fe}$
20	32/9	−12/9	166/54	5/06
150	137/09	12/9	166/54	1/21
40	27/09	12/9	166/54	6/1
100	112/9	−12/9	166/54	1/48
				$x^2=13/8$

ضریب فی$=\sqrt{\frac{x^2}{n}}=\sqrt{\frac{13/8}{310}}=0/21$

بخش دوم: آمار استنباطی

فصل هفتم: آمار استنباطی

- کارکرد آمار استنباطی (برآورد- آزمون فرضیه)
- برآورد پارامترهای جامعه (خطای نمونه‌گیری- محدوده اطمینان- خطای استاندارد میانگین)
- آزمون فرضیه (فرضیه صفر- فرضیه مخالف- آماره آزمون- مقدار بحرانی)
- خطای نوع اول و نوع دوم
- انواع آزمون‌های آماری استنباطی پارامتریک:

الف. آزمون **F**

ب. آزمون **t**

آمار استنباطی

تا اینجا درباره روش‌های آمار توصیفی بحث شد و در این بحث، شاخص‌های مرکزی، شاخص‌های پراکندگی و ضرایب همبستگی و رگرسیون بررسی شد.

حال وقت آن رسیده است که به بررسی آن دسته از روش‌های آماری بپردازیم که از طریق آنها ویژگی‌های یک جامعه از روی قسمتی از آن جامعه به نام نمونه فهمیده می‌شود، به عبارت دیگر روش‌هایی که از طریق داده‌های جمع‌آوری شده پارامترهای جامعه را برآورد می‌کند.

استنباط آماری با دو نوع مسأله سروکار دارد: تخمین پارامترهای جامعه و آزمون فرضیه‌ها.

در استنباط آماری کوشش ما بر این است که ببینیم چگونه می‌توانیم درباره تعداد زیادی از رویدادها بر اساس مشاهده بخشی از آن رویدادها ، نتایجی استخراج کنیم. در واقع شیوه‌های استنباط آماری به هر کوششی برای استخراج نتایج از شواهد فراهم شده به وسیله گروه‌های منظم را نظم می‌بخشد. منطق این شیوه‌ها، برخی از شرایطی را که برای جمع‌آوری شواهد لازم است دیکته می‌کند، و آزمون‌های آماری مشخص می‌کنند که میزان تفاوت مشاهده شده چقدر باید باشد تا بتوان اطمینان داشت که این تفاوت نشان‌دهنده تفاوت‌های واقعی در جمعیتی است که نمونه از آن گرفته شده است.

یک مسأله در استنباط آماری این است که برحسب احتمالات معلوم کنیم که آیا تفاوت‌های مشاهده شده بین دو نمونه ، نشان‌دهنده این است که جمعیت‌های اصلی این دو نمونه نیز واقعاً با یکدیگر تفاوت دارند یا نه. لیکن باید توجه داشت که

هر زمان ما دو گروه نمرات به طور تصادفی به دست بیاوریم، این دو نمونه تا حدودی با یکدیگر تفاوت دارند. این تفـاوت معمولاً به خاطر جمع‌آوری اطلاعات و شانس و تصادف به دست می‌آید. با چنین زمینه‌ای حال باید دید که چگونه مـی-توانیم حدس بزنیم که آیا تفاوت‌های مشاهده شده حاصل شانس و تصادف است یا نه؟ شیوه استنباط آمـاری مـا را قـادر می‌سازد که بر حسب احتمال معلوم کنیم که آیا تفاوت‌های مشاهده شده در حدودی هست که بتوان آنها را به شانس و تصادف نسبت داد یا اینکه تفاوت آنقدر زیاد است که باید گفت نمونه‌های مشاهده شـده از دو جمعیـت متفـاوت گرفتـه شده‌اند.

در ایجاد روش‌های آماری جدید، نخستین فنون استنباط پیدا شده آنهایی بودند که تعداد زیادی فـرض، دربـارهٔ ماهیـت جامعه‌ای که نمونه از آن گرفته شده است عنوان می‌کردند. از آنجا که مقادیر مربوط به جامعه را پارامتر مـی‌گوینـد، ایـن فنون آماری را روش‌های پارامتری می-گویند.

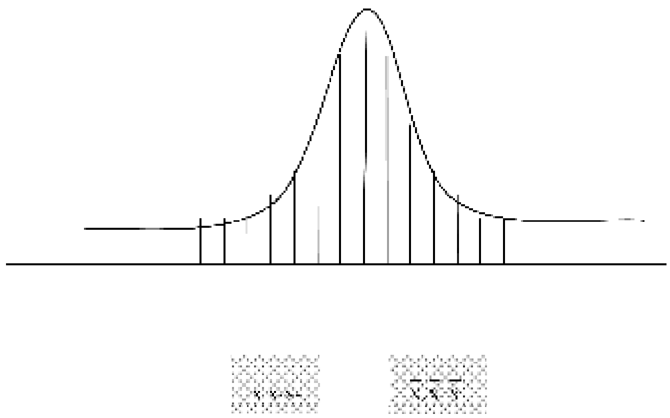
از آنجا که استنباط آماری مستلزم برآورد پارامترهای جامعه از طریق نمونه‌ای است که از جامعـه مـورد مطالعـه انتخـاب می‌شود، بنابراین ابتدا دربارهٔ خطای نمونه‌گیری بحث می‌کنیم.

خطای نمونه‌گیری

چنانچه از جامعه‌ای نمونه‌های متفاوتی را به طور تصادفی انتخاب کنیم، ملاحظه خواهیم کرد که همه آنها دارای ویژگی-های یکسانی نیستند و حتی گاهی اوقات ویژگی‌های آنها مشابه جامعه هم نیست. علت این امر وجـود نـوعی خطـا بنـام «خطای نمونه‌گیری» است.

خطای نمونه‌گیری ضرورتاً نتیجه اشتباه انسانی در نمونه‌گیری نیست. این خطا ممکن است به علت تفاوت‌های فاحش در بین آزمودنی‌ها باشد. به عنوان مثال اگر نمونه‌های متفاوت و نسبتاً زیادی از جامعه‌ای انتخاب کنیم و میانگین‌های آنها را محاسبه کنیم، ملاحظه خواهیم کرد که میانگین‌های محاسبه شده با هم متفاوتند. بعضی از آنها خیلی بزرگ و برخـی از آنها خیلی کوچک و تعداد زیادی از آنها از نظر مقداری در حد متوسط‌اند. چرا این مسأله اتفاق می‌افتد؟

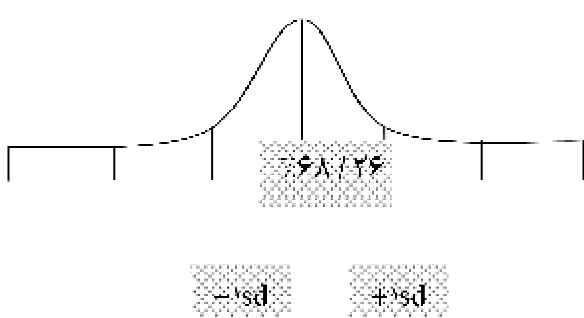
این پدیده را می‌توان با استفاده از قضیه حدّ مرکزی که خود از قانون احتمالات گرفته شده است، توضیح داد. نظریه حـدّ مرکزی بیانگر این مطلب است که اگر از یک جامعه به صورت تصادفی نمونه‌های زیادی را با اندازه‌هـای مسـاوی انتخـاب کنیم و میانگین‌های این نمونه را محاسبه کنیم، توزیع این میانگین‌ها یک توزیع طبیعی خواهد بود و میانگینِ میـانگین-های انتخاب شده تقریباً برابر میانگین جامعه خواهد بود. این نظریه را می‌توان با استفاده از شکل 1-1 نشان داد.



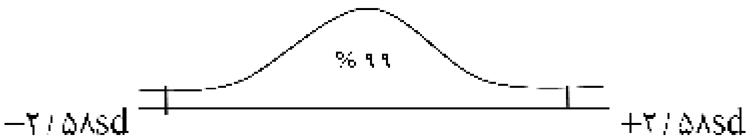
شکل1-1 توزیع میانگین‌های نمونه‌های انتخاب شده

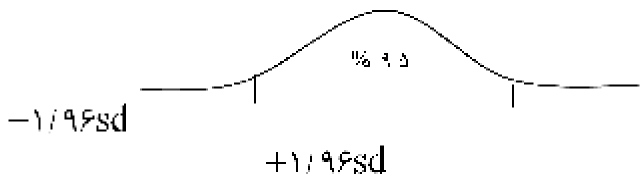
این شکل با توجه به قضیه حدّ مرکزی، توزیع میانگین‌های نمونه‌های انتخاب شده و پراکندگی آنها را در اطراف میـانگین جامعه نشان می‌دهد. همانطور که ملاحظه می‌شود میانگین آن با توجه به یکی از ویژگی‌های جامعه محسوب می‌شود.

به علاوه به دلیل خطای نمونه‌گیری، میانگین نمونه با میانگین جامعه‌ای که نمونه از آن انتخاب شده است، برابر نیست. با توجه به ویژگی‌های توزیع طبیعی می‌دانیم 68/26 درصد احتمال می‌رود نمره‌ها بین ±1 انحراف استاندارد از میـانگین قرار گیرند. به عبارت دیگر می‌دانیم که میانگین یک نمونه بـا 68/26 درصـد اطمینـان بـین ±1 انحـراف اسـتاندارد از میانگین جامعه قرار خواهد گرفت. (شکل زیر



به علاوه میانگین نمونه با 95 درصد اطمینان بین ±1/96 انحراف استاندارد و با 99 درصد اطمینان بین ±2/58 انحراف استاندارد از میانگین جامعه قرار دارد . همچنین با 99/73 درصد اطمینان بین ±3 انحراف استاندارد از میانگین جامعه قرار خواهد داشت.





به علاوه با توجه به شکل می‌توان ملاحظه کرد که میانگین نمونه با 95 درصد اطمینان بین $\pm 1/96$ انحراف استاندارد و با 99 درصد اطمینان بین $\pm 2/58$ انحراف استاندارد از میانگین جامعه قرار دارد ($z = 1/96$ و $z = 2/58$). همچنین با 99/73 درصد اطمینان بین ± 3 انحراف استاندارد از میانگین جامعه قرار خواهد داشت ($z = \pm 3$).

خطای استاندارد میانگین

واضح است که چنانچه انحراف استاندارد توزیع نظری میانگین‌های نمونه را محاسبه کنیم، قادر خواهیم بود درجه اطمینان یا تغییرات را پیش‌بینی کنیم. انحراف استاندارد توزیع نظری میانگین‌ها شاخصی است که به وسیله نمونه‌گیری اندازه‌گیری می‌شود و آن را خطای استاندارد میانگین می‌نامند.

خطای استاندارد میانگین‌ها را با $S_{\bar{x}}$ یا **S.E.D** نمایش می‌دهند و آن را با استفاده از فرمول زیر محاسبه می‌کنند:

$$S_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

$S_{\bar{x}}$ = خطای استاندارد میانگین

s_x = انحراف استاندارد نمونه

n = اندازه یا حجم نمونه

به عبارت دقیق‌تر فرمول خطای استاندارد میانگین برابر است با :

$$S_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}$$

σ_X = انحراف استاندارد جامعه

اما چون در عمل قادر به محاسبه انحراف استاندارد جامعه نیستیم، معمولاً برای محاسبه خطای استاندارد میانگین، از فرمول قبلی استفاده می‌کنیم. به منظور آشنایی با موارد استفاده از فرمول مورد بحث، به مثال زیر توجه کنید:

چنانچه میانگین قد 36 نفر کودک، 75 و انحراف استاندارد قد آنها 18 سانتی‌متر باشد، میانگین قد این 36 نفر در چه فاصله‌ای از میانگین قد کلیه دانش‌آموزانی که این 36 نفر از میان آنها انتخاب شده‌اند، قرار دارد؟

$$s_{\bar{x}} = \frac{S}{\sqrt{N}} \Rightarrow s_{\bar{x}} = \frac{18}{\sqrt{36}} = 3$$

1- میانگین جامعه با 68/26 درصد اطمینان بین 78 و 72 قرار دارد یعنی:

$\bar{x} \pm 1(s_{\bar{x}})$ یا $75 \pm (1)(3)$ یا 75 ± 3

2- میانگین جامعه با 95 درصد اطمینان بین 80/88 و 69/12 قرار دارد یعنی:

$\bar{x} \pm (1/96)(S_{\bar{x}})$

$75 \pm (1/96)(3)$ یا $75 \pm 5/88$

$69/12 \langle \mu \langle 80/88$

3- میانگین جامعه با 99 درصد اطمینان بین 82/84 و 67/62 قرار دارد یعنی:

$\bar{x} \pm (2/58)(S_{\bar{x}})$

$75 \pm (2/58)(3)$ یا $75 \pm 7/74$

$67/26 \langle \mu \langle 82/74$

توجه داشته باشید که میانگین جامعه دارای مقدار ثابتی است و دستخوش تغییر نمی‌شود. آنچه که حدود اطمینان بیان می‌کند، این است که میانگین جامعه ثابت و ناشناخته است و در دامنه معینی قرار دارد.

به آخرین حدود اطمینان توجه کنید. با استفاده از 82/84 - 67/62 می‌توانیم بگوییم یک درصد احتمال یا شانس ($1 = 99 - 100$) وجود دارد که میانگین جامعه در فاصله‌ای بیشتر از 7/74 ($2/58 \times 3 = 7/74$) از میانگین نمونه قرار داشته باشد. می‌توان از سایر محاسبات نتایج مشابهی را به دست آورد.

برآورد

روش‌های آمار استنباطی به منظور برآورد پارامتر جامعه به کار برده می‌شوند. خطای استاندارد شاخص‌های آماری نیز به خاطر دقت در برآورد، محاسبه و بررسی می‌شوند.

چنانچه از جامعه‌ای نمونه‌ای را انتخاب کنیم و میانگین ($\bar{x}$) این نمونه را به منظور برآورد میانگین جامعه (μ) محاسبه کنیم، درباره میانگین جامعه‌ای که نمونه را از آن انتخاب کرده‌ایم، یک گمان یا برآورد خواهیم داشت. بنابراین میانگین جامعه را از طریق میانگین نمونه برآورد می‌کنیم.

آزمون فرضیه

هدف آمار استنباطی عبارتست از نتیجه‌گیری دربارهٔ ویژگی‌های جامعه‌ای که نمونه مورد مطالعه از آن استخراج شده است. در آمار استنباطی یک اصل اساسی که لازم است به آن توجه شود، این است که نتایج بدست آمده از نمونه متفاوت است. در آمار استنباطی دو هدف وجود دارد: 1- برآورد 2- آزمون فرضیه . تا اینجا درباره برآورد به اختصار صحبت کردیم. در این قسمت آزمون فرضیه بررسی خواهد شد :

آزمون فرضیه یک فرآیند استنتاجی است که هدف آن برآورد پارامتر جامعه‌ای که است که نمونه از آن انتخاب شده است. آزمون فرضیه با صورت‌بندی کردن فرضیه‌های آماری آغاز می‌شود. به عبارت دیگر، در آزمون فرضیه، یک ارزش معین در ذهن وجود دارد و فرض می‌شود که این ارزش پارامتر جامعه مورد نظر را برآورد می‌کند.

سؤالی که در اینجا مطرح است این است که آیا ارزش یا مقداری که به کمک اطلاعات جمع‌آوری شده از نمونه، برای پارامتر برآورد جامعه می‌شود، درست است یا خیر.

آزمون فرضیه با بیان فرضیه‌های آماری آغاز می‌شود. فرضیه آماری جمله یا عبارتی است که پیرامون ویژگی‌های جامعه بیان می‌شود و امکان دارد درست نباشد، ولی پژوهشگر صرفاً به خاطر برقرار کردن شرایط قابل آزمایش آن را مطرح می‌کند. به عبارت دیگر فرضیه آماری یک بیان کمّی درباره پارامتر جامعه است و اصولاً بدون فرضیه آماری امکان انجام یک آزمون آماری دشوار است. فرضیه آماری به دو دسته تقسیم می‌شود : الف) فرضیه صفر ب) فرضیه خلاف

فرضیه صفر

فرضیه صفر را با H_0 نشان می‌دهند. این فرض، اصل را بر این قرار می‌دهد که بین پارامترهای مورد مطالعه اختلاف یا ارتباط معناداری وجود ندارد. این فرض به صورت پارامتر صورت‌بندی یا بیان می‌شود. به عنوان مثال فرض کنید قصد آزمون این فرضیه را داشته باشیم که میانگین قد دانشجویان دانشکده افسری برابر 168 سانتی‌متر است. در این مثال فرضیه صفر به صورت زیر بیان می‌شود :

$H_0 = m = 168$ یا $H_0 = \mu - 168 = 0$

فرضیه خلاف

فرضیه خلاف را با H_1 یا H_A نشان می‌دهند. این فرضیه مخالف فرضیه صفر است و در اکثر مـوارد بـا فرضـیه پژهشـی مطابقت دارد. به این معنی که فرضیه خلاف انتظار پژوهشگر را پیرامون نتایج آتی پژوهش بیان می‌کند. فرضیه خلاف نیز همانند فرضیه صفر به صورت پارامتر صورت‌بندی می‌شود. برایِ مثال قبل فرضیه خلاف بدین صورت بیان می‌شود :

$H_1 = \mu - 168 > 0$ یا $H_1 = \mu > 168$

فرمانده دانشکده مورد پژوهش به منظور آزمون فرضیه تدوین شده، در بین دانشجویان به صـورت تصـادفی نمونـه‌ای بـه حجم 120 نفر دانشجو انتخاب می‌کند و قد آنها را اندازه می‌گیرد. چون فرمانده دانشکده، اطلاعات لازم را از نمونه جمع-آوری می‌کند، باید نتایج بدست آمده را به جامعه‌ای که نمونه از آن انتخاب شده اسـت (یعنـی دانشـکده) تعمـیم دهـد. میانگین قد نمونه ممکن است بزرگتر یا کوچکتر از میانگین جامعه باشد و در نتیجه ممکن است او در تصمیم‌گیری خود دچار اشتباه شود، اما از طرفی چون میانگین جامعه را نمی‌داند و از طرف دیگر مجبور به اتخاذ تصمیم اسـت، تصـمیم او ممکن است درست یا اشتباه باشد. به عبارت دیگر در تصمیم‌گیری برای تأیید یا رد فرضیه صفر، دو نوع خطا ممکن است اتفاق بیافتد که به خطای نوع اول و نوع دوم معروفند.

خطای نوع اول :

هدف از اجرای هر آزمون آماری این است که آیا نتیجه تحقیق (متغیر وابسته) ناشی از اجرای متغیر مستقل است یـا بـر اثر عوامل تصادفی ایجاد شده است. در صورتی که فرضیه صفر رد شود، پژوهشـگر بـرای ایـن نتیجـه‌گیـری کـه حاصـل پژوهش او ، ناشی از متغیر مستقل است، دلیلی خواهد داشت. همانطور که در بالا ذکر شد، در جریان تصمیم‌گیری امکان ارتکاب دو نوع خطا وجود دارد. خطای نوع اول وقتی اتفاق می‌افتد که فرضیه صفر درست باشد ولی محقق آن را رد کند. احتمال ارتکاب خطای نوع اول با سطح معنادار بودن آزمون یا درجه اطمینان رابطه مستقیم دارد (سـطح معنـادار بـودن عبارتست از به دست آوردن یک ارزش آماری که به رد فرضیه صفر منجر شود، در شرایطی که این فرضیه درست باشد). احتمال ارتکاب خطای نوع اول را با α نشان می‌دهند، بنابریان مقدار α (آلفا) احتمال ارتکاب این خطا را تعیـین مـی-کند. در نتیجه هنگامی که در یک آزمون آماری مقدار تعیین می‌شود، در حقیقت، احتمال ارتکاب خطای نوع اول تعیـین می‌شود. به عنوان مثال زمانی که سطح معنادار بودن آزمون 0/05باشد، احتمال اینکه فرض صـفر بـه طـور نادرسـت رد شود، 0/05 است. به همین ترتیب اگر مقدار α مساوی 0/01 باشد، یعنی 0/01 احتمال ارتکاب خطای نـوع اول وجـود دارد. بنابراین برای کاهش خطای نوع اول باید α را دقیق‌تر انتخاب کرد.

خطای نوع دوم :

خطای نوع دوم وقتی اتفاق می‌افتد که فرضیه صفر، غلط باشد، ولی پژوهشگر این فرضیه را قبول کند. به عبارت دیگر، قبول یا تأیید فرضیه صفر زمانی که این فرضیه در واقع غلط باشد و باید رد شود، خطای نوع دوم نامیده می‌شود. خطای نوع دوم را با β نشان می‌دهند. با توجه به مطالب فوق، در صورتی که خطای نوع اول افزایش پیدا کند، احتمال ارتکاب خطای نوع دوم کاهش پیدا می‌کند. غالب پژوهشگران سعی می‌کنند که احتمال ارتکاب خطای نوع اول را تا حد ممکن کاهش دهند، خطای نوع اول و دوم را می‌توان به صورت زیر نشان داد :

	H_0 درست است	H_1 درست است
قبول H_1	α	تصمیم صحیح
قبول H_0	تصمیم صحیح	β

هنگامیکه H_1 را فرضیه درستی است قبول می‌کنیم تصمیم صحیح اتخاذ شده است. به همین ترتیب هنگامیکه H_0 درست است و ما آن را می‌پذیریم تصمیم ما درست بوده است.

احتمال ارتکاب خطای نوع دوم به چهار عامل بستگی دارد :

۱- سطح معنادار بودن :

احتمال خطای نوع دوم به مقدار سطح معنادار بودن آزمون آماری بستگی دارد. تغییر در مقدار α در تعیین محدودهٔ ناحیه رد که برای آزمون آماری انتخاب می‌شود، تأثیر دارد. در صورتیکه مقدار α کاهش پیدا کند، به عنوان مثال از 0/05 به 0/01 برسد، ارزش بحرانی آزمون آماری یا ناحیه رد فرضیه صفر، افزایش پیدا کرد.

۲- اندازه تأثیر متغیر مستقل :

بین تأثیر متغیر مستقل و خطای نوع دوم، رابطه معکوس وجود دارد.

۳- مقدار پراکندگی موجود در متغیر وابسته :

احتمال ارتکاب خطای نوع دوم به مقدار پراکندگی موجود در نمره‌های متغیر وابسته بستگی دارد. بنابراین برای کاهش احتمال ارتکاب خطای نوع دوم و در نتیجه افزایش توان آزمون، پژوهشگر باید سعی کند پراکندگی متغیر وابسته را به حداکثر مقدار (یا پراکندگی متغیرهای خطا را به حداقل) برساند.

۴- اندازه حجم نمونه :

اندازه نمونه در احتمال ارتکاب خطای نوع دوم تأثیر دارد. برای روشن شدن این تأثیر به خاطر بیاورید که قبلاً گفته شد در آمار استنباطی پژوهشگر به کمک میانگین نمونه ($\bar{x}$) میانگین جامعه (μ) را برآورد می‌کند. به خاطر بیاورید که پژوهشگر در این برآورد مرتکب خطا می‌شود که به آن خطای استاندارد میانگین نیز گفته می‌شود. مقدار این خطا با استفاده از فرمول زیر محاسبه می‌شود : $S_{\bar{x}} = \frac{s}{\sqrt{n}}$

همانطور که از فرمول فوق استنباط می‌شود بین اندازه نمونه و خطای مورد بحث ، رابطه معکوس وجود دارد. به این معنی که با افزایش اندازه نمونه، خطای استاندارد میانگین کاهش می‌یابد. بنابراین در صورتیکه اندازه نمونه به اندازه کافی بزرگ باشد، تأثیر متغیر مورد آزمایش را می‌توان تعیین کرد.

با توجه به مطالب مطرح شده در مورد خطای نوع اول و دوم نتایج زیر حاصل می‌شود :

1- بین خطای نوع اول و نوع دوم رابطه معکوس وجود دارد .

2- اندازه خطای نوع اول با تنظیم α یا مقدار بحرانی کاهش می‌یابد.

3- افزایش حجم نمونه مقادیر α و β را همزمان کاهش می‌دهد.

توان آزمون

توان یک آزمون آماری برابر است با احتمال رد H_0 هنگامی که این فرضیه واقعاً غلط است.. به عبارت دیگر توان آزمون مساوی است با احتمال قبول H_1 وقتی این فرضیه واقعاً درست است. توان آزمون مساوی است با $1-\beta$، چون بین β و توان آزمون رابطه وجود دارد، باید توجه داشت که هر عاملی یا شرایطی که موجب شود β کاهش یابد، توان آزمون را افزایش خواهد داد.

سطح معنی‌دار بودن

احتمال ارتکاب خطای نوع اول یا α را سطح معنادار‌بودن آزمون می‌نامند. انتخاب این سطح اختیاری است. ولی غالباً پژوهشگران سطوح 0/05و0/01 را به عنوان سطوح قابل قبول انتخاب می‌کنند. بعد از اینکه مشخص شد دو متغیر رابطه دارند باید دید این رابطه‌ای که در نمونه مشاهده شده‌است، به همان شدت در جمعیتی که نمونه از آن گرفته شده‌است، نیز وجود دارد یانه؟ باید دید می‌توان نتایج حاصله از نمونه را با اطمینان تعمیم داد یا نه؟. بدین منظور از آزمون‌های معنی‌داری سود می‌جوییم. منطقی که در پس این آماره‌ها نهفته‌است، چیست؟ معمولاً ابتدا فرض می‌کنیم در جمعیت بین دو متغیر رابطه‌ای وجود ندارد(یعنی $r=0$) این فرضیه را فر ضیه صفر می‌گیریم. چنانچه مشاهده شد در نمونه بین دو متغیر رابطه‌ای وجود دارد(مثلاً $r=0/40$) تفاوت فرض عدم رابطه در جمعیت با رابطه مشاهده شده در نمونه به دو صورت قابل تفسیر است:

1- نمونه ما نمونه معرف نیست.

2- فرض عدم رابطه در جمعیت(H_0)فرض غلطی است.

اگر تفسیر اول را بپذیریم و حاصل نمونه را($r=0/40$) را ناشی از خطای نمونه‌گیری بدانیم، آن‌گاه می‌گوییم بعید است که رابطه این دو متغیر در جمعیت از صفر تجاوز کند(یعنی صفر را قبول می‌کنیم) و فرض می‌کنیم رابطه مشاهده شده در نمونه($r=0/40$) ناشی از خطای نمونه‌گیری است. اما اگر تفسیر دوم را بپذیریم و فرضیه اولیه عدم رابطه در جمعیت بزرگتر را غلط بدانیم، نتیجه حاصل از نمونه را بازتاب رابطه واقعی موجود در جمعیت تفسیر کنیم، چنان عمل کنیم که گویی رابطه مشاهده شده در نمونه واقعی است و محصول نمونه غیر معرف نیست. با کمک آزمون‌های معنی‌داری می‌توان پی برد کدام تفسیر درست است. منطق این آزمون‌ها ساده است. اگر دو متغیر در جمعیت فاقد رابطه باشند، احتمال اینکه نمونه تصادفی ما مبین رابطه‌ای بین این دو متغیر باشد چقدر است؟(یعنی احتمال دقیق نبودن نمونه تصادفی ما چقدر است؟) به عنوان مثال اگر صد بار نمونه‌گیری تصادفی انجام دهیم احتمال اینکه یکی از آنها نمونه غیر معرف باشد، یعنی بیانگر رابطه‌ای باشد که واقعاً در جمعیت مرجع نمونه وجود ندارد، چقدر است؟ معمولاً گفته می‌شود آنجا که احتمالاً از هر صد نفر نمونه بیشتر از 5 نمونه مبین رابطه‌ای باشند که ناشی از خطای نمونه‌گیری است، احتمال نادرست بودن نمونه بالاست. چه بسا نمونه خاص ما یکی از این 5 نمونه باشد. لاجرم باید گفت که به احتمال

زیاد رابطه مشاهده شده دو متغیر درنمونه، ناشی از خطای نمونه‌گیری است و فرض فقدان رابطه در جمعیت مرجع واقعی است. با کمک نظریه احتمالات می‌توان احتمال واقعی نبودن رابطه مشاهده شده در نمونه را(یعنی احتمال اینکه رابطه مشاهده شده در نمونه ناشی از خطای نمونه‌گیری باشد) برآورد کرد. آزمون معنی‌داری آماری در واقع برآورد همین احتمال است. دامنه مقادیر این آزمون‌ها از 0/0000تا 1/0000است و آنها را سطوح معناداری می‌نامیم. معنای این ارقام چیست؟ گیریم سطح منی‌داری 0/5 است. این بدان معنی است که در 50 نمونه از هر 100 نمونه فقط براثر خطای نمونه‌گیری رابطه‌ای به قوت رابطه‌ای که ما در نمونه خاص خود مشاهده کردیم، دیده می‌شود. در این صورت احتیاط حکم می‌کند رابطه مشاهده شده در نمونه خود را به احتمال زیاد واقعی ندانیم، فرض عدم رابطه در جمعیت تأیید می‌شود.

اگر سطح معنی‌داری 0/05 باشد بدین معنی‌است که فقط 5 نمونه از هر 100 نمونه بر حسب تصادف به رابطه مشاهده شده در نمونه ما، می‌کشد. اگر سطح معنی‌داری 0/01 باشد، به معنی واقعی نبودن رابطه در یک نمونه از 100 نمونه است. پیداست هر چه سطح معنی‌داری پائین‌تر باشد، می‌توان اطمینان بیشتری به واقعی‌بودن رابطه مشاهده شده در نمونه داشت. در گزارش‌های تحقیقی سه سطح معناداری مورد استفاده قرار می‌گیرد: 0/05، 0/01،0/001، اینها به معنای آنند که شانس‌های به دست آوردن پیوستگی ناشی از خطای نمونه‌گیری به ترتیب عبارتند از $\frac{5}{100}$، $\frac{1}{100}$ و $\frac{1}{1000}$. به خاطر داشته باشید آزمون‌های معنی‌داری آماری این احتمال را می‌-سنجند که ارتباط بین متغیرها صرفاً ناشی از خطای نمونه‌گیری باشد، اگر نمونه‌گیری در کار نباشد، خطای نمونه‌گیری هم وجود نخواهد داشت.

تفسیر تأیید یا رد فرض صفر

تأیید شدن فرض صفر به این معنی است که مدارک کافی برای اینکه نشان دهد که این فرض غلط است، وجود ندارد. تأیید فرض صفر به این معنی نیست که این فرض واقعاً درست است. هنگامی که فرض صفر رد می‌شود معنی آن این است که بین شاخص‌های آماری مورد مقایسه از نظر آماری اختلاف معنی‌داری وجود دارد. به عبارت دیگر چنانچه فرض صفر واقعاً درست باشد، احتمال اینکه اختلاف بین شاخص‌های مورد مقایسه از روی شانس باشد برابر α (خطای نوع اول) است.

انواع آزمون‌های آماری

در این قسمت درباره آزمون‌هایی که به منظور تعیین معنادار بودن تفاوت بین میانگین‌ها به کار برده می‌شود، بحث خواهیم کرد. این آزمون‌ها عبارتند از:

1- آزمون **t**

2- آزمون **f**

مقایسه میانگین‌ها

چنانچه داده‌های مربوط به متغیر وابسته از نوع کمّی با مقیاس فاصله‌ای یا نسبی و داده‌های متغیر مستقل یا گروه‌بندی از نوع کیفی با مقیاس اسمی یا ترتیبی باشد، برای بررسی تفاوت‌ها می‌توان به مقایسه میانگین‌ها پرداخت و معنی‌دار بودن تفاوت‌های موجود بین طبقات یا گروه‌ها را مورد بررسی قرار داد. برای این منظور دو روش پر کاربرد تحت عنوان آزمون**t** وآزمون**f** وجود دارد. چنانچه متغیر مستقل یا متغیر گروه‌بندی تنها دو گروه داشته‌باشد(به عنوان مثال اگر بخواهیم در آمد زنان و مردان را با هم مقایسه کنیم) در این صورت باید از آزمون **t** استفاده کرد. امّا اگر تعداد گروه‌ها بیشتر از دو باشد؛ در این صورت آزمون**t** کاربرد نخواهد داشت و باید آزمون **f**که به تحلیل واریانس یا ANOVA نیز معروف است، استفاده شود.(به عنوان مثال اگر بخواهیم میزان درآمد گروه‌های شغلی کارگر، کارمند و کشاورز را با همدیگر مقایسه کنیم، باید از آزمون **f** استفاده کنیم)

انتخاب آزمودنی‌های مناسب برای مقایسه میانگین‌ها

آزمون‌های **t** و**f** عمده‌ترین آزمون‌های آماری برای مقایسه میانگینِ گروه‌ها می‌باشد. از آنجا که گروه‌های مورد بررسی ممکن است مستقل یا همبسته(از یک گروه ثابت دو آزمون بگیریم) باشند، بنابراین هر یک از آزمون‌های فوق به دو بخش مستقل و همبسته تقسیم می‌شوند. تصمیم‌گیری در خصوص اینکه در چه مواقعی باید از آزمون‌های **t**یا **f** مستقل یا همبسته استفاده کرد، مهم‌ترین مسأله در تحلیل داده‌های کمّی است. برای این منظور درخت تصمیم‌گیری در شکل زیر ارائه شده‌است که به محقق کمک می‌کند تا مناسب‌ترین آزمون آماری را انتخاب و به کار گیرد.

سوال اول ⇐ آیا گروه‌های مورد مطالعه مستقل هستند یا همبسته؟

سوال دوم ⇐چند گروه یا متغیر را می‌خواهید با یکدیگر مقایسه کنید؟

بر این اساس اگر محقق تنها دو سوال را مطرح و پاسخ‌های لازم را ارائه دهد، قادر بـه انتخـاب آزمـون مناسـب خواهد بود. پاسخ هر یک از سوالات فوق و آزمون مناسب برای آنها در شکل بالا آمده است.

پیش‌فرض‌های آزمون‌های پارامتری(fوt)

آزمون‌های پارامتری **f**و**t** را زمانی می‌توان مورد استفاده قرار داد که پیش فرض‌های ذیل در خصوص داده‌هـای مورد بررسی صادق باشد. این پیش فرض‌ها عبارتند از:

1- مشاهدات از یک جامعه نرمال انتخاب شده باشد.

2- اطلاعاتی که باهم مقایسه می‌شوند، باید تقریباً واریانس یکسانی داشته باشند. اگر گروه‌های مورد بررسـی اندازه یکسانی داشته باشند در این صورت این فرض چندان مهم نیست.

3- داده‌های گردآوری شده دارای مقیاس فاصله‌ای یا نسبی باشند.

تحلیل واریانس

واریانس، معدل مجذورات انحراف مجموعه‌ای از اندازه‌ها از میانگین است. مقدار واریانس هر چه کوچکتر باشـد، به همان نسبت انحراف اندازه‌ها از میانگین، کمتر است. ترکیب اساسی تحلیـل واریـانس را مجمـوع مجـذورات تشکیل می‌دهد. شاخص مجموع مجذورات را با $\sum (x-\bar{x})^2$ یا ss_{xi} نشان می‌دهند.

کاربرد تحلیل واریانس

تحلیل واریانس یک روش آماری است که به منظور تحلیل تفاوت بـین میـانگین‌هـای متغیرهـای مقولـه‌ای یـا تفاوت بین میانگین‌های دو یا چند نمونه آماری بکار می‌رود. پیش‌فرض مهم این تکنیک آن است که نمونـه‌هـا از دو یا چند جمعیت آماری بصورت تصادفی انتخاب شده باشند.

فرض صفر در تحلیل واریانس بصورت زیر است:

$$H_0 : \overline{x}_1 = \overline{x}_2 = \overline{x}_i$$

عدم تفاوت میان میانگین‌های صفت مورد مطالعه دلالت بر فرض صفر می‌کند.

تحلیل واریانس زمانی که برای مقایسه میانگین‌های دو نمونه بکار برده می‌شود، معادل آزمون **t** است. بیشترین استفاده از تحلیل واریانس هنگامی است که پژوهشگر قصد مقایسه بیش از دو میانگین را داشته باشد.

در تحلیل واریانس، سطح سنجش متغیر وابسته (**y**) فاصله‌ای است و متغیر و متغیرهای مستقل، اسمی یا ترتیبی هستند. این تکنیک کمک می‌کند که محقق مجبور به تقلیل سطح سنجش متغیر **y** از فاصله‌ای به رتبه‌ای به منظور محاسبه آماره‌های ترتیبی نباشد.

فصل هشتم: آزمونهای بی پارامتری

آزمونهای بی پارامتری یا آزمونهای با توزیع آزاد در موارد زیر به کار می روند:

1- بهنجار بودن توزیع جامعه ای که نمونه از آن انتخاب می شود، معلوم نباشد.

2- متغیرها بصورت اسمی بیان شوند.

3- متغیرها بصورت ترتیبی بیان شوند.

این آزمونها از دقت کمتری برخوردارند و احتمالاً فرضیه صفر را هنگامی که اشتباه است رد نمی کنند.

آزمون مجذور خی(کای)χ^2

آزمون مجذور خی فقط در مورد داده های گسسته یا ناپیوسته یعنی داده های شمرده نشده و نه مقادیر اندازه گیری شده به کار می رود. آزمون مجذور خی آزمون نبودن وابستگی است، به این معنا که یک متغیر از یک متغیر دیگر تأثیر نمی پذیرد و یا رابطه ای بین آن وجود ندارد. صفات مورد مطالعه در آزمون خی دوکمی است. مقدار $\mathbf{x}^2$ از فرمول زیر محاسبه می شود:

$$\mathbf{x}^2 = \sum\left(\frac{(fo-fe)}{fe}\right)^2$$

$$\text{(فراوانی مورد انتظار)}\ \mathbf{F_e} = \frac{\text{(مجموع فراوانی های افقی) (مجوع فراوانی های عمودی)}}{\text{مجموع کل}}$$

- درجه آزادی $(\mathbf{df})$ در این آزمون از فرمول زیر محاسبه می شود:

$$\text{(درجه آزادی)}\ \mathbf{df} = \text{(تعداد ستوها -1) (تعداد ردیف ها -1)}$$

اگر $\mathbf{x_{ob}}^2$ (محاسبه شده) از $\mathbf{x_{er}}^2$ (جدول) با توجه به $\mathbf{df}$ و مقدار a بزرگتر یا مساوی باشد، فرض صفر رد می شود .

در مورد جدول 2×2 که دارای 4 خانه است، از فرمول زیر استفاده می شود. در این مورد محاسبه فراوانی نظری یا مورد انتظار ضرورتی ندارد.

$$\mathbf{x}^2 = \frac{N\left[|AD-BC|\right]^2}{(A+B)(C+D)(A+C)(B+D)}$$

- نکته: درجه آزادی $(\mathbf{df})$ در جدول 2×2، همیشه برابر با یک است.

تصحیح یتیس[5] برای پیوستگی داده ها

در محاسبه مجذور خی برای جدول 2×2 و با یک درجه آزادی در صورتی که فراوانی مورد نظـر در هریـک از خانـه هـا کمتر از 5 باشدفرمول مذکور به صورت زیر خواهد شد:

$$\mathbf{x}^2 = \frac{N\left[|AD - BC| - \frac{N}{2}\right]^2}{(A+B)(C+D)(A+C)(B+D)}$$

تصحیح خی دو

توزیع نمونه گیری خی دو یک منحنی نظری است. در شرایطی که درجات آزادی برابر 1 باشـد، ایـن منحنـی پیوسـته قادر به بر آورد دقیق احتمال واقعی نیست، یعنی $\mathbf{x}^2$ در معرض

خطا قرار می گیرد.

برای اجتناب از این خطا در فرمول خی دو برازندگی تصحیح صورت می گیرد، یعنی:

$$\mathbf{x}^2 = \frac{(|o - E| - 0/5)^2}{E}$$

آزمونهای احتمال نسبت[6]

یکی دیگر از روش های تجزیه و تحلیل داده های اسمی یا طبقه ای بر احتمال نسـبت اسـتوار اسـت. آزمـون احتمـال نسبت در مقایسه با آزمون های مجذور کای دارای 2درجۀ آزادی و یا بیشتر هستندو کمتـر از تعـداد نمونـه هـای مـورد مطالعه تأثیر می پذیرند. با استفاده از آزمون نسبت های احتمال می توان آزمونی را طراحـی کـرد کـه اغلـب ((احتمـال بیشینهχ^2)) نامیده

می شود. چنین آزمونی را می توان هم در مورد داده های چند بعدی و یا جداول توافقی بکار برد.

- فرمول محاسبه $\mathbf{x}^2$ برای داده های یک بعدی:

$$\mathbf{x}^2 = \sum o1_{in}(\frac{o_i}{E_i})$$

---

[5] Yetes

[6] Likehood of ratio tests

در این فرمول o_i فراوانی های مشاهده شده، E_i فراوانی های مورد انتظار در هر خانه، 1_n علامت لگاریتم طبیعی است.

(df)درجه آزادی در این مورد برابر با **k-1** است یعنی تعداد طبقه های متغیر منهای یک.

فرمول محاسبه x^2 برای داده های چند بعدی یا جداول توافقی:

$$X^2 = 2\sum o1_{ijn}\left(\frac{o_{ij}}{E_{ij}}\right)$$

درجه آزادی در این فرمول از رابطه $df = (C-1)(R-1)$ بدست می آید.

شاخص های ارتباط

۱- ضریب توافقی یا ضریب وابستگی (c)

اگر داده ها به صورت جدول توافقی دو یا چند بعدی تنظیم شوند، ضریب توافقی را می توان به عنوان شاخصی از ارتباط بین متغیرها به کار برد که از فرمول زیر محاسبه می شود:

$$c = \sqrt{\frac{x^2}{n+x^2}}$$

۲- ضریب فای (j)

خی دو محاسبه شده از یک جدول توافقی 2×2 را می توان به ضریب فی تبدیل کرد. ضریب فی از فرمول زیر محاسبه می شود:

$$j = \sqrt{\frac{x^2}{n}}$$

۳- ضریب فای کرامر

اشکال ضریب فای یا فی این است که فقط در مورد جداول 2×2 به کار می رود و در مورد جداول توافقی با طبقات بیشتر قابل اجرا نیست. ضریب کرامر این مشکل را بر طرف می سازد.

فرمول آن بصورت زیر است:

$$f c = \sqrt{\frac{x^2}{n(k-1)}}$$

آزمون میانه

این آزمون غیر پارامتری، معنی دار بودن تفاوت بین میانه های دو گروه مستقل را تعیین می کنـد و فرمـول آن بصـورت زیر است:

$$\mathbf{x}^2 = \frac{N\left[\left|AD - BC\right|\right]^2}{(A+B)(C+D)(A+C)(B+D)}$$

آزمون U مان – ویتنی

آزمون **U** مان ویت نی برای آزمودن تفاوت بین دو گروه مستقل که بر اساس نمونه گیری تصادفی از جامعه انتخاب شده اند، طرح ریزی می شود. این آزمون معادل بی پارامتری آزمون پارامتری **t** است. این آزمون از آزمون میانه قویتر اسـت. و می توان آن را در آزمون های یک دامنه ای و دو دامنه ای برای آزمایش فرضیه صفر به کار برد. اندازه های **U** از فرمول های زیر محاسبه می شوند:

الف - $\mathbf{U}_1 = N_1N_2 + \frac{N_1(N_1+1)}{2} - \sum R$

ب - $\mathbf{U}_1 = N_1N_2 + \frac{N_2(N_2+1)}{2} - \sum R_2$ یا $\mathbf{U}_1 = \mathbf{N}_1\mathbf{N}_2 - \mathbf{U}_2$

هنگامی که از جدول **U** مان – ویتنی استفاده می شود باید اندازه ی **U** کوچکتر را در نظر گرفت. اندازه **z** را بـرای **U** می توان از فرمول زیر محاسبه کرد:

$$z = \frac{\mathbf{U} - \frac{N_1N_2}{2}}{\sqrt{\frac{(N_1)(N_2)(N_1+N_2+1)}{12}}}$$

آزمون نشانه

آزمون نشانه را می تـوان بـرای ارزشـیابی تـأثیر یـک روش آزمایشـی ، هنگـامی کـه شـرایط محـدود کننـدۀ خاصـی حکمفرماست به کار برد. فرمول آن بصورت زیر است:

$$z = \frac{(o \pm 0/50) - \frac{N}{2}}{\sqrt{\frac{N}{4}}}$$

آزمون ویل کاکسون

آزمون ویل کاکسون درباره نمونه های همتا و متشکل از جفت های همتا به کار می رود و از آن در مورد نمونه های مستقل استفاده نمی شود. مقیاس سنجش در این آزمون، رتبه ای است. فرمول آن بصورت زیر است:

$$z = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$$

در فرمول بالا **N** تعداد رتبه بندی شده و **T** مجموع رتبه های هم علامت و کوچکتر است.

آزمون کروسکال - والیس[7]

آزمون کروسکال والیس بسط آزمون **U** مان – ویتنی در مورد سه گروه و یا بیشتر است. این آزمون را می توان همتای تحلیل واریانس یک طرفه در مورد داده های بی پارامتری دانست. فرمول آن به شکل زیر است:

$$H = \frac{12}{N(N+1)}\left[\sum \frac{R^2_i}{n_i}\right] - 3(N+1)$$

در این فرمول **N** تعداد کل افراد مورد مطالعه، R_i^2 مجذور مجموع رتبه های هر گروه و n_i تعداد افراد هر گروه است.درجه آزادی در این فرمول **k-1** می باشد.

آزمون نسبت

آزمون نسبت برای آزمون معنی داربودن تفاوت نسبت ها به کار می رود. در این آزمون فرض صفر به این معناست که بین دو نسبت مفروض تفاوت معنی داری وجود ندارد و تفاوت ظاهری بین نسبت ها ناشی از خطای نمونه گیری است.

۱- آزمون تفاوت نسبت صفت در یک جامعه با یک نسبت معین

سؤال محقق در این آزمون این است که آیا نسبت خصیصهٔ مورد نظر در نمونهٔ مورد مطالعه با نسبت خصیصه در جامعه یکسان است یا خیر. برای آزمودن تفاوت نسبت ها دراین آزمون از فرمول زیر استفاده می شود:

$$z = \frac{p - po}{\sqrt{\frac{p(1-p)}{n}}}$$

در این فرمول **p** نسبت صفت مشاهده شده در گروه نمونه، **po** نسبت صفت در جامعه و **n** تعداد افراد گروه نمونه است.

[7] Kruskal wallis tes

۲- آزمون تفاوت نسبت دو صفت دریک جامعه ی واحد

برای آزمون تفاوت این دو نسبت از فرمول زیر استفاده می شود:

$$z=\frac{|p_1-p_2|}{\sqrt{\frac{p(1-p)}{n}}}$$

در فرمول فوق p_1 نسبت صفت اول در بین افراد گروه نمونه، p_2 نسبت صفت دوم دربین افراد گروه نمونه، p متوسط یا میانگین نسبت دو صفت در گروه نمونه و n تعداد افراد گروه نمونه است.

۳- آزمون تفاوت نسبت یک صفت در دو جامعه ی مستقل

فرض صفر در این آزمون بدان معناست که بین نسبت های مشاهده شده در نمونه ها، در جامعه های مورد مطالعه تفاوت معنی داری وجود ندارد. برای انجام این آزمون از فرمول زیر استفاده می شود:

$$z=\frac{|p_1-p_2|}{\sqrt{p(p-1)(\frac{1}{n_1}+\frac{1}{n_2})}}$$

عناصر این فرمول همانند فرمول بالاست و تنها n_2 تعداد نمونهٔ دوم می باشد.

نکته: اگر تعداد نمونه کمتر از 30 نفر باشد بجای آزمون z از آزمون t استفاده می شود:

آزمون تفاوت بین دو نسبت همبسته

الف- طبقه بندی ناسازگار: برای آزمون تفاوت نسبت ها در چنین مواردی از فرمول زیر استفاده می شود:

$$z=\frac{p_1-p_2}{\sqrt{\frac{1}{n}(p_1q_1+p_2q_2+2p_1q_2)}}$$

ب- طبقه بندی همپوش: برای آزمون تفاوت نسبت ها در این مورد از فرمول زیر استفاده می شود:

$$z=\frac{p_1-p_2}{\sqrt{\frac{1}{n}(p_1q_1+p_2q_2+2(p_1p_2-p_{12})}}$$

دراین فرمول p_{12} نسبت کسانی است که هم در گزینه 1 و هم گزینه 2 را انتخاب کرده اند.

آزمون کولموگورف - اسمیرنوف[8]

این آزمون را می توان برای مقایسه دو مجموعهٔ درصدهای حاصل از یک مقیاس درجه بنـدی کـه در مـورد دو نمونـه مستقل به دست آمده است، به کار برد. فرمول این آزمون بصورت زیر است:

$$\mathbf{x}^2 = 4\mathbf{D}^2\left[\frac{n_1 n_2}{n_1+n_2}\right]$$

[8] Kolomogorov Smirnov

تست های فصل هشتم

۱- برای آزمون ارتباط بین دو متغیر مذهب و رضایت از زندگی ((راضی، ناراضی)) از کدام آزمون آماری استفاده می شود؟

الف) آزمون T ب) آزمون f ج) آزمون خی دو د) آزمون فی دو

۲- برای آزمون تفاوت دو میانگین و سه میانگین به ترتیب از کدام آزمون های آماری استفاده میشود؟

الف) F,T ب) T,F ج) T,Z د) Z,T

۳- اگر در مطالعه ای اختلاف فاحشی بین در آمد زنان شاغل و در آمد مردان شاغل مشاهده شود برای تعیین شدت رابطه متغیر جنس با درآمد از چه آماره ای استفاده می کنیم؟

الف) تای گودمن و کراسکال (T) ب) ضریب تعیین (r^2)

ج) گاما (l) د) نسبت همبستگی $(eta)^2$

۴- در تحقیقی مقدار اشتباه (خطا) از نوع اول a را ۰/۳ انتخاب کرده اند. کدامیک از موارد زیر ممکن است رخ دهد؟

الف)پذیرش H_0 آسان می شود.

ب) پذیرش H_A آسان میشود.

ج) پذیرش H_o,H_A هر دو سخت می شود.

د) پذیرش H_o,H_A هر دو آسان می شود.

۵- در یک شرکت ۹ نفره، ۴ نفر طرفدار حزب ((الف)) هستند، ۳ نفر طرفدار حزب ((ب)) و ۲ نفر طرفدار حزب ((ج)). شاخص تغییر کیفی (IQV) گرایش سیاسی این ۹ نفر چقدر است؟

الف) $\frac{6}{27}$ ب) $\frac{8}{27}$ ج) $\frac{12}{27}$ د) $\frac{26}{27}$

۶- کدامیک از فرضیه های زیر می پذیرد که هیچ بستگی میان دو متغیر دجود ندارد؟

الف) فرضیه تحقیق ب) فرضیه صفر ج) فرضیه ثانویه د) فرضیه اولیه

۷- تفاوت آشکار تحلیل یک جانبه و دو جانبه واریانس چیست؟

الف) در تحلیل دو جانبه بیش از یک متغیر وارد می شود.

ب) در تحلیل دو جانبه تنها یک متغیر وارد می شود.

ج) در تحلیل یک جانبه توصیف مدنظر است.

د) در تحلیل یک جانبه تفاوت میان متغیر جنس وسایر متغیر ها مهم است.

۸- خطای نوع اول زمانی اتفاق می افتد که محقق یک فرضیه صفر . . . نماید.

الف) درست را رد ب) درست را قبول ج) نادرست را رد د) نادرست را قبول

۹- در آزمایش خاصی شنونده ای را وارد کرده ایم که دو بار متوالی با چشمهای بسته ۲۰ خط ده سانتی متری رسم کنند. برای مقایسه میانگین های خطوط ترسیمی در دو مرحله از کدام آزمون آماری باید استفاده کنیم؟

الف)t استودنت برای گروههای مستقل ب) تحلیل واریانس یک طرفه

ج) t استودنت برای گروههای وابسته د) آزمون مجذور خی

۱۰- توان یک آزمون آماری عبارت است از احتمال . . . است.

الف) رد فرضیه صفر موقعی که فرضیه مزبور درست

ب) رد فرضیه صفر، موقعی که فرضیه مزبور نادرست

ج) پذیرش فرضیه صفر موقعی که فرضیه مزبور درست

د)پذیرش فرضیه صفرموقعی که فرضیه مزبور نادرست

۱۱- در آزمون ANOVA برآورد واریانس که از پراکندگی میانگین های گروه در اطراف میانگین کل ناشی می شود، عبارت است از :

الف) نسبتF ب)MS_W ج) MS_B د) MS_E

۱۲- برای تحلیل داده های جمع آوری شده از طرح $R_E \; T_1 \times T_2$ / $R_c \; T_1 - T_2$ **کدامیک از آزمون های آماری زیر مناسب تر است؟**

الف) تحلیل واریانس   ب) تحلیل کوواریانس

ج) دو میانگین مستقل   د) دو میانگین همبسته

۱۳- کدام یک از شاخص های آماری زیر غیر پارامتری است؟

الف) مجذورخی   ب) آزمون تی تست   ج) تحلیل واریانس   د) ضریب همبستگی

۱۴- مهمترین جایگزین برای ANOVA درطرح های پیش آزمون – پس آزمون کدام است؟

الف) فریدمن   ب) طرح های بلوکی   ج) تحلیل رگرسیون   د) تحلیل کوواریانس

۱۵- محققی می خواهد نسبت دانشجویان دختر و پسر رشته روان شناسی را با نسبت دختر و پسر در تمام دانشگاهمقایسه و تعیین کند که آیا بین این نسبت ها تفاوت معنی داری وجود دارد یا نه؟ این محقق از چه روشی باید استفاده کند؟

الف) آزمون تی   ب) تحلیل واریانس   ج) ضریب همبستگی   د) مجذور خی

۱۶- میانگین دو گروه آزمایشی و کنترل را می خواهیم مقایسه کنیم. میانگین گروه آزمایشی ۳۶/۵۰ و میانگین گروه کنترل ۳۲/۷۵ است. اگر میزان خطای استاندارد تفاوت دو میانگین ۲/۸۰ باشد، مقدار t چقدر است؟

الف) 1/11   ب) 1/33   ج) 1/875   د) 3/75

۱۷- مناسبترین آزمون آماری برای فرضیه پژوهشی زیر چیست؟

((بین نمره عزت نفس و اضطراب دانش آموزان ابتدایی، رابطه وجود دارد.))

الف) همبستگی   ب) مقایسه دو میانگین همبسته

ج) خی دو   د) مقایسه دو میانگین مستقل

۱۸- برای تعیین اینکه دو متغیر گسسته مستقل از یکدیگر هستند، از کدام آزمون باید استفاده کرد؟

الف) دو واریانس مستقل   ب) دو نسبت مستقل

ج) دو میانگین مستقل   د) مجذور کای

۱۹- برای مقایسه یک نسبت با یک نسبت ادعا شده کدامیک از آزمون های زیر مناسبتر است؟

الف) t ب) χ^2 ج) F د) z

۲۰- در صورتی که اطلاعات جمع آوری شده در یک جدول ۵×۴ تدوین شده باشند و بخواهیم این اطلاعات را با استفاده از χ^{2} (خی دو) تحلیل کنیم، درجات آزادی مساوی با کدام یک از مقادیر زیر می گردد؟

الف) 10 ب) 12 ج) 14 د) 16

۲۱- فرض کنید تفاوت میانگین های دو گروه برابر با ۸/۴ و میزان خطای استاندارد و میانگین $\sqrt{16}$ باشد در این صورت مقدار تی چقدر خواهد بود؟

الف) 2/1 ب) 2/2 ج) 4/2 د) 4/4

۲۲- درجات آزادی یک جدول توافقی ۶× ۳ که داده های ۱۳۶ آزمودنی به صورت تصادفی در آن جایگزین شده اند. کدام است؟

الف) 7 ب) 10 ج) 16 د) 18

۲۳- قاعده تصمیم گیری برای پذیرش فرض صفر (H_0) یا بکارگیری آزمون آماری f عبارت است از اینکه مقدار F مشاهده شده . . . ارزش بحرانی F باشد.

الف) کوچکتراز ب) برابر ج) بزرگتر از د) برابر یا بزرگتر از

۲۴- احتمال رد یک H_0 غلط یا تأیید یک H_0 درست چه نام دارد؟

الف) خطای نوع اول ب) خطای نوع دوم

ج) توان آزمون د) تایید یک فرضیه نادرست

۲۵- برای تعیین این موضوع که مجموعه ای از نسبت های مشاهده شده با نسبت های مورد انتظار متناظر خود تفاوت دارند یا نه، از کدام آزمون باید استفاده کرد؟

الف) برای دو نمونه همبسته ب) تحلیل واریانس

ج) ضریب همبستگی د) مجذور کای(خی دو)

۲۶- در صورتی که نسبت t مشاهده شده برای تفاوت بین دو میانگین تقریباً صفر باشد، می توان نتیجه گرفت که:

الف) ابزار اندازه گیری مورد استفاده از اعتبار (روایی) لازم برخوردار نبوده است.

ب) حجم نمونه خیلی کوچک بوده است.

ج) دو جامعه، میانگین مساوی دارند.

د) دلیل قانع کننده ای مبنی بر وجود تفاوت بین میانگین های دو جامعه به دست نیامده است.

۲۷- در یک تحلیل واریانس یک راهه اگر n_1 =۱۹ و n_2=۲۱ و ۲۳= n_3 ، درجات آزادی درون گروهی مساوی کدام مقدار است؟

الف)2 ب)3 ج) 60 د)63

۲۸- در روش آماری مجذور خی محقق به دنبال تحلیل داده های . . . و . . . است.

الف) سنجشی و کیفی ب) اسمی و شمارشی

ج) کمی و رتبه ای د) فاصله ای و نسبی

۲۹- زمانی که فرضیه ای را در مورد توزیع میانگین های نمونه در اطراف میانگین جامعه آزمون می کنیم، درجه آزادیچقدر است؟

الف) مساوی اندازه نمونه

ب) مساوی اندازه نمونه منهای یک

ج) درجه آزادی ارتباطی با اندازه نمونه ندارد.

د)هرچه قدرنمونه افزایش یابد،درجه آزادی کم میشود.

۳۰- عبارت ((بین نمره دانش آموزان در آزمون ورودی دانشگاه و انگیزه موفقیت رابطه معنا دار وجود ندارد)) چه نوع فرضیه ای است؟

الف) آماری یا تحقیقی ب) جهت دار ج) صفر یا پوچ د) غیر جهت دار

۳۱- در طح ۵×۴×۳ چند متغیر مستقل مورد بررسی قرار می گیرد؟

الف) 3متغیر ب) 4 متغیر ج) 24متغیر د) 60 متغیر

پاسخنامه تست های فصل هشتم

1- گزینه د

2- گزینه الف

3- گزینه د

4- گزینه ب

5- گزینه د

$$IQV = \frac{c(x^2 \sum xy)}{x^2(c-1)} = \frac{3[(9)^2-(4)^2+3^2+2^2]}{9^2(3-1)} = 0/096$$

6- گزینه ب

7- گزینه الف

8- گزینه الف

9- گزینه ج

10- گزینه ب

11- گزینه ج

12- گزینه الف

13- گزینه الف

14- گزینه د

15- گزینه د

16- گزینه الف

17- گزینه الف

18- گزینه د

19- گزینه د

20- گزینه ب

21- گزینه ب

22- گزینه ب

23- گزینه الف

24- گزینه ج

25- گزینه د

26- گزینه د

27- گزینه ج

$\mathbf{df_w} = (19+21+23) - 3 = 60$

28- گزینه ب

29- گزینه ب

30- گزینه ج

31- گزینه الف

فصل نهم: تحلیل واریانس

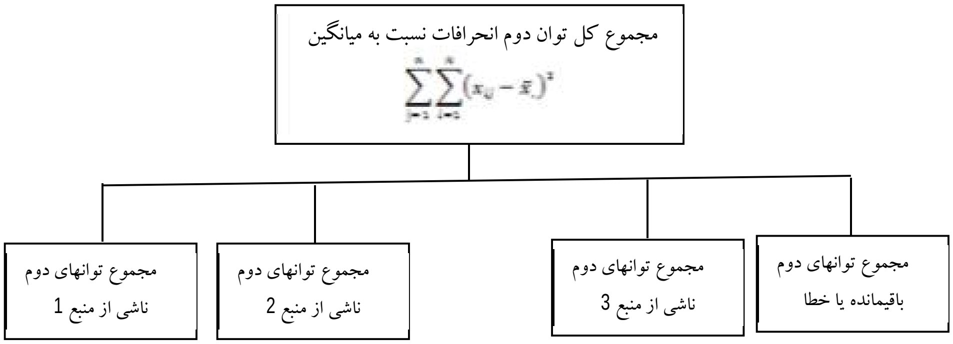
مقدمه

در فصل قبل دربارهٔ تفاوتهای مشاهده شده بین دو میانگین نمونه ای بحث کردیم. در این فصل به بررسی و تحلیل تفاوت بین بیش از دو میانگین نمونه ای می پردازیم. معمولاً به جای اینکه چندین امتحان مقایسه را دو به دو انجام دهند، هم از نظر زمان و هم از نظر هزینه به صرفه تر است که به طور همزمان میانگین ها را با هم مقایسه کنند. در این فصل می خواهیم بررسی کنیم که آیا بین میانگین های نمونه ای که از جامعه های مختلف گرفتیم تفاوتهای واقعی وجود دارد و یا آن مقدار تفاوت قابل اغماض بوده و می توان آن را معلول تصادف دانست. برای مثال می خواهیم بر مبنای اطلاعات نمونه ای قضاوت کنیم آیا بین کارآیی سه روش تدریس ریاضی تفاوت معنی داری هست یا خیر. ممکن است بخواهیم ببینیم آیا بین میزان محصول ناشی از چهار نوع بذر گندم تفاوت چشمگیری وجود دارد یا خیر و یا ممکن است هدف ما مقایسه تفاوت تأثیر سه نوع آگهی در جذب مشتریان بالقوه باشد. روشی که ما برای این منظور استفاده خواهیم کرد ابزار قدرتمند آماری است که ((تحلیل واریانس)) نامیده می شود.

تحلیل واریانس روشی است که می توان میزان انحرافات کل در مجموعهٔ داده ها را به مؤلفه های افراز کرد. هر مؤلفه به دلیل قابل تشخیص بوده، می توان آن را به یک منبع انحراف نسبت داد. علاوه بر این، یک مؤلفه، انحراف حاصل از عاملهای کنترل نشده و خطاهای تصادفی مربوط به اندازه های پاسخها را نشان می دهد. برای مثال در مقایسهٔ انحرافات بین کارایی سه روش تدریس مشخص می شود که چه مقدار از انحرافات، ناشی از تدریس و چه مقدار ناشی از عوامل دیگری چون استعداد تحصیلی دانشجویان، فضای آموزشی و عوامل ناشناخته دیگر است یا در مقایسهٔ میزان محصول دانشجویان، فضای آموزشی و عوامل ناشناخته دیگر است یا در مقایسهٔ میزان محصول ناشی از چهار نوع بذر گندم چه اندازه از انحرافات ناشی از نوع بذر و چه اندازه ناشی از عوامل غیر قابل کنترلی چون خاک، دما، شیب زمین است. یا در تأثیر سه نوع آگهی چه میزان از تفاوت در جذب مشتریان ناشی از نوع آگهی و چه میزان ناشی عوامل دیگری چون زمان آگهی و عوامل تصادفی و غیر قابل کنترل است. در این فصل دربارهٔ بعضی از سؤالات مربوط به طرح آزمایشها بحث می کنیم تا با درجه اطمینان معقولی بتوانیم نتایج معنی دار آماری را به علل (یا عواملی) مشخص نسبت دهیم.

فرض کنید داده های x_{ij} برای $i=1,2,\dots,K$ و $j=1,2,\dots,n$ بوده، میانگین کل آنها $\bar{x}_{..}$ باشد، در این صورت کل تغییر نسبت به میانگین به صورت $\sum_{j=1}^{n}\sum_{i=1}^{K}(x_{ij}-\bar{x}_{..})^2$ مجموع توانهای دوم انحرافات در می آید که مجموع کل توان

دوم انحرافات نامیده می شود. روش تحلیل واریانس این مقدار را به قسمتهایی تجزیه می کنـد کـه در شـکل 5-1 ، سـه منبع تغییر قابل شناسایی بعلاوه مؤلفهٔ خطا مشخص شده است.

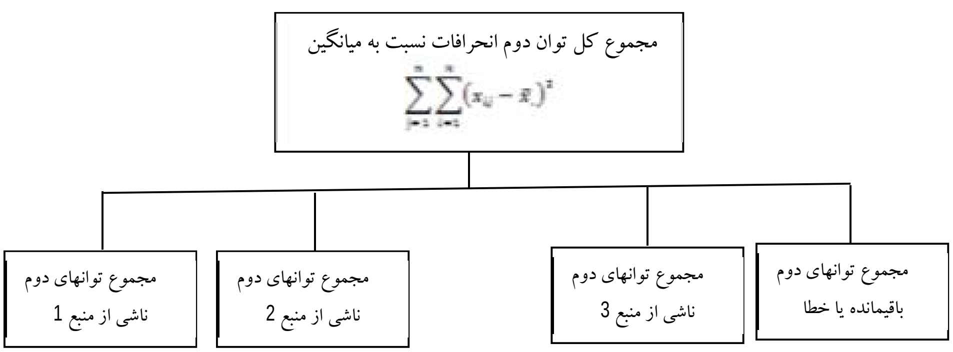


شکل 9-1 افراز تغییرات به چند منبع تغییر

.

تعداد منابع تغییر قابل شناسایی و فرمول مجموعهای توانهای دوم مؤلفه ها اساساً به طرح آزمایشـی کـه در جمـع آوری داده ها به کار رفته و به مدل آماری ای که برای تحلیل مناسب تشخیص داده شده است، بستگی دارد.

چنانچه تحلیلی واریانس بر اساس مشاهداتی صورت گیرد که بر مبنای معیار واحدی – برای مثال نوع بذر- طبقه بنـدی شده اند، آن را تحلیل واریانس یک عامله گویند؛ ولی اگر بر اساس مشاهداتی صورت گیرد که بر مبنای دو معیار – مانند نوع بذر و نوع کود – طبقه بندی شده اند، آن را تحلیل واریانس دو عامله گویند.



شکل9-2 افراز تغییرات به چند منبع تغییر

تحلیل واریانس یک عامله

برای آکه روش تحلیل واریانس یک عامله را تشریح کنیم، مثالی می آوریم. فرض کنید می خواهیم تعـداد ضـایعات سـه ماشین را با هم مقایسه کنیم. ضایعات هر کدام از این ماشینها در پنج روز به این صورت بوده است:

ماشین اول	86	79	81	70	4
ماشین دوم	89	82	88	76	90
ماشین سوم	82	68	73	71	81

میانگین ضایعات این سه ماشین به ترتیب 80 ، 85 و 75 است. می خواهیم بدانیم آیا تفاوت معنی داری بین آنها وجـود دارد یا اینکه می توان اختلاف بین آنها را معلول تصادف دانست.

در حالت کلی، در چنین مسائلی **k** نمونه تصادفی به اندازهٔ **n** از **k** جامعه داریم. مقدار مشـاهدهٔ **j** ام را بـا x_{ij} نشـان مـی دهیم. فرض می کنیم متغیرهای تصادفی متناظر (یعنی x_{ij}) که همه مستقلند، توزیعهای نرمال با میانگین هـای m_i و واریانس مشترک s^2 دارند. با این فرضها هر مشاهده ای را می توان به صـورت $x_{ij} = m_i + e_{ij}$ بـه ازای $i = 1,2,....K$ و $j = 1,2,...,n$ نشان داد که در آن e_{ij}ها مقادیر متغیر تصادفی مستقل نرمال با میانگین صـفر و واریـانس مشـترک s^2 هستند و باری امکان تعمیم آن به وضعیت های پیچیده تر، معمولاً آن را به صورت $X_{ij} = m + a_i + e_{ij}$ نمایش می دهند که به m ((میانگین کل)) و به a_i ((اثر تیماری)) گویند؛ به طوری که $\sum_{i=1}^{K} a_i = 0$ است.

علت نامیدن جامعه های مختلف با عنوان تیمارهای مختلف آن است که بسیاری از تکنیک‌های واریانس، ابتدا، در رابطـه با آزمایشهای کشاورزی مطرح شدند که در آنها، مثلاً کودهای مخلف را به عنوان تیمار در نظر گرفته، به خاک اضافه می کردند. ما هم از این اصطلاح استفاده کرده، هر یک از ماشین ها را یک تیمار می نامیم و تأثیر آنها را روی تعداد ضایعات مشخص می کنیم.

فرض صفری که باشد آن را آزمون کنیم، عبارت است از :

$$H_0 : m_1 = m_2 = ... = m_K$$

یا:

$$H_0 : a_1 = a_2 = ... = a_K = 0$$

یعنی تمام میانگین های جامعه با هم برابرند (و یا همهٔ اثرهای تیماری صفرند). فرض مقابل عبارت است از :

H_1: دست کم دو تا از میانگینها برابر نیستند.

H_1: دست کم یکی از اثرهای تیماری مخالف صفر است.

آزمون مبتنی است بر تحلیل تغییر پذیری کل داده های تلفیق شده که با این فرمول می توان نشان داد:

$$\sum_{j=1}^{n} \sum_{i=1}^{K} (x_{ij} - \overline{x}..)^2$$

که در آن:

$$\overline{x}.. = \frac{1}{nk} \sum_{i=1}^{K} \sum_{j=1}^{n} x_{ij}$$

است. اگر فرض صفر درست باشد، همهٔ این تغییر پذیری ناشی از شانس است؛ اما اگر درست نباشد، بخشی از مجموع توانهای دوم بالا ناشی از اختلافهای بین تیمارها خواهد بود. حال تغییرپذیری (انحرافات) کل فوق را به دو جزء انحرافات حاصل از تیمار و انحرافات باقیمانده (خطا) تفکیک میکنیم:

مجموع توان دو باقیمانده + مجموع توان دوم حاصل از تیمار =مجموع کل توان دوم

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}..)^2 = n.\sum_{i=1}^{K}(\bar{x}_{i.}-\bar{x}..)^2+\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}.)^2$$

که در آن $\bar{x}_{i.}$ میانگین مشاهدات جامعه **i** ام و $\bar{x}..$ میانگین همهٔ **nk** مشاهده است. فرمول فوق را به صورت : $SST = SS(Tr) + SSE$ نشان می دهیم.

در اینجا انحرافات (تغییر پذیری) کل داده های تلفیق شده (**SST**) را به دو جزء تفکیک کردیم: جزء دوم-**SST-** تغییر شانس یعنی تغییرات داخل نمونه ها، و جزء اول-**SS(Tr)**– تغییر تصادفی را وقتی که فرض صفر درست باشد اندازه می گیرد. در ضمن این جزء تغییر بین میانگینهای جامعه ای را وقتی فرض صفر درست باشد اندازه می گیرد. درجه آزادی مجموع توان دوم تیمارها **k-1** و درجه آزادی مجموع توان دوم خطا **k(n-1)** است. در این صورت، میانگین توان دوم تیمارها و میانگین توان دوم خطا که آنها را به ترتیب با **MS (Tr)** و **MSE** نشان می دهیم، به این صورت خواهند بود:

$$MS(Tr) = \frac{SS(Tr)}{k-1}$$
$$MSE = \frac{SSE}{K(n-1)}$$

چنانچه میانگین توان دوم تیمارها نسبت به میانگین توان دوم خطا کم باشد، نتیجه می گیریم که میانگینهای جامعه ها با هم تفاوت ندارند و فرض H_0 (تساوی میانگینها) پذیرفته می شود. چنانچه میانگین توانهای دوم تیمارها نسبت به میانگین توان دوم خطا زیاد باشد، نتیجه می گیریم که میانگینهای جامعه ها با هم تفاوت دارند و فرض H_0 (تساوی میانگین ها) رد می شود. نظریه آماری می گوید طبق فرض صفر، نسبتِ:

$$\frac{\text{میانگین توان دوم تیمار ها}}{\text{میانگین توان دوم خطاها}} = \frac{MS(Tr)}{MSE}$$

که آن را با **F** نشان می دهیم، به این صورت محاسبه می شود:

$$F=\frac{MS(Tr)}{MSE}=\frac{SS(Tr)/(k-1)}{SSE/K(n-1)}$$

که دارای توزیع **F** با درجات آزادی **k-1** برای صورت و **k (n-1)** برای مخرج است. روشی را که در این قسمت ارائه کردیم تحلیل واریانس یک عامله می نامند. معمول است که تجزیهٔ مجموع توانهای دوم و درجات آزادی به همراه میانگینهای توان دوم را به شکل جدولی به نام ((جدول تحلیل واریانس)) یا به صورت ساده تر جدول ((ت.و)) ارائه می کنند. این جدول ستون اضافی دیگری دارد که مقدار **F** را نیز حساب می‌کند.

جدول9- 1جدول « ت. و » یک عامله برای مقایسهٔ **k** تیمار

منبع تغییرات	مجموع توانهای دوم	درجه آزادی منبع	میانگین توانهای دوم	F
تیمارها	$SS(Tr)$	$k-1$	$MS(Tr)=\frac{SS(Tr)}{k-1}$	$\frac{MS(Tr)}{MSE}$
خطا	SSE	$K(n-1)$	$MSE=\frac{SSE}{K(n-1)}$	
جمع	SST	$Kn-1$		

قبلاً گفتیم که $SST=\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}..)^2$ و $SS(Tr)=n.\sum_{i=1}^{K}(\bar{x}_{i.}-\bar{x}..)^2$ و $SSE=\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}.)^2$ است. برای ساده تر کردن محاسبهٔ مجموع توانهای دوم فوق ، معمولاً از این فرمولهای محاسباتی استفاده می شود:

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}{}^2-\frac{1}{Kn}.T..^2$$
$$SS(Tr)=\frac{1}{n}.\sum_{i=1}^{K}T_i^2-\frac{1}{Kn}.T..^2$$
$$SSE=SST-SS(Tr)$$

$T_{i.}$مجموع مقادیر حاصل برای تیمار **i** ام و .. **T** مجموع کلِ **nk** مشاهده است.

مثال - با مراجعه به مثال مبحث 2-2 در سطح معنی دار $a=0/05$ آزمون کنید که آیا اختلاف بین میانگینهای تعداد ضایعات ماشینها معنی دار است یا نه.

1. فرضها: $H_0=a_1=a_2=a_3=0$

دست کم یکی از اثرهای تیماری غیر صفر است:H_1

2. آماره آزمون:

جدول ۹-۲ محاسبات اولیهٔ مثال قبل

ماشینها (تیمارها)	تعداد ضایعات					جمع ضایعات $(T_{i.})$	میانگین
اول	86	79	81	70	84	400	80
دوم	89	82	88	76	90	425	85
سوم	82	68	73	71	81	375	75

بنابراین $T_{1.}=400$ ، $T_{2.}=425$ ، $T_{3.}=375$ ، $T_{..}=1200$ و :

$$\sum_{i=1}^{3}\sum_{j=1}^{5}x_{ij}^{2}=86^2+79^2+\ldots+81^2=96698$$

$$SST=\sum_{i=1}^{3}\sum_{j=1}^{5}x_{ij}^{2}-\frac{T_{..}^{2}}{Kn}$$

$$=96698-\frac{1}{3\times 5}(1200)^2=698$$

$$SS(Tr)=\frac{1}{n}.\sum_{i=1}^{K}T_i^2-\frac{1}{Kn}.T_{..}^{2}$$

$$\frac{1}{5}(400^2+425^2+375^2)-\frac{1}{3\times 5}(1200)^2=250=$$

$$SSE=SST-SS(Tr)$$

$$698-250=448 \quad =$$

همانند جدول9-1 جدولی تشکیل و محاسبات باقیمانده را در آن انجام می دهیم.

جدول 2-3 جدول «ت.و» برای مثال 5-1

منبع تغییرات	مجموع توانهای دوم	درجه آزادی منبع	میانگین توانهای دوم	F
تیمارها	250	$3-1=2$	$\frac{250}{2}=125$	$\frac{125}{37/33}=3/35$
خطا	448	$3(5-1)=12$	$\frac{448}{12}=37/33$	
جمع	698	$3\times 5-1=14$		

3. مقدار بحرانی : داریم $a=0/05$ و $n=5$ و $K=3$ پس:

$$F_{a,(K-1),K(n-1)} = F_{0/05,2,12} = 3/89$$

4. تصمیم گیری : چون $F = 3/35$ از $F_{0/05,2,12} = 3/89$ کمتر است، فرض صفر (فرض تساوی میانگینهای جامعه) در سطح معنی دار 5 درصد را نمی توان رد کرد و اختلاف در میانگینها شانسی و تصادفی است.

در مثال قبل تعداد نمونه گرفته شده از هر جامعه با تعداد نمونه جامعه دیگر برابر بود. اگر به دلیل تعداد نمونه های گرفته شده از جامعه های مختلف برابر نباشد و برای تیمار **i** ام، n_i نمونه در دست باشد، در این صورت خواهیم داشت:

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{x}_{ij}-\bar{x})^2 = \sum_{i=1}^{K} n_i(\bar{x}_{i.}-\bar{x}_{..})^2 + \sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{x}_{ij}-\bar{x}_{i.})^2$$

و درجات آزادی برای **SST** و**SS (Tr)** و **SSE** به ترتیب عبارتند از **N-1** ، **k-1** و **N-k** که در آنها است. همچنین فرمولهای محاسباتی نیز به این صورت خواهند بود:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^{\ 2} - \frac{1}{N}.T_{..}^{\ 2}$$

$$SS(Tr) = \sum_{i=1}^{K}\frac{T_{i.}^2}{n_1} - \frac{1}{N}.T_{..}^2$$

$$SSE = SST - SS(Tr)$$

مثال- در مثال قبل، فرض کنید ضایعات سه ماشین در 5 روز، 2 روز و 3 روز به ترتیب به این صورت باشد:

ماشین اول	75	74	83	85	68
ماشین دوم	60	30			
ماشین سوم	71	71	68		

در سطح معنی دار 5 درصد آزمون کنید که آیا بین میانگین های فوق که به ترتیب 77 ، 45 و 70 هستند، تفاوت معنی داری وجود دارد؟

۱. فرضها:

$$H_0 = a_1 = a_2 = a_3 = 0$$

H_1:دست کم یکی از اثرهای تیماری غیر صفر است

۲. آمارهٔ آزمون:

جدول 9-4 محاسبات اولیهٔ مثال

ماشینها (تیمارها)	تعداد ضایعات					جمع ضایعات $(T_{i.})$	میانگین
1	75	74	83	85	68	385	77
2	60	30				90	45
3	71	71	68			210	70

بنابراین $T_{1.}=385$ ، $T_{2.}=90$ ، $T_{3.}=210$ و $T_{..}=685$ است و :

$$\sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij}{}^2 = 75^2+74^2+\ldots+68^2=49045$$

$$SST=\sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij}{}^2-\frac{1}{N}.T_{..}^2$$

$$=49045-\frac{1}{10}(685)^2=2122/5$$

$$SS(Tr)=\sum_{i=1}^{3}\frac{T_{i.}^2}{n_i}-\frac{1}{N}.T_{..}{}^2$$

$$=(\frac{385^2}{5}+\frac{90^2}{2}+\frac{210^2}{3}+\frac{1}{10}(685)^2=1472/5$$

$$SSE=SST-SS(Tr)$$

$$=2122/5-1472/5=650$$

حال جدول ((ت.و)) را برای این مثال تنظیم می‌کنیم.

جدول 9-5 جدول ((ت.و)) برای مثال حاضر

منبع تغییرات	مجموع توانهای دوم	درجه آزادی منبع	میانگین توانهای دوم	F
تیمارها	$1472/5$	$3-1=2$	$\frac{1472/5}{2}=736/25$	$\frac{736/25}{92/86}=7/93$
خطا	$\frac{650}{7}=92/86$	$10-3=7$	$\frac{650}{7}=92/86$	
جمع	$2122/5$	$10-1=9$		

3. مقدار بحرانی: داریم $k=3$، $N=n_1+n_2+n_3=5+2+3=10$ و $a\ 0/05$ پس:

$$F_{\alpha,K-1,N-K}=F_{0/05,2,7}=4/74$$

4. تصمیم گیری: چون $F = 7/93$ از $F_{0/05,2,7} = 4/74$ بزرگ تر است، فرض صفر (تساوی میانگینهای جامعه یـا صـفر بودن اثرهای تیماری) در سطح معنی دار 5 درصد رد می شود. می توان ادعا کرد که تفاوت بـین میانگینهـا در سـطح 5 درصد معنی دار است.

طرح آزمایشها

در مثال اولین مثال طرح شده در این قصل نتیجه گرفتیم که تعداد ضایعات تولیدی سه ماشین به یک انـدازه هسـتند، ولی این نتیجه گیری شاید منطقی نباشد، چرا که ممکن است، مثلاً تعداد ضایعات ماشین اول بـه دلیـل تجربـه ناکـافیِ متصدی ماشین، نامرغوب بودن قطعات یدکی ای که روی آنها کار کرده، یا حتی ابزارهـایی کـه قطعـات را انـدازه گیـری کرده و در هنگام تعیین ضایعات این ماشین از تنظیم خارج شده و تعداد ضایعات را بیشتر نشان داده است، باشد.

البته امکان دارد که اختلاف بین سه میانگین نمونه ای عمدتاً معلول نوع ماشین باشد، هرچند ممکن است عوامـل دیگـر نیز در این اختلاف دخیل باشند. به خاطر داشته باشید که آزمون معنی دار می توانـد نشـان دهـد کـه اختلافهـای بـین میانگین های نمونه ای بزرگ تر از آن هستند که بتوان آنها را معلول تصادف دانسـت، امـا نمـی توانـد بگویـد چـرا ایـن اختلافها پیش آمده اند.

به طور کی اگر بخواهیم عاملی را، از بین عوامل مختلف، علت پدیده ای بـدانیم بایـد تـأثیر عوامـل دیگـر را بـه حـداقل برسانیم. در این صورت است که می توان گفت، آزمایش قویاً عوامل دیگر را به حداقل برسانیم. در این صـورت اسـت کـه می توان گفت، آزمایش قویاً کنترل شده است. مثلاً در خصوص اولین مثال فصل می توانیم از متصدی ماشـین بخـوهیم به تناوب روی سه ماشین کارکند، قطعات ورودی را از لحاظ مرغوبیت کنترل کنیم تا مطمئن شویم ورودی هر دستگاه با دیگری برابر است، و ابزار اندازه گیری را پس از هر بار استفاده، بازرسی و در صورت لزوم تنظیم کنیم. اگر بتوانیم به ایـن صورت تأثیر عوامل مختلف را (بجز عامل مورد نظر) حذف کنیم آنگاه اگر اختلاف معنی داری بین میانگینها باشد، قابل استناد یه عامل مورد نظر است.

در اغلب موارد، آزمایشهای ((قویاً کنترل شده ای)) نظیر آنچه در بالا توصیف شـد، غیـر عملـی اسـت و یـا نمـی توانـد اطلاعات مورد نظر را در اختیار ما قرار دهد؛ بنابراین باید دنبال راه دیگری بود. این راه که نقطه مقابل روش فـوق اسـت، ((تصادفی کردن)) است؛ یعنی آزمایشها چنان طراحی شود که تغییرات ناشی از عوامل غیر مربوط بـه صـورت تصـادفی توزیع شود. با تصادفی کردن، خود را در مقابل عوامل غیر مربوط محافظت می کنیم. وقتـی کـه همـهٔ تغییـرات ناشـی از عوامل غیر مربوط را بتوان بدین سان به صورت یکنواخت توزیع کرد، به این طـرح، ((طـرح آزمایشـی کـاملاً تصـادفی))

گویند. مثلاً به صورت تصادفی متصدیان یا قطعات ورودی را به ماشینها اختصاص داد و یا انتخاب ترتیب بازرسی کالاها به صورت تصادفی انجام گیرد.

برای معرفی مفهوم مهم دیگری در طرح آزمایشها، داده هایی مربوط به زمان لازم (برحسب دقیقه) برای شخصی را در نظر می گیریم که با اتومبیل خود از شنبه تا چهارشنبه با چهار مسیر مختلف به سر کار می رود:

مسیر اول	22	26	25	25	31
مسیر دوم	25	27	28	26	29
مسیر سوم	26	29	33	30	33
مسیر چهارم	26	28	27	30	30

میانگینهای این چهار نمونه عبارتند از 25/8 ، 27 ، 30/2 ، 28/2 و چون اختلافهای بین آنها نسبتاً زیاد است، منطقی است نتیجه بگیریم که بین میانگین زمان لازم برای آنکه شخصی با رانندگی از چهار مسیر مختلف به سر کار برود، اختلاف وجود دارد؛ ولی این امر از تحلیل واریانس یک عامله نتیجه نمی شود، چرا که مقدار $F = 2/80$ را به دست می آوریم و چون این مقدار از $F_{0/05,3,16} = 3/24$ بیشتر نیست، فرض صفر را نمی توان رد کرد.

البته فرض صفر ممکن است درست باشد، ولی ملاحظه می کنید که نه تنها بین چهار میانگین، بلکه بین مقادیرِ خود نمونه ها هم اختلافهای زیادی وجود دارد. در نمونه اول این مقادیر از 22 تا 31، در نمونه دوم از 25 تا 29، در نمونه سوم از 26 تا 33 و در نمونه چهارم از 26 تا 30 تغییر می کند. بعلاوه، در هر نمونه مقدار اول کوچک تر از همه و مقدار آخر بزرگ تر از همه است. از مطلب اخیر چنین بر می آید که تغییر در داخل نمونه ها ناشی از اختلاف شرایط رانندگی در مجموع توان دوم خطای تحلیل واریانس یک عامله منظور شده و مخرج آمارۀ **F** بزرگ شده است و این می تواند دلیل آن باشد که چرا نتایج معنی دار نبوده اند.

برای احتراز از چنین وضعی، می توان عامل غیر مربوط را ثابت گرفت، ولی با این که بندرت به اطلاعاتی که لازم داریم می رسیم. در مثال فوق، می توانستیم مطالعه را به شرایط رانندگی در روز شنبه محدود کنیم، ولی در این صورت نمی توانستیم مطمئن باشیم که می توان نتایج را به شرایط رانندگی در یکشنبه یا روز دیگر تعمیم داد. امکان دیگر آن است که عامل غیر مربوط را به عنوان یک عامل مهم در نظر گرفته، آزمایش را طوری طراحی کنیم که بتوانیم اثر آن را اندازه بگیریم. در این صورت ((تحلیل واریانس دو عامله)) به کمک ما می آید که در آن تغییرات کل داده ها به سه جزء افراز می شود: تیمارها (در مثال مذکور، چهار مسیر)، عامل غیر مربوط (در مثال مذکور، شرایط رانندگی در روزهای مختلف هفته) و خطای آزمایش یا تصادف.

روشی که پیشنهاد کردیم ((بلوک بندی)) نام دارد و به روزهای مختلف هفته عنوان بلوک اطـلاق مـی شـود. در حالـت کلی، بلوکها سطوحی هستند که عامل غیر مربوط را در آن ثابت می گیریم تا بتوانیم سهم آن را در مجموع کل تغییـرات داده ها اندازه بگیریم.

تحلیل واریانس دو عامله

همان طور که گفتیم، تحلیل واریانس دو عامله به بررسی اثر دو عال در ایجاد تغییرات می پردازد؛ مثلاً عامل نـوع بـذر و نوع کود در میزان محصول، عامل نوع ماشین و متصدی ضایعات، عامل مسیر و روز در میزان صرف شـده بـرای رفـتن از خانه به محل کار.

مفهوم اساسی که درتحلیل واریانس دو عامله باید معرفی شود ((تأثیر متقابل)) است. تأثیر متقابل در آزمـایش دو عاملـه بدین معناست که دو تیمار مستقل نیستند و اثر خاص سطح مختلف تیمار در یک عامل بر اسـاس سـطوح عامـل دیگـر تغییر می کند. مثلاً اثر بخشی روشهای مختلف تدریس وابسته به سطوح استعداد دانشجویان اسـت، یـا انجـام کـار یـک متصدی روی ماشین خاصی باعث افزایش ضایعات و در ماشین دیگر باعث کاهش ضایعات می شود. در این حالتهـا مـی گوییم که بین روشهای تدریس سطوح توانایی دانشجویان و یا متصدی و نوع ماشین تأثیر متقابل وجود دارد.

تحلیل واریانس دو عمله بدون تأثیر متقابل

در تحلیل وارینس دو عامله می توانیم دو متغیر (عامل) را تیمارها و بلوکها – یا عامل $\mathbf{A}$ و عامل $\mathbf{B}$– بنامیم و تیمارهـا را در سطر و بلوکها را در ستون نشان دهیم. فرض کنیـد x_{ij} بـه ازای $i=1,2,...,k$ و $j=1,2,...,n$ مقـادیر $\mathbf{n}$ متغیـر تصادفی مستقل نرمال با میاگینهای m_{ij} و واریانس مشترک s^2 باشند؛ این ماتریس را که قصد داریم تحلیل واریانس دو عامله (بدون تأثیر متقابل) را روی آن انجام دهیم در نظر بگیرید:

x_{11}	x_{12}	...	x_{1n}
x_{12}	x_{22}	...	x_{2n}
$\vdots$	$\vdots$	...	$\vdots$
x_{K1}	x_{K2}	...	x_{Kn}

هر مشاهده را می توان به این صورت نوشت:

$$x_{ij} = m + a_i + b_j + e_{ij}$$

که در آن m میانگین کل، a_i اثرهای تیماری، $\sum_{i=1}^{K} a_i = 0$ است. همچنین b_j اثرهای بلوکی و $\sum_{i=1}^{K} b_j = 0$ است. نیز e_{ij} نشان دهندهٔ مقادیر متغیرهای تصادفی مستقل نرمال با میانگین صفر و واریانس مشترک s^2 است.

دو فرض صفری که آزمون خواهیم کرد، عبارتند از صفر بودن اثرهای تیماری و اثرهای بلوکی، یعنی :

$$H_0 : a_1 = a_2 = ... = a_K = 0$$

و:

$$H'_0 : b_1 = b_2 = ... = b_n = 0$$

فرض مقابل H_0 آن است که همهٔ اثرهای تیماری برابر صفر نیستند و فرض مقابل H'_0 آن است که همهٔ اثرهای بلوکی صفر نیستند، یعنی :

$H_1 : a_i \neq 0$ ، **i** دست کم به ازای یک مقدار

$H'_1 : b_j \neq 0$ ، **j** دست کم به ازای یک مقدار

می توان براحتی اثبات کرد که :

مجموع توان دوم انحرافات حاصل از بلوک +مجموع توان دوم انحرافات حاصل از تیمار =مجموع توان دوم انحرافات

+ مجموع توان دوم انحرافات باقیمانده

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})^2 = n.\sum_{i=1}^{K} n_i(\bar{x}_{i.}-\bar{x}_{..})^2 + K.\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.})^2 + \sum_{i=1}^{K}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.}-\bar{x}_{j.}+\bar{x}_{..})^2$$

$\bar{x}_{i.}$ میانگین مشاهدات برای تیمار **i** ام، $\bar{x}_{.j}$ میانگین مشاهدات برای بلوک **j** ام، و $\bar{x}_{..}$ میانگین همهٔ **nk** مشاهده است؛ یا:

$$SST = SS(Tr) + SSB + SSE$$

تنها تفاوت فرمول فوق با اولین فرمول صفحه 149 در آن است کـه اثـر بلـوکی در انحرافـات را از مجمـوع توانهـای دوم باقیمانده به دست آوردیم. محاسبهٔ **SST** و **SS (Tr)** همانند قبل است، ولی **SSE** فرق می کند. فرمولهای محاسـباتی برای تفکیک انحرافات به سه جزء (بدون تأثیر متقابل) بدین ترتیب است:

$$SST = \sum_{i=1}^{K}\sum_{j=1}^{n} x_{ij}^{\ 2} - \frac{1}{Kn}.T_{..}^2$$

$$SS(Tr) = \frac{1}{n}.\sum_{i=1}^{K} T_{i.}^2 - \frac{1}{Kn}.T_{..}^2$$

$$SSB = \frac{1}{K}.\sum_{J=1}^{n} T_{.j}^2 - \frac{1}{Kn}.T_{..}^2$$

که در آن $T_{.j}$ مجموع مقادیر بلوک **j** ام و $T_{..}$ مجموع کل تمام **nk** مشاهده است، همچنین:

$$SSE = SST - SS(Tr) - SSB$$

در این صورت جدول ((ت.و)) به صورت جدول در می آید.

جدول «ت.و» برای تحلیل وارانس دو عامله (بدون تأثیر متقابل)

منبع تغییرات	مجموع توانهای دوم	درجه آزادی منبع	میانگین توانهای دوم	F
تیمارها	$SS(Tr)$	$k-1$	$MS(Tr)=\dfrac{SS(Tr)}{k-1}$	$\dfrac{MS(Tr)}{MSE}$
بلوکها	SSB	$(n-1)$	$MSB=\dfrac{SSB}{n-1}$	$F_B=\dfrac{MSB}{MSE}$
خطا	SSE	$K(n-1)$	$MSE=\dfrac{SSE}{K(n-1)}$	
جمع	SST	$Kn-1$		

در این حالت دو **F** محاسبه می شود. فرض H_0 را در صورتی رد می کنیم که $F_{Tr} \geq F_{a,k-1,(n-1)(k-1)}$ باشد و فرض H_0' را در صورتی رد می کنیم که $F_B \geq F_{a,n-1,(n-1)(k-1)}$ باشد.

مثال با مراجعه به مثال صفحهٔ 157، در سطح معنی دار 5 درصد آزمون کنید که آیا اختلافهای بین میانگینهای اصل برای مسیرهای مختلف (تیمارها) معنی دارند یا نه و نیز آیا اختلافهای بین میانگینهای حاصل برای روزهای مختلف هفته (بلوکها) معنی دارند یا نه؟

1. فرضها:

$$H_0 = a_1 = a_2 = a_3 = a_4 = 0$$

$$H_0' : b_1 = b_2 = b_3 = b_4 = 0$$

دست کم به ازای یک مقدار **i**، $H_1 : a_i \neq 0$

دست کم به ازای یک مقدار **j**، $H_1' : b_j \neq 0$

۲. آماره آزمون:

جدول 9-7 محاسبات اولیه برای مثال حاضر

مسیرها (تیمارها)	روزهای هفته (بلوکها) شنبه	یکشنبه	دوشنبه	سه شنبه	چهارشنبه	جمع، $T_{i.}$	میانگین
مسیر اول	22	26	25	25	31	129	25/8
مسیر دوم	25	27	28	26	29	135	27
مسیر سوم	26	29	33	30	33	151	30/2
مسیر چهارم	26	28	27	30	30	141	28/2
جمع، $T_{.j}$	99	110	113	111	123	$T_{..}=556$	
میانگین	24/75	28/5	28/25	27/75	30/75		$\bar{x}_{..}=27/8$

بنابراین،

$$\sum_{i=1}^{4}\sum_{j=1}^{5} x_{ij}^{2} = 22^2 + 26^2 + \ldots + 30^2 = 15610$$

$$\begin{aligned} SST &= 15610 - \frac{1}{5\times 4}(556)^2 \\ &= 153/2 \end{aligned}$$

$$\begin{aligned} SS(Tr) &= \frac{1}{5}(129^2 + 135^2 + 151^2 + 141^2) - \frac{1}{5\times 4}(556)^2 \\ &= 52/8 \end{aligned}$$

$$\begin{aligned} SSB &= \frac{1}{4}(99^2 + 110^2 + 113^2 + 111^2 + 123^2) - \frac{1}{5\times 4}(556)^2 \\ &= 73/2 \end{aligned}$$

$$\begin{aligned} SSE &= 153/2 - 52/8 - 73/2 \\ &= 27/2 \end{aligned}$$

بقیهٔ محاسبات را در جدول 9-8 ببینید.

جدول 9-8 جدول «ت.و» برای مثال حاضر

منبع تغییرات	مجموع توانهای دوم	درجه آزادی منبع	میانگین توانهای دوم	F
تیمارها	52/8	3	$\frac{52/8}{3}=17/6$	$\frac{17/6}{2/27}=7/75$
بلوکها	73/2	4	$\frac{73/2}{4}=18/3$	$\frac{18/3}{2/27}=8/06$
خطا	27/2	12	$\frac{27/2}{12}=2/27$	
جمع	153/2	19		

3.مقادیر بحرانی: که داریم $K = 4$ و $n = 5$ و $a = 0/05$؛ پس:

برای تیمارها : $F_{a,k-1,(n-1)(k-1)} = F_{0/05,3,12} = 3/49$

برای بلوکها: $F_{a,n-1,(n-1)(k-1)} = F_{0/05,4,12} = 3/26$

4.تصمیم گیریها: چون $F_{Tr} = 7/75$ از $F_{0/05,3,12} = 3/49$ بیشتر است و $F_B = 8/06$ از $F_{0/05,4,12} = 3/26$ پس نتیجه میگیریم که هر دو فرض صفر باید رد شوند. به عبارت دیگر، اختلافهای بین میانگینهای حاصل برای چهار مسیر مختلف و روزهای مختلف معنی دار است و ناشی از تصادف نیست.

تحلیل واریانس دو عامله با تاثیر متقابل (چند مشاهده در هر سلول)

در قست قبل فرض بر این بود که اثرهای تیماری و بلوکی مستقلند و اگر نتوان آن را صحیح آنست باید از تحلیل واریانس دو عامله با تأثیر متقابل استفاده کرد. مثلاً آیا تأثیر متقابلی بین روشهای تدریس و موضوع درس وجود دارد یا خیر (به عبارت دیگر، آیا می توان گفت برای درس مشخصی، روش تدریس خاصی موفقتر است)، یا می توان گفت که طی مسیر خانه تا محل کار مثلاً در روز شنبه استفاده از مسیر سوم بهتر است (به عبارت دیگر آیا تأثیر متقابلی بین روز و مسیر وجود دارد یا نه)؟

در تحلیل واریانس دو عامله با تأثیر متقابل، از جامعه ای که توزیع نرمال با میانگینهای m_{ij} و واریانس مشترک s^2 دارد برای هر تیمار **i** و هر بلوک **j**– که آن را یک سلول می نامیم- **m** مشاهده تصادفی داریم؛ از اندیس **r** استفاده کرده و مشاهدهٔ x_{ijr} را مشاهدهٔ **r** ام از تیمار **i** (سطر **i**) و بلوک **j** (ستون **j**) می نامیم. این ماتریس را که در هر سلول آن **m** مشاهده وجود دارد در نظر بگیرید:

$$
\begin{array}{ccc}
\hline
x_{111} & & x_{1n1} \\
x_{112} & \cdots & x_{1n2} \\
\vdots & \cdots & \vdots \\
x_{11m} & & x_{1nm} \\
\hline
\vdots & \cdots & \vdots \\
\hline
x_{k11} & & x_{kn1} \\
x_{k12} & & x_{kn2} \\
\vdots & \cdots & \vdots \\
x_{k1m} & & x_{knm} \\
\hline
\end{array}
$$

هر مشاهده ای را می توان به این صورت نوشت:

$$x_{ijr} = m + a_i + b_i + (ab)_{ij} + e_{ijr}$$

در مشاهده فوق m میانگین کل، α_i اثرهای تیماری که در نتیجه $\sum_{i=1}^{k}\alpha_i = 0$ است، b_i اثرهای بلوکی که $\sum_{j=1}^{k}\beta_j = 0$ است، $(ab)_{ij}$ اثرهای متقابل تیمار و بلوک که $\sum_{i=1}^{k}(\alpha\beta)_{ij} = 0$ و $\sum_{j=1}^{n}(\alpha\beta)_{ij} = 0$ است و e_{ijr} نشان دهندهٔ مقادیر متغیرهای تصادفی نرمال با میانگین صفر و واریانس مشترک s^2 است.

سه فرض صفری که آزمون خواهیم کرد عبارتند از صفر بودن اثرای تیماری، اثرهای بلوکی تأثیرات تیمار و بلوک، یعنی:

$H_0 : a_i = 0$ به ازای تمام مقادیر **i**،

$H_0' : b_i = 0$ به ازای تمام مقادیر **j**،

$H_0'' : (\alpha\beta)_{ij} = 0$ به ازای تمام ترکیبات **i** و **j**،

فرض صفر مقابل H_0 و H_0' و H_0'' به ترتیب عبارتند از اینکه همه اثرهای تیماری، همه اثرهای بلوکی و همه تأثیرات متقابل تیماری بلوکی صفر نیستند، یعنی :

دست کم به ازای یک مقدار **i**، $H_1 : \alpha_i \neq 0$

دست کم به ازای یک مقدار **j**، $H_1' : \beta_j \neq 0$

دست کم به ازای یک ترکیب **i**، **j** $H_1'' : (\alpha\beta)_{ij} \neq 0$

در تفکیک انحرافات به چهار جزء (با تأثیر متقابل) می توان ثابت کرد که :

مجموع توان دوم انحرافات حاصل از بلوک +مجموع توان دوم انحرافات حاصل از تیمار=مجموع توان دوم انحرافات

مجموع توان دوم انحرافات باقیمانده +مجموع توان دوم انحرافات حاصل از تأثیر متقابل بلوک و تیمار +

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}(x_{ijr} - \bar{x}_{..})^2 = nm.\sum_{i=1}^{k}(\bar{x}_{i..} - \bar{x}_{...})^2 + km.\sum_{j=1}^{n}(\bar{x}_{.j.} - \bar{x}_{i...})^2$$
$$+ m.\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}(x_{ijk} - \bar{x}_{ij.})^2$$

$$SST = SS(Tr) + SSB + SS(Tr.B) + SSE$$

نحوهٔ محاسبات مجموع توانهای دوم به این صورت است:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m} x^2_{ijr} - \frac{1}{knm}(T^2...)$$

$$SS(Tr) = \frac{1}{mn}.\sum_{i=1}^{k} T_{i..}^{2} - \frac{1}{knm}(T^{2}...)$$

$$SSB = \frac{1}{km}.\sum_{j=1}^{n} T_{.j.}^{2} - \frac{1}{knm}(T^{2}...)$$

$$SS(Tr.B) = \frac{1}{m}.\sum_{i=1}^{k}\sum_{j=1}^{n} T_{ij.}^{2} - \frac{1}{nm}.\sum_{i=1}^{k} T_{i..}^{2} - \frac{1}{km}.\sum_{j=1}^{n} T_{.j.}^{2} + \frac{1}{knm}(T^{2}...)$$

$$SSE = SST - SS(Tr) - SSB - SS(Tr.B)$$

در این صورت جدول تحلیل واریانس به صورت جدول 9-9 خواهد بود.

جدول 92-9 جدول «ت.و» برای تحلیل واریانس دو عامله(با تأثیر متقابل)

منبع تغییرات	مجموع توان دوم	درجه آزادی	میانگین توانهای دوم	F
تیمارها	$SS(Tr)$	$k-1$	$MS(Tr) = \frac{SS(Tr)}{k-1}$	$F_{Tr} = \frac{MS(Tr)}{MSE}$
بلوک	SSB	$(n-1)$	$MSB = \frac{SSB}{n-1}$	$F_{B} = \frac{MSB}{MSE}$
تأثیر متقابل	$SS(Tr.B)$	$(k-1)(n-1)$	$MS(Tr.B) = \frac{SS(Tr.B)}{(k-1)(n-1)}$	$F_{Tr.B} = \frac{MS(Tr.B)}{MSE}$
خطا	SSE	$K(m-1)$	$MSE = \frac{SSE}{kn(m-1)}$	
جمع	SST	$knm-1$		

در این حالت، سه **F** محاسبه می شود. فرض H_0 را در صورتی رد می کنیم که $F_{Tr} \geq F_{a,k-1,kn(m-1)}$ باشد، فرض H_0' را در صورتی رد می کنیم که $F_{B} \geq F_{a,n-1,kn(m-1)}$ باشد و فرض H_0'' را در صورتی رد می کنیم که $F_{Tr.B} \geq F_{a,(k-1)(n-1),kn(m-1)}$ باشد.

مثال می خواهیم تأثیر سه روش مختلف تدریس و چهار درس را بر نمره های دانشجویان بررسی کنیم. بدین منظور نمونه ای متشکل از 36 نفر انتخاب کرده ایم که در آن هر سه نفر در یک ترکیب روش تدریس- درس جای داشتند. اطلاعات را رد جدول صفحه بعد خلاصه کردیم:

آیا در سطح معنی دار 5درصد می توان ادعا کرد که: الف) اثربخشی روشهای مختلف تدریس متفاوت است، ب) میانگین نمره های دروس مختلف با هم متفاوت است و ج) روشهای مختلف تدریس با نوع درس اثر متقابل دارند؟

1. فرضها: $H_0 = a_1 = a_2 = a_3 = 0$

$H'_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

به ازای تمام ترکیبات مختلف **i** و **j**، $H'_0 : (ab)_{ij} = 0$

روش تدریس	درس 1	2	3	4	جمع	میانگین
الف	70	77	82	85		
	79	81	78	90	960	80
	72	79	80	87		
ب	83	77	94	84		
	89	87	83	90	1020	85
	78	88	79	88		
ج	81	74	72	68		
	86	69	79	71	900	75
	79	77	75	69		
جمع	717	709	722	732	2880	
میانگین	79/7	78/8	80/2	81/3		80

و : دست کم به ازای یک مقدار **i**، $H_1 : a_i \neq 0$

دست کم به ازای یک مقدار ، $H'_1 : b_j \neq 0$

دست کم به ازای یکی از ترکیبات **i** و **j**، $H''_1 : (ab)_{ij} \neq 0$

2. آمارهٔ آزمون:

$$\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{r=1}^{3} x_{ijr}{}^2 = 70^2 + 79^2 + \ldots + 69^2 = 232000$$

$$SST = 232000 - \frac{1}{36}(2880)^2 = 1600$$

$$SS(Tr) = \frac{1}{12}(960^2 + 1020^2 + 900^2) - \frac{1}{36}(2880)^2 = 600$$

$$SSB = \frac{1}{9}(717^2 + 709^2 + 722^2 + 732^2) - \frac{1}{36}(2880)^2 = 30/8$$

$$SS(Tr.B) = \frac{1}{9}\left[(70+79+72+)^2 + \ldots + (68+71+69)^2\right] - \frac{1}{12}(960^2 + 1020^2 + 900^2)$$
$$-\frac{1}{9}(717^2 + 709^2 + 722^2 + 732^2) - \frac{1}{36}(2880)^2 = 533/9$$

$$SSE = 1600 - 600 - 30/8 = 533/9 = 435/3$$

جدول 9-10 جدول «ت.و» برای مثال حاضر

منبع تغییرات	مجموع توانهای دوم	درجه آزادی	میانگین توانهای دوم	F
تیمارها	$52/8$	2	$\frac{600}{2}=300/0$	$\frac{300/0}{18/1}=16/57$
بلوک	$30/8$	3	$\frac{30/8}{3}=10/3$	$\frac{10/3}{18/1}=0/57$
تأثیر متقابل	$533/9$	6	$\frac{533/9}{6}=89/0$	$\frac{89/0}{18/1}=4/92$
خطا	$435/3$	24	$\frac{435/3}{24}=18/1$	
جمع	$1600/0$	35		

مقادیر بحرانی: داریم $K=3$ و $n=4$ و $a=0/05$ پس:

برای تیمارها : $F_{a,k-1,kn(m-1)}=F_{0/05,2,24}=3/40$

برای بلوکها: $F_{a,n-1,kn(m-1)}=F_{0/05,3,24}=3/01$

برای تأثیرات متقابل تیمار با بلوک : $F_{a,(k-1)(n-1),kn(m-1)}=F_{0/05,6,24}=2/51$

تصمیم گیریها: چون $F_{Tr}=16/57$ از $F_{0/05,2,24}=3/40$ بیشتر است می توان ادعا کرد کـه بـین میـانگین نمـره هـای حاصل از روشهای مختلف تدریس تفاوت معنی داری وجود دارد، چون $F_B=0/57$ از $F_{0/05,3,24}=3/01$ بیشتر نیست و فرض H_0' پذیرفته مـی شـود (بـین میـانگین نمـره هـای دروس مختلـف تفـاوت معنـی داری وجـود نـدارد)، چـون $F_{Tr.B}=4/92$ از $F_{0/05,6,24}=2/51$ بزرگ تر است می توان ادعا کرد که روشهای تدریس با نوع درس اثر متقابل دارند. اگر به مثال قبل مراجعه کنید می بینید که روش ((ب)) و ((الف)) برای درس 4 به حصول نمره های بالایی منجر شـده، ولی مثلاً برای درس 1، روش ((ج)) مؤثر بوده است. دقت نمایید که روش تحلیل واریانس می گوید که روشهای مختلـف تدریس با نوع درس اثر متقابل دارند، ولی اگر بخواهیم قدمی فراتر از این بگذاریم و بگوییم کدام روش متناسب بـا کـدام درس (دروس) است، باید یکی از به اصطلاح ((آزمونهای مقایسهٔ چندگانه)) را به کار ببـریم کـه از حوصلهٔ ایـن فصـل خارج است.

ملاحظات بیشتر

مطالبی که در این فصل مطرح شد برای مدلهای با آثار ثابت کاربرد دارد. یک ((مدل بـا آثـار ثابـت))، مـدلی اسـت کـه سطوح مختلف تیمار برای عاملی مشخص در آزمایش وارد می شود. مثلاً در مثال آخر به صورت ضـمنی فـرض شـد کـه

تنها روشهای تدریس ((الف))، ((ب)) و ((ج)) وجود دارد و وارد مدل شده اند. نقطـه مقابـل ایـن مـدل، ((مـدل بـا آثـار تصادفی)) است که تنها چند نمونه تصادفی از سطوح مختلف تیمار برای عاملی مشخص وارد می شوند. مثلاً ممکن است 10 روش تدریس وجود داشته باشد ولی 3 تای آنها به صورت تصادفی انتخاب شوند. ر چنین وضعی روشهای محاسـباتی دیگری لازم است، چرا که فرضیه صفر باید مدعی آن باشد که تفاوت معنی داری بین تمام روشهای تدریس، نـه آنهـایی که به صورت تصادفی انتخاب شده اند، وجود ندارد. در بیشتر آزمایشها، مدلهای با آثار ثابت مناسب است و بـدین جهـت مطالب این فصل به این مدلها محدود می شود.

مفاهیم این فصل را برای بیش از دو تیمار یا عامل نیز می توان بسط داد. طرحهایی کـه درگیـر چنـین امـری اسـت بـه ((طرحهای عاملی)) معروفند، و بسیاری از آماردانها، تحلیل واریانس دو عاملی با چند مشـاهده در هـر سـلول را در ایـن طبقه جای می دهند. هر چند می توان تعداد فرضیه های صفر بسیاری را با یک سـری از داده هـا بـه وسـیلهٔ طرحهـای عامی آزمود، ولی توسعه چنین طرحهایی می تواند به سلولهای زیادی در جدول داده ها منجر شده، کـار بـا آن را دشـوار کند. به دلیل چنین مشکلاتی، طرحهایی تهیه شده اند که در آنها دیگر لازم نیست هر ترکیبی از سطوح تیمارهـای یـک عامل را در تحلیل وارد کرد. طرحهایی همانند ((طرح مربع لاتین)) و ((طرحهای بلوکی غیر کامل)) از آن جمله اند.

توجه نمایید که صرف نظر از هر طرحی که از آن استفاده می شود، رد فرض صفر در تحلیل واریانس اغلب برای تحلیلگـر نمی تواند مبنای قدرتمندی برای تصمیم گیری نهایی باشد، چون چنین ردی نمی تواند به طور دقیق مبین تفـاوت بـین سطوح مختلف تیماری باشد. مثلاً فرض کنید که تفـاوت معنـی داری بـین موفقیـت دانشـجویان در سـه روش تـدریس مختلف وجود دارد، ما در قدم بعد می خواهیم ببینیم کدام یک از زوج روشهای تدریس با دیگران تفـاوت دارد. ((آزمـون توکی)) روشی برای مقایسه های دو به دو است که پس از رد فرض صفر به کار می رود.

مقایسه های پس از تجربه

از مقایسه های تبعی یا پس از تجربه زمانی استفاده می شود که در تحلیل واریانس، فرض صفر (H_0) رد شود و فرضـیه مخالف (یعنی H_1) پذیرفته شود؛ یعنی دست کم میانگین دو گروه (جامعه) با هم اختلاف داشته باشند. در شروع کـار، پژوهشگر ممکن است هیچ پرسشی در مورد اینکه اختلاف میانگین کدام گروهها معنی دار است نداشته باشد، ولی بعد از رد H_0 علاقه مند به این موضوع شود. پژوهشگران تازه کار بعد از رد H_0 در تحلیل واریانس ممکن اسـت کـار خـود را خاتمه یافته تلقی کنند، ولی پژوهشگران کهنه کار رد H_0 را شروع دیگری می دانند. در اینجا سه روشِ **LSD**، **HSD** و شفه مطرح می شود.

روش LSD

روش **LSD** یا کمترین تفاوت معنی دار را فیشر ارائه کرده و مستلزم محاسبه کوچک ترین تفاوت معنی دار ممکن بین دو میانگین است. این روش دقیقاً متنی بر همان شیوه ای است که از طریق آزمون **t** استیودنت و با به کار بردن واریانس درون گروهها (واریانس ناشی از خطا) در مقایسه های 2 تایی طرح ریزی شده است.

دو شرط برای استفاده از این روش وجود دارد:

1. تعداد محدودی از مقایسه های مستقل از قبل طراحی شده باشد.
2. همه میانگینهای گروههای آزمایشی با یک گروه کنترل مقایسه شود.

در این روش برای هر مقایسه، یک صفر را به گونهٔ زیر برای میانگینها آزمون می کنیم.

$$\begin{cases} H_0 = C = 0 \\ H_1 = C \neq 0 \end{cases}$$

که **C** عبارت است از تفاوت بین میانگینها. در این روش باید برای هر مقایسه 2 تایی مقدار $\hat{C}$ را بـه دسـت آوریـم کـه $\hat{C} = \bar{x}_i - \bar{x}_j$... است.

برای هر مقایسه 2 تایی لازم است خطای استاندارد تفاوت بین میانگینها، S_d، را حساب کنیم. اگر دو گروه از هم مستقل باشند و حجم نمونه دو گروه یکسان باشد، در این صورت:

$$\begin{aligned} V &= (\bar{X}_i - \bar{X}_j) = V(\bar{X}_i) + V(\bar{X}_j) - 2\mathrm{cov}(\bar{X}_i, \bar{X}_j) \\ &= V(\bar{X}_i) + V(\bar{X}_j) - 0 \\ &= \frac{MSE}{n} + \frac{MSE}{n} = \frac{2MSE}{n} \end{aligned}$$

بنابراین:

$$S_d = \sqrt{\frac{2MSE}{n}}$$

مقدار بحرانی یا کمترین تفاوت معنی دار **(LSD)** عبارت است از:

$$LSD_a = t_{(a, df)} \times S_d$$

اگر حجم نمونه ها برابر نباشند، S_d به صورت زیر محاسبه می شود:

$$S_d = \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

بنابراین برای مواقعی که حجم نمونه ها برابر نیست لازم چند مقدار بحرانی **(LSD)** برای مقایسه ها محاسبه شود.

قاعده تصمیم نیز به این گونه است که اگر $\left|\bar{x}_i - \bar{x}_j\right| \geq LSD$ باشد میانگین دو جامعهٔ m_i و m_j با هم تفاوت دارند.

مثال– فرض کنید مهندسی در بخش تولید کارخانه نساجی می خواهد قدرت کشش نـوعی پارچـه را بـه ازای ترکیبـات مختلف پنبه به کار رفته در آنها بررسی کند. برای این منظور، از هر کدام از پارچه ها در سطوح مختلف پنبه به کار رفته در آنها 5 بار به صورت تصادفی در حجمهای مختلف نمونه گیری کرده و مشاهداتی به صـورت جـدول 5-11 بـه دسـت آورده است.

جدول 9-11 قدرت کشش پنبه به ازای درصد وزن پنبه در 5 بار نمونه گیری

	تیمارها	نمونه						
	(درصد پنبه)	۱	۲	۳	۴	۵	T_i	$\overline{X}_i$
A	15	7	7	15	11	9	49	9/8
B	20	12	17	12	18	18	77	15/4
C	25	14	18	18	19	19	88	17/6
D	30	19	25	22	19	23	108	21/6
E	35	7	20	11	15	11	54	10/8

مطلوب است:

الف) آیا قدرت کشش پارچه ها با درصدهای متفاوت پنبه یکسان است؟ $(a = 0/05)$

ب) در صورت منفی بودن جواب بند الف، کدام دو با هم تفاوت معنی دارای دارند؟ از روش **LSD** استفاده کنید.

1. فرضها: $H_0 = \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

دست کم دو میانگین تفاوت دارند: H_1

2. آمارهٔ آزمون:

$$SST = \sum x^2_{ij} - \frac{1}{kn} \times T^2 ... = 636/96$$

$$SS(Tr) = \frac{1}{n} . \sum T_{i.}^2 - \frac{1}{kn} \times T^2 .. = 475/76$$

$$SSE = SST - SS(Tr) = 161/20$$

بنابراین جدول تحلیل واریانس به صورت جدول 9-12 خواهد بود.

جدول 9-12 جدول تحلیل واریانس مثال حاضر

منبع تغییرات	SS	df	MS	F
تیمارها	475/76	4	118/94	14/76
خطا	161/20	20	8/06	

مقدار بحرانی: داریم $n=5$ و $k=5$ و $\alpha=0/05$ پس:

$$F_{0/05,4,5(5-1)}=F_{0/05,4,20}=2/87$$

4. تصمیم گیری: چون 2/87<14/76 است، فرض صفر(H_0) رد می شود، یعنی اختلاف بین میانگینها ناشی از تصادف (خطا) نیست.

حال که H_0 رد شد، با استفاده از **LSD** می خواهیم منبع اختلاف منبع اختلاف را بررسی کنیم. برای این منظـور، هـر کدام از میانگینهای **A**، **B**، **C**، **D** و **E** را به عنوان گروه کنترل قرار داده، با سایر میانگینها دو به دو مقایسه می کنیم.

$$\hat{C}_1=\overline{X}_A-\overline{X}_B=9/8-15/4=5/6^{*} \qquad \hat{C}_2=\overline{X}_A-\overline{X}_C=9/8-17/6=-7/8^{*}$$

$$\hat{C}_3=\overline{X}_A-\overline{X}_D=9/8-21/6=-11/8^{*} \quad \hat{C}_4=\overline{X}_A-\overline{X}_E=9/8-10/8=-1$$

$$\hat{C}_5=\overline{X}_B-\overline{X}_C=15/4-17/6=-2/2 \quad \hat{C}_6=\overline{X}_B-\overline{X}_D=15/4-21/6=-6/2^{*}$$

$$\hat{C}_7=\overline{X}_B-\overline{X}_E=15/4-10/8=4/7^{*} \qquad \hat{C}_8=\overline{X}_C-\overline{X}_D=17/6-21/6=-4^{*}$$

$$\hat{C}_9=\overline{X}_C-\overline{X}_E=17/6-10/8=6/8^{*} \quad \hat{C}_{10}=\overline{X}_E-\overline{X}_F=21/6-10/8=-10/8^{*}$$

$$S_d=\sqrt{\frac{2\times 806}{5}}=1/795 \qquad , \qquad t_{(0/05,20)}=2/086$$

$$\Rightarrow LSD=2/086\times 1/795=3/75$$

در هر کدام از مقایسه ای دهگانه مذکور که با علامت * نشانه گذاری شده اند، فرض $H_0:m_i=m_j$ رد می شود، زیـرا قدر مطلق تفاوت از مقدار بحرانی **(LSD)** بزرگ تر یا مساوی با آن است.

روش HSD

روش **HSD** یا تفاوت معنی دار راستین که توسط توکی ارائه شد، همانند روش **LSD** برای آزمون این فرضیه که همـه مقایسه های 2 تایی ممکن بین میانگینها در سطح a با یکدیگر برابر هستند، به کار می رود. در این روش نیز اگـر قـدر

مطلق تفاوت بین میانگینهای نمونه، بزرگ تر از مقدار **HSD** یا مساوی با آن باشد، این تفاوت در سطح معین a معنی دار خواهد بود. مقدار بحرانی در روش **HSD** نسبت به روش**LSD** باید بزرگ تر باشد تا نتیجه معنی دار شود. بنابراین، این روش نست به روش **LSD** محافظه کارتر است.

اگر حجم گروههای نمونه برابر باشد مقدار بحرانی از طریق این فرمول محاسبه می شود.

$$\mathrm{HSD} = \mathrm{q}_{(a\,,\mathrm{k.df})} \times \sqrt{\frac{MSD}{n}}$$

مزیت اصلی روش **HSD** نسبت به **LSD** این است که اگر حجم گروههای نمونه برابر نباشد نیازی نیست که چندین مقدار بحرانی محاسبه شود، زیرا این روش برخلاف **LSD** نوعی آزمون با دامنه ثابت است و بنا به پیشنهاد کرک مقدار بحرانی از این فرمول به دست می آید:

$$\mathrm{lSD} = \mathrm{q}_{(a\,,\mathrm{k.df})} \times \sqrt{\frac{MSD}{k} + \frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_k}}$$

توان دوم خطاها با درجه آزادی $df = k(n-1)$ است. **q** نیز معرف نقطه بالایی دامنهٔ توزیعی مشابه با **t** استیودنت بوده که در سطح معنی دار a برای کل میانگینها**(k)** است که با استفاده از جداول 9 و 10 با توجه به مقادیر a، **k** و **df** به دست می آید.

مثال – با مراجعه به مثال قبل و با استفاده از آزمون **HSD** توکی، مقدار بحرانی را محاسبه و مشخص کنید اختلاف میانگین کدام گروهها معنی دار است.

$$a = 0/05\ ,\quad k = 5\ ,\quad n = 5,\quad df = 5(5-1) = 20\ \ , MSE = 8/06$$

$$\mathrm{HSD} = \mathrm{q}_{(0/05,5,20)} \times \sqrt{\frac{8/06}{5}} = 4/232 \times \sqrt{\frac{8/06}{5}} = 5/373$$

جدول 9-13 مقایسه ها را نشان می دهد؛ مقایسه هایی که با علامت * مشخص شده اند معنی دار هستند.

جدول 9-13 نتایج آزمون توکی برای مثال 5-12

		A	B	C	D	E
	میانگن نمونه	9/8	15/4	17/6	21/6	10/8
A	9/8	-	5/6*	6/8*	11/8*	1
B	15/4	-	-	1/2	7/2*	4/6*
C	17/6	-	-	-	5	5/8*
D	21/6	-	-	-	-	10/8*
E	10/8	-	-	-	-	-

روش شِفه

در آزمون شِفه برای هر مقابله، $\hat{C} = \overline{X}_i - \overline{X}_j$ یک فاصله اطمینان محاسبه می شود. اگـر ایـن فاصـله عـدد صـفر را در برگیرد، در این صورت عدد صفر یکی از مقادیر معقول و منطقی برای مقایسه مذکور خواهد بود (یعنی تفاوت بـین m_i و m_j می تواند صفر باشد). اما اگر فاصله اطمینان عدد صفر را شامل نشود، نشانه آن است که صفر نمی تواند یـک مقـدار منطقی برای تفاوت بین دو میانگین باشد و $m_i \neq m_j$ است. این مطلب شبیه به بحث مربوط بـه فاصـله هـای اطمینـان برای تفاوت بین میانگینهای دو جامعه در رابطه با آزمون **t** برای دو گروه مستقل است. اما روش شِفه روش مطلوبتری بـه شمار می آید، زیرا در آن نه فقط دو جامعه بلکه چندین مقایسه پیچیده را می توان انجام داد.

در این روش نیز خطای استاندارد مقایسه c_j از فرمول زیر به دست می آید:

$$Sd_{c_j} = \sqrt{MSE \times \sum_{j=1}^{k} \frac{c_j^2}{n_j}}$$

که در آن c_j^2 مربع هر کدام از وزنهای داده شده به میانگین مورد مقایسه، و n_j نیـز حجـم نمونـه از هـر گـروه (جامعه) است.

مقدار بحرانی این آزمون از این فرمول به دست می آید:

$$S_{a,c_j} = Sd_{c_j} \times \sqrt{(k-1) \times F_{(a,k-1,k(n-1))}}$$

مقدار **F** بر اساس سطح خطای a در درجه آزادی $k-1$ برای صورت، و $k(n-1)$ برای مخـرج از جـدول **F** اسـتخراج می شود. بنابراین فاصله اطمینان برای هر مقایسه با این فرمول محاسبه می شود:

$$\hat{C} - S_{a,c_j} \leq C \leq \hat{C} + S_{a,c_j}$$

که در آن $\hat{C}$ نیز همان تفاوتهای مشاهده شده بین میانگینهای نمونه در مقایسه مورد نظر است که از فرمول زیر پیروی می‌کند

$$\hat{C} = \sum_{j=1}^{k} c_j \overline{X}_j \qquad \text{و} \qquad \sum_{j=1}^{k} c_j = 0$$

که در آن c_j وزن داده شده به هر یک از میانگینهاست.

مثال با مراجعه به مثال قبل و با استفاده از آزمون شِفه به سؤالهای ذیل پاسخ دهید:

آیا میانگین **D** با **E** تفاوت دارد؟

آیا میانگین **E** با متوسط میانگینهای چهارتای دیگر تفاوت دارد؟

مقایسه ها	A	B	C	D	E
1	9/8		17/6	21/6	10/8
2	0	0	0	1	−1
3	$\frac{-1}{4}$	$\frac{-1}{4}$	$\frac{-1}{4}$	$\frac{-1}{4}$	1

$$C_1=\hat{C}_1=\overline{X}_D-\overline{X}_E=21/6-10/8=10/8$$

$$C_2=\hat{C}_2=\overline{X}_E-\frac{\overline{X}_A+\overline{X}_B+\overline{X}_C+\overline{X}_D}{4}$$

$$=10/8-\frac{1}{4}(9/8+15/4+17/6+21/6)=-5/3$$

$\hat{C}_1$ برآورد خطای استاندارد برای $=\sqrt{8/06\times(\frac{1^2}{5}+\frac{(-1)^2}{5})}=\sqrt{3/224}=1/795$

$\hat{C}_2$ برآورد خطای استاندارد برای $=\sqrt{8/06\times(\frac{1^2}{5}+\frac{(-\frac{1}{4})^2}{5}+\frac{(-\frac{1}{4})^2}{5}+\frac{(-\frac{1}{4})^2}{5}+\frac{(-\frac{1}{4})^2}{5})}=1/419$

$\hat{C}_1$ مقدار بحرانی برای $=1/795\times\sqrt{(5-1)\times F_{(0/05,4,20)}}=1/795\sqrt{4\times 2/87}=6/1$

$\hat{C}_1$ مقدار بحرانی برای $=1/419\times\sqrt{(5-1)\times F_{(0/05,4,20)}}=1/419\sqrt{4\times 2/87}=4/8$

فاصله اطمینان برای مقایسه اول $\Rightarrow 10/8-6/1\le C_1\le 10/8+6/1\Rightarrow 4/7\le C_1\le 16/9$

چون مقدار صفر در این فاصله قرار ندارد و در ضمن $\left|\hat{C}_1\right|>6/1$ است، می توان فرض صفر را مبنی بر اینکه تفاوتی بین میانگین **D** و **E** وجود ندارد رد کرد و گفت که تفاوت معنی دار است.

فاصله اطمینان برای مقایسه دوم $\Rightarrow 5/8-4/8\le C_2\le 5/8+4/8\Rightarrow -10/1\le C_2\le -0/5$

چون مقدار صفر در این فاصله قرار می گیرد و $4/8>\left|-5/3\right|$ است، نتیجه می گیریم که تفاوت بین میانگین **E** با متوسط میانگینهای چهار نوع دیگر می تواند صفر باشد و معنی دار نیست.

نمونه سوالات تستی فصل نهم

۱- تحلیل یک طرفه واریانس چه موقع مناسب است؟

1)زمانی که میانگین معلوم است .
2)زمانی که متغیر مستقل مشخص است.
3) زمانی که متغیر مستقل مشخص نیست.
4) زمانی که متغیر وابسته جنسیت است.

۲- نسبت Fدر تحلیل واریانس زمانی که ۴ باشد و احتمال۰۰۶ F/. باشد را تفسیر کنید

1) فرضیه صفر به احتمال 6 در هزار مورد تحلیل واریانس قرار می‌گیرد.
2) از 6 مورد در هزار می‌توان به توزیع داده‌ها اعتماد کرد.
3) اگر فرضیه صفر صحیح می‌بود به احتمال 6 در 1000 **f** را معادل 4 یا بیشتر می‌دیدیم.
4) متغیرهای مورد استفاده دارای ارزش‌های کیفی‌اند که چنین نسبتی از**f** به دست آمده است.

۳- بزرگتر شدن مقدار Fناشی از چیست؟

1) بزرگتر شدن واریانس کل
2) کوچکتر شدن واریانس کل
3) بزرگتر شدن واریانس بین گروه‌ها نسبت به واریانس داخل گروه‌ها
4) کوچکتر شدن واریانس بین گروه‌ها نسبت به واریانس داخل گروه‌ها

۴- محققی می‌خواهد فرضیه:« متوسط سن ابتلا به اضطراب کمتر از متوسط سنی ابتلا به افسردگی می‌باشد» را مورد آزمون قرار دهد، کدام‌یک از آزمون‌های زیر لازم است انجام شود؟

1) مقایسه دو نسبت مستقل
2) مقایسه دو میانگین مستقل
3)مقایسه میانگین و انحراف معیار
4) مقایسه و میانگین همبسته

۵- محققی می‌خواهد نسبت دانشجویان دختر و پسر را در دانشکده علوم اجتماعی با همین نسبت در کل دانشگاه مقایسه کند آزمون آماری مناسب چه می‌باشد؟

1) آزمون**T-Test**
2) ضریب همبستگی**pearson**
3) مجذور خی(کی دو$\mathbf{X}2$)
4) آنالیز واریانسی(**Anova**)

۶- برای سنجش تفاوت بین میانگین‌ها، کدام آزمون مناسب است؟

1)آزمون گاما 2)آزمون لاندا

3)آزمون‌های **F**و**T** 4)آزمون‌های کای اسکور **X**2

۷- اگر میانگین واریانس بین گروه‌ها برابر ۴۰۰ و میانگین واریانس داخل گروه‌ها برابر ۱۰۰ باشد، مقدار آزمون F چقدر خواهد بود؟

3(1 4(2 0/25(3 0/33(4

۸- پژوهشگری می‌خواهد تأثیر دو روش ارائه پوسترهای آموزشی و ارائه برنامه‌های کوتاه نمایش در کاهش تخلفات رانندگی را مقایسه نماید، آزمون آماری مورد استفاده در ارزیابی یافته‌ها چه می‌باشد؟

1)مقایسه میانگین‌ها 2)آزمون **T** زوج(**Paired- test**)

3)تست**T**(**T-Test**) 4)آزمون کا اسکوئر **X**2

۹- برای آزمون فرضیه این که میانگین نمرات درس«مبانی تحلیل محتوا» دانشجویان دو دانشگاه مختلف با هم تفاوت دارند یا خیر، کدام‌یک از آزمون‌های آماری زیر مناسب می‌باشد؟

1)آزمون مقایسه زوج‌ها(**Paired-TTesty**) 2)آزمون**Z**

3) آزمون مجذور کا اسکوئر **X**2 4)**T-Test**

۱۰- اگر در آنالیز واریانس، واریانس بین گروه‌ها برابر ۴۵۰۰ و واریانس داخل گروه‌ها برابر ۱۰۰۰ باشد، مقدار f چقدر است؟

4/5(1 5/5(2 35(3 45(4

۱۱- چنانچه قصد داشته باشیم میانگین مشارکت در فعالیت‌های اجتماعی را در نزد دو گروه شهری و روستایی مورد مقایسه قرار دهیم. کدام‌یک از آزمون‌های زیر مناسب‌تر است؟

1) آزمون مقایسه دو میانگین همبسته 2)آزمون**z** برای یک گروه

3)آزمون تحلیل واریانس 4)آزمون **t**برای دو میانگین مستقل

۱۲- در صورتی که فرضیه تحقیقی در یک پژوهش جهت‌دار باشد، کدام‌یک از آزمون‌های زیر مناسب‌تر است؟

1) دو دامنه 2)یک دامنه 3)بدون دامنه 4)هم دامنه

۱۳- چنانچه قصد داشته باشیم میانگین درآمد یک گروه روزنامه نگار را با میانگین درآمد یک گروه خبرنگار مقایسه کنیم، کدام‌یک از فرضیه‌های زیر مناسب‌ترین است؟

بین میانگین‌ها اختلاف معنی‌داری وجود دارد.

1) $\begin{cases} H_0: \bar{X}_1 - \bar{X}_2 = 0 \\ H_A: \bar{X}_1 - \bar{X}_2 \neq 0 \end{cases}$ 2) $\begin{cases} H_0: u_1 - u_2 = 0 \\ H_A: u_1 - u_2 \neq 0 \end{cases}$

3) $\begin{cases} H_0: \bar{X}_1 - \bar{X}_2 \leq 0 \\ H_A: \bar{X}_1 - \bar{X}_2 \neq 0 \end{cases}$ 4) $\begin{cases} H_0: u_1 - u_2 \leq 0 \\ H_A: u_1 - u_2 \neq 0 \end{cases}$

۱۴- اگر در نمونه‌ای به حجم ۲۰۰، واریانس بین گروهی۱۲ و واریانس دو گروهی ۲۴ باشد، مقدار F چقدر خواهد بود؟ (سراسری۸۸)

1) 0/10 2) 50 3) 10 4) 0/50

۱۵- درآمد سالانه نمونه‌ای احتمالی از خانواده‌های سه شهر در جدول توزیع درآمد ارائه شده‌است؟برای آزمون معنی داری تفاوت متوسط درآمدها در این سه شهر چه آزمونی مناسب است؟

شهر الف	شهر ب	شهر ج
4	1	2
8	3	6

1) **F**

2) **Z**

3) X^2

4) **T**

۱۶- برای بررسی تفاوت تحصیلی(نمرات) ۲۰ دانشجو در دو ترم متوالی کدام آزمون مناسب است؟

1) رگرسیون یک متغیره 2)آزمون **T** با نمونه‌های وابسته

3) آزمون **T** با نمونه‌های مستقل 4) رگرسیون چند متغیره

۱۷- روابط کدام‌یک از متغیرهای زیر در آزمون T تست می‌شود؟

1) اسمی – فاصله‌ای 2) رتبه‌ای- فاصله‌ای 3) فاصله‌ای – نسبی 4) اسمی- اسمی

۱۸- پر کاربردترین آزمون و سطح سنجش در تحلیل محتوا کدام است؟

1) نسبت – نسبی 2) آزمون X^2 - نسبی

3) آزمون X^2 - اسمی 4)نسبت – اسمی

۱۹- در صورتی که یک محقق قصد برآورد واریانس جامعه را از طریق نمونه‌ای که از این جامعه انتخاب کـرده است، داشته باشد، مجموع مجذورات انحراف از میانگین را بر کدام‌یک از مقادیر زیر تقسیم کند.

1) $\mathbf{n}$ 2) $n-2$ 3) $n-1$ 4) $N-1$

۲۰- تحلیل واریانس با انتخاب و جایگزینی تصادفی با کدامیک از مقیاس‌های زیر بکار برده می‌شود؟

1)ترتیبی 2)اسمی 3)لیکرت و نسبی 4)حداقل مقیاس فاصله‌ای

۲۱- در تجزیه و تحلیل برای سه گروه که در هر گروه ده آزمودنی در نظر گرفته شده مجموع مجذورات بـین گروه‌ها ۴۵ و مجموع مجذورات بین درون‌گروه‌ها ۱۸۹ محاسبه شده است. عدد F کدام است؟

1) $2/84$ 2) $2/97$ 3) $3/15$ 4) $3/21$

۲۲- به منظور آزمون معنی‌دار بودن ضریب همبستگی ، از کدام توزیع استفاده می‌شود؟

1) $\mathbf{T}$ 2) $\mathbf{F}$ 3)کای دو 4)پواسون

۲۳- جدول مقابل نمرات دو گروه در آزمون اندازه‌گیـری مهـارت آنهـا اسـت، آمـاره آزمـون T در مقایسـه میانگین‌های این دو گروه کدام است؟

گروه اول	7	6	8	9	5
گروه دوم	4	8	5	7	

1) $\frac{\sqrt{5}}{3}$

2) $\frac{\sqrt{7}}{3}$

3) $\frac{\sqrt{5}}{4}$

4) $\frac{\sqrt{7}}{4}$

۲۴- اگر Z_1 تا Z_K متغیرهای استاندارد صفر و یک باشند، آنگاه توزیع $\sum_{i=1}^{K} z_i^2$ کدام است؟

$\mathbf{F}$(1   2) کای دو   2) نرمال   4)$\mathbf{t}$استیودنت

۲۵- در صورتی که قصد مقایسه میانگین یک گروه ۶۰ نفره در یک طرح پیش آزمون و پس آزمون را داشته باشیم، درجات آزادی مساوی است با

58 (1   59 (2   120(3   118(4

۲۶- درجات آزادی میانگین مجموع مجذورات بین گروه‌های مساوی کدامیک از عبارت‌های زیر است؟

$N-J$ (1 $N-1$(2 $J-1$(3 $N-1$(4

۲۷- رعایت مفروضه همسانی واریانس در کدامیک از آزمون‌های زیر با اصلاح درجات آزادی صورت می‌گیرد؟

1) میانگین‌های همبسته   2) ضریب همبستگی   3) میانگین مستقل   4) تحلیل واریانس یک طرفه

۲۸- آزمون‌های آماری با توجه به کدامیک از فعالیت‌های زیر تعیین می‌شود؟

1) هدف تحقیق   2) فرضیه‌های آماری

3) فرضیه‌های قیاسی   4) مسأله پژوهشی

پاسخنامه سوالات تستی

۱- گزینه۲ صحیح است.

تحلیل واریانس به دو دسته تقسیم می‌شود:

1. تحلیل واریانس یک طرفه(**one- way**)

2. تحلیل واریانس دو طرفه(**two-way**)

تحلیل واریانس یک طرفه روشی است به منظور آزمون معنا دار بودن تفاوت بین میانگین‌ها، هنگامی کـه فقـط یک متغیر مستقل دستکاری می‌شود، و طی آن اثر یک متغیر مستقل روی متغیـر وابسـته مـورد بررسـی قـرار می‌گیرد. به عبارت دیگر در این نوع تحلیل واریانس یک متغیر مستقل و یک متغیر وابسته وجود دارد.

۲- گزینه ۳ صحیح است.

۳- گزینه ۲ صحیح است.

$$F = \frac{(Msb)}{(Msw)}$$

میانگین واریانس بین گروه‌ها $\rightarrow$ Msb

میانگین واریانس درون گروه‌ها $\rightarrow$ Msw

با توجه به فورمول مذکور، بزرگتر شدن مقدار F ناشی از بزرگتر شدن واریانس بین گروه‌ها، نسبت به واریـانس درون گروه‌هاست.

۴- گزینه ۴ صحیح است.

نمونه‌های همبسته در طرح‌های پژوهش زیر به کار برده می‌شوند:

1. اندازه‌گیری مکرر: در این نوع طرح برای هر یک از آزمودنی‌ها در نمونه مورد مطالعه دو بار اندازه‌گیری مـی-شود.

2. طرح جفت‌های همتراز شده: در این نوع پژوهش آزمودنی‌های هر دو نمونه براساس یک یا چند متغیر که بـا متغیر وابسته یا متغیر ملاک رابطه دارند، همتراز می‌شوند

۵- گزینه ۳ صحیح است.

آزمون خی دو، برای محاسبه رابطه معنی‌دار نسبت‌ها و درصدها در سطح سنجش اسمی بکار می‌رود، در این سوال جنسیت در سطح اسمی است

۶- گزینه ۳ صحیح است.

آزمون **F**(تحلیل واریانس)، یک روش آماری است که به منظور تحلیل تفاوت بین میانگین‌های دو یا چند نمونه آماری به کار می‌رود، در ضمن تحلیل واریانس زمانی که برای مقایسه میانگین‌های دو نمونه بکار برده شود، معادل آزمون **T**است. آزمون لاندا وآزمون گاما از ضرایب همبستگی می‌باشد. آزمون خی دو هم به منظور آزمون فرضیه درباره استقلال فراوانی‌هایی که در طبقه‌های مختلف قرار گرفته‌اند، بکار برده می‌شود.

۷- گزینه ۲ صحیح است.

$$F=\frac{Msb}{Msw}=\frac{400}{100}=4$$

میانگین واریانس بین گروه‌ها ← Msb

میانگین واریانس درون گروه‌ها ← Msw

۸- گزینه۳ صحیح است.

برای بدست آوردن آزمون معنی داری در این سوال از مقایسه میانگین‌ها استفاده می‌شود، مقایسه میانگین‌ها یا برای دو گروه بکار می‌رود یا برای بیش از دو گروه . در این سوال چون قصد داریم دو روش ارائه را با هم مقایسه کنیم، دو گروه هستند بنابراین از آزمون **t** بهره می‌گیریم. ضمناً مورد استفاده آزمون **T** مبتنی بر گروه‌های وابسته(زوج) زمانی است که از یک گروه آزمودنی دو بار آزمون گرفته شود ما در اینجا دو گروه آزمودنی داریم پس از آزمون**T** مبتنی بر گروه‌های مستقل استفاده می‌شود.

۹- گزینه ۴ صحیح است.

آزمون **T(T-Test)** برای مقایسه میانگین بین دو نمونه یا دو گروه مناسب است.

۱۰-گزینه ۱ صحیح است.

$$\Rightarrow\frac{4500}{1000}=4/5\ \ F=\frac{Msb}{Msw}$$

$\rightarrow$ Msb میانگین واریانس بین گروه‌ها

$\rightarrow$ Msw میانگین واریانس درون گروه‌ها

۱۱-گزینه ۴ صحیح است.

میانگین نمرات دو گروه مستقل (شهری – روستایی) توسط آزمون **t** مستقل انجام می‌شود.

۱۲- گزینه ۲ صحیح است.

زمانی که محقق جهت تأثیر متغیر مستقل بر متغیر وابسته را قبل از تحقیق معلوم کند، فرضیه ما جهت دار است و دراین موارد باید ار آزمون یک دامنه استفاده شود.

۱۳- گزینه ۲ صحیح است.

در آمار استنباطی فرضیه‌های صفر و خلاف به صورت پارامتر صورت‌بندی می‌شود، یعنی باید با حروف یونانی نوشته شوند. فرضیه صفر را با H_0 نشان می‌دهند. این فرض، اصل را بر این قرار می‌دهد که بین پارامترهای مورد مطالعه اختلاف یا ارتباط معناداری وجود ندارد. $\{H_0 : u_1 - u_2 = 0\}$فرضیه خلاف را با H_1 یا H_A نشان می‌دهند. این فرضیه مخالف فرضیه صفر است و در اکثر موارد با فرضیه پژوهشی مطابقت دارد. به این معنی که فرضیه خلاف انتظار پژوهشگر را پیرامون نتایج آتی پژوهش بیان می‌کند. $\{H_a : u_1 - u_2 \neq 0\}$

۱۴- گزینه ۴ صحیح است

$$F = \frac{Msb}{Msw} = \frac{12}{24} = 0/5$$

$\rightarrow$ Msb میانگین واریانس بین گروه‌ها

$\rightarrow$ Msw میانگین واریانس درون گروه‌ها

۱۵-گزینه ۱ صحیح است.

آزمون **F** برای بررسی تفاوت میانگین‌های سه گروه و بیشتر مورد استفاده قرار می‌گیرد. (دلاور، 1387: 350)

۱۶- گزینه ۲ صحیح است.

چون در این سوال دو متغیر مستقل بر روی یک گروه ثابت اندازه‌گیری می‌شود، پس باید از **T** همبسته استفاده کرد.

۱۷- گزینه ۱ صحیح است.

آنجا که متغیر وابسته در سطح فاصله‌ای باشد و تعداد طبقات متغیر مستقل کم باشد (اسمی) بهتر است در تحلیل داده‌ها به جای جدول توافقی از مقایسه میانگین‌های پاره‌گروه‌ها(آزمون **T**) استفاده کنیم.

۱۸- گزینه ۳ صحیح است.

پرکاربرد ترین آزمون تحلیل محتوا، آزمون خی دو است. که این آزمون بر معنــی‌داری بــین دو متغیــر در ســطح اســمی دلالت دارد.

۱۹- گزینه۴ صحیح است.

درجات آزادی مجموع مجذورات کل مطابق زیر، از جمع کردن درجات آزادی بین گروه‌ها و درون گروه‌ها بدست می‌آید

$$d.ft = dfw + dfb = (N-k) + (k-1) = N-1$$

۲۰- گزینه ۴ صحیح است.

از مفروضه‌های تحلیل واریانس این است که مقیاس اندازه‌گیری حداقل فاصله‌ای باشد.

۲۱-گزینه ۴ صحیح است.

در این سوال ابتدا باید مجموع مجذورات بین گروهی و درون گروهی را به واریانس بین گروهــی و درون گروهــی تبــدیل کنیم و سپس هر کدام را بر درجه آزادی خاص خود تقسیم کنیم.

$$F = \frac{Msb}{Msw} = \frac{\frac{ssb}{k-1}}{\frac{ssw}{N-1}}$$

$$SS_W = 189$$
$$df_w = N-k$$
$$SS_b = 45$$
$$df_b = k-1$$

$$F = \frac{\frac{45}{3-1}}{\frac{189}{30-3}} = \frac{22/5}{7} = 3/21$$
$$F = 3/21$$

قابل ذکر است که df_w از تفاضل تعداد کل آزمودنی‌ها(30 نفر) به تعداد گروه‌هــا(یعنی 3) بــه دســت مــی‌آیــد و df_b از تفاضل تعداد گروه‌ها از یک بدست می‌آید.

۲۲- گزینه ۱ صحیح است.

رد یا اثبات فرضیه (معنی‌داری همبستگی) از فرمول $t = r_s\sqrt{\frac{n-2}{1-r_s^2}}$ انجام می‌گیرد.

۲۳- گزینه صحیح وجود ندارد.

x_1	x_2	$(x_1-\bar{x}_1)$	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)$	$(x_2-\bar{x}_2)^2$
7	4	0	0	-2	4
6	8	-1	1	2	4
8	5	1	1	-1	1
9	7	2	4	1	1
5		-2	4		
35	24		10		10

گروه دوم $= x_2$ گروه اول $= x_1$

$$s^2x_1=\frac{\sum(x_1-\bar{x}_1)^2}{N}=\frac{10}{5}=2$$

$$s^2x_2=\frac{\sum(x_2-\bar{x}_2)^2}{N}=\frac{10}{4}=2/5$$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}}}\Rightarrow\frac{35-24}{\sqrt{\frac{2}{5}+\frac{2/5}{4}}}=\frac{11}{\sqrt{0/4+0/62}}=\frac{11}{\sqrt{1/02}}=\frac{11}{1/009}=10/9$$

$$t=10/9$$

۲۴- گزینه ۳ صحیح است.

۲۵- گزینه ۲ صحیح است.

زیرا در این سوال برای مقایسه میانگین این گروه باید از آزمون **t** همبسته استفاده نمائیم که درجه آزادی در **t**همبسته **n-1** می‌باشد، یعنی $60-1=59$

۲۶- گزینه ۳ صحیح است.

درجه آزادی میانگین مجموع مجذورات $k-1$ یا $j-1$ می‌باشد.

۲۷- گزینه ۴ صحیح است.

به تجربه ثابت شده است که رعایت نکردن مفروضه‌های طبیعی بودن توزیع متغیر و همسانی واریانس‌ها بر اعتبار آزمون **F** تأثیر زیادی ندارد. صحت این مسأله زمانی که اندازه نمونه‌ها مساوی باشد بیشتر است.

۲۸- گزینه ۲ صحیح است.

منابع

1- ساروخانی ، باقر . روش های تحقیق در علوم اجتماعی جلد 1و2. تهران

2- شریفی ، حسن پاشا; و نجفی زند ، جعفر . (1382) . روش های آماری در علوم رفتاری. چاپ دوازدهم. تهران: سخن

3- شریفی ، حسن پاشا; و نجفی زند ، جعفر . (1373) . روش های آماری در روان شناسی، علوم تربیتی، اجتماعی و ......(علوم رفتاری). چاپ چهارم. تهران:دانا

4– دلاور ، علی.(1375) . روش تحقیق در روان شناسی و علوم تربیتی . تهران : نشر ارسباران

5- سرمد ، زهره ; بازرگان عباس ; حجازی الهه . (1376) . روش های تحقیق در علوم رفتاری .تهران : آگاه

6 – فرگوسن ، جرج ا; تاکانه یوشیو. (1989). . تحلیل آماری در روان شناسی و علوم تربیتی . ترجمه دلاورو نقشبندی. چاپ دوم ، 1380 . تهران : نشر ارسباران

7 – هومن ، حیدر علی .(1376). شناخت روش علمی در علوم رفتاری(پایه های پژوهش) . چاپ سوم . تهران : نشر پارسا

8- هومن ، حیدر علی .(1370) . استنباط آماری در پژوهش رفتاری . چاپ اول . تهران : پیک فرهنگ

9- نقشبندی ، سیامک . (1383) . آمارو روش تحقیق . چاپ دوم . تهران . پوران پژوهش