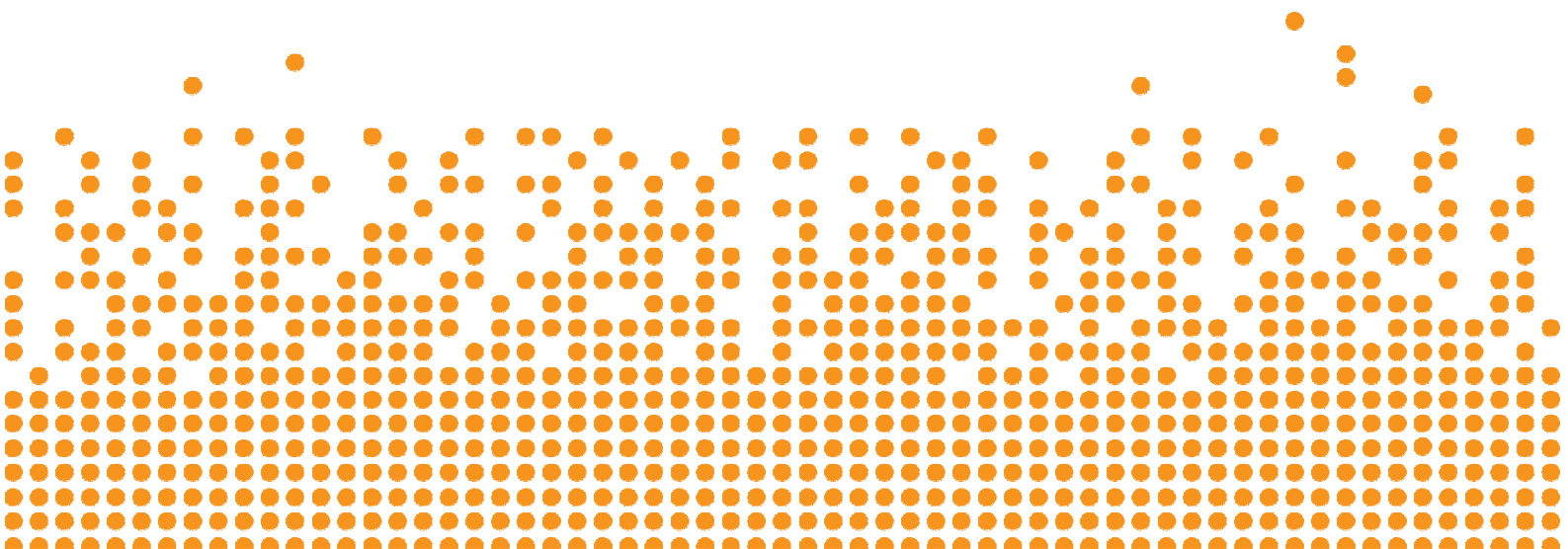


آمار و مدل سازی



آمار:

آمار علم جمع‌آوری اطلاعات، سازمان‌دهی، تفصیل، تجزیه و تحلیل آن‌ها و نتیجه‌گیری اطلاعاتی می‌باشد، به عبارت دیگر آمار جمع‌آوری و طبقه‌بندی همراه با استنباط است. هم‌چنین آمار را هنر تصمیم‌گیری در شرایط نامطمئن نیز گفته‌اند.

فصل اول: اندازه‌گیری و مدل‌سازی

بیان مسأله به زبان ریاضی را مدل‌سازی ریاضی می‌گویند. هر چقدر مفاهیم ریاضی به‌کار برده شده ساده‌تر و ابتدایی‌تر و نتیجه‌ی کار به پدیده مورد نظر نزدیک‌تر باشد، مدل‌سازی با ارزش‌تر است.

اندازه‌گیری عبارت است از تخصیص معیار عددی به یک صفت. اولین قدم برای رسیدن به اطلاعات عددی اندازه‌گیری است. در مدل‌سازی ریاضی با اندازه‌گیری سر و کار داریم و در اندازه‌گیری همواره خطا وجود دارد. خطای اندازه‌گیری تفاضل مقدار واقعی داده و مقدار اندازه‌گیری شده می‌باشد. این خطا لزوماً از واحد اندازه‌گیری کمتر است.

مثال: اگر قد شخصی برابر $180/6$ اندازه‌گیری شده باشد، مدل قد او را معین کنید.

کحل:

$$H = 180/6 + E \quad |E| < 0/1$$

چون توانستیم تا $0/1$ متر را اندازه‌گیری کنیم پس باید خطا از واحد اندازه‌گیری کوچکتر باشد.

مثال: اگر شعاع کره ای به صورت $R = 3 + E$ که در آن E مقدار خطاست مدل‌سازی شده باشد، مدلی برای حجم کره بنویسید.

کحل:

$$V = \frac{4}{3}\pi(R)^3 = \frac{4}{3}\pi(3+E)^3 = \frac{4}{3}\pi(27+3 \times 9E) = 36\pi + 36\pi E = 36\pi + E_1$$

مثال: در یک استوانه مدل شعاع قاعده $R = 4 + E_1$ و مدل ارتفاع $h = 3 + E_2$ است. در این صورت حجم استوانه از چه

مدلی پیروی می‌کند؟

کحل:

$$V = \pi R^2 h = \pi(4 + E_1)^2 (3 + E_2) = 48\pi + 24\pi E_1 + 16\pi E_2$$

فصل دوم: جامعه و نمونه

جامعه‌ی آماری مجموعه‌ای از افراد یا اشیا است که درباره‌ی اعضای آن می‌خواهیم موضوع یا موضوعاتی را مطالعه کنیم. تعداد اعضای جامعه را اندازه‌ی جامعه می‌گوییم. اگر تمام افراد جامعه را مورد مطالعه قرار دهیم می‌گوییم سرشماری کرده‌ایم.

مشکلات سرشماری:

- ۱- در دسترس نبودن تمام اعضای جامعه
 - ۲- وقت‌گیر بودن دسترسی به تمام اعضای جامعه
 - ۳- گران تمام شدن بررسی تمام اعضای جامعه
 - ۴- از بین رفتن جامعه در برخی از مطالعات
- نمونه زیرمجموعه‌ای از جامعه‌ی آماری است.

نمونه تصادفی:

اگر جامعه‌ی آماری را که می‌خواهیم روی موضوعات آن مطالعاتی انجام دهیم کوچک باشد، معمولاً مطالعه را به صورت سرشماری انجام می‌دهیم.

نمونه: گاهی به علت وسعت جامعه نمی‌توانیم سرشماری (بررسی کلیه اجزای جامعه) انجام دهیم. در این مواقع اقدام به بررسی تصادفی بخشی از جامعه‌ی آماری می‌کنیم و نتایج حاصل را به صورت استقرایی به تمام جامعه تعمیم می‌دهیم. نمونه باید به قسمی انتخاب شود که بتواند بیانگر جامعه باشد.

لذا باید روش انتخاب نمونه به گونه‌ای باشد که:

۱- امکان انتخاب هر فرد به عنوان عضوی از نمونه امکان‌پذیر باشد.

۲- قبل از انتخاب نمونه، نتوانیم با اطمینان بیشتر درباره‌ی حضور یا عدم حضور عده‌ای در نمونه قضاوت کنیم.

بدیهی است نمونه‌ی مورد نظر باید به گونه‌ای انتخاب شود که تعمیم نتایج حاصل از نمونه‌گیری غیرواقعی نباشد. اطلاعات حاصل از نمونه‌گیری را **داده** می‌نامیم و تعداد افراد نمونه را **حجم** یا **اندازه‌ی نمونه** می‌نامند.

مثال: در کدام بررسی، اندازه نمونه برابر اندازه جامعه است؟

(۱) نمونه تصادفی (۲) دسته‌بندی (۳) سرشماری (۴) با متغیر کیفی

✓ حل: گزینه ۳ پاسخ است.

نمونه‌گیری تصادفی ساده به روش‌های مختلف انجام می‌گیرد، یکی از این روش‌ها استفاده از اعداد تصادفی به کمک ماشین حساب است (اعداد تصادفی، اعداد بین صفر و یک هستند که به وسیله‌ی کلید RAN تولید می‌شوند). پس از تولید عدد تصادفی به وسیله‌ی ماشین حساب آن را در حجم جامعه‌ی آماری ضرب کرده و اولین عدد بزرگتر از یا مساوی با عدد تولید شده را به عنوان عدد تصادفی در نظر می‌گیریم. سپس عضوی از جامعه را که عدد به دست آمده متناظر با اوست به عنوان نمونه در نظر می‌گیریم.

مثال: در یک جامعه به حجم ۲۰۰ می‌خواهیم به کمک ماشین حساب نمونه‌گیری انجام دهیم. اعداد تصادفی ۰/۲۹۱ و ۰/۶۵۰ به دست آمده‌اند. چه شماره‌هایی متناظر با این اعداد به ترتیب باید انتخاب شوند؟
✓ حل:

$$۰/۲۹۱ \times ۲۰۰ = ۵۸/۲ \rightarrow \text{نمونه} = ۵۹$$

$$۰/۶۵۰ \times ۲۰۰ = ۱۳۰ \rightarrow \text{نمونه} = ۱۳۰$$

روش‌های جمع‌آوری داده‌ها:

داده: نتایج حاصل از اندازه‌گیری و یا بررسی نمونه را داده می‌گوییم.

روش‌های جمع‌آوری داده:

- ۱- استفاده از داده‌های از پیش تهیه شده
- ۲- از طریق پرسش کتبی یا شفاهی
- ۳- از طریق مشاهده و ثبت وقایع
- ۴- از طریق انجام آزمایش

روش طراحی پرسش‌نامه:

- ۱- قبل از پرسش سؤال، محتوای پرسش‌نامه باید سازماندهی شود.
- ۲- هدف بررسی باید روشن باشد.
- ۳- تهیه فهرست از عناوینی که باید راجع به آن اطلاعات جمع‌آوری شود.
- ۴- خودداری از جمع‌آوری اطلاعات اضافی

- ۵- سؤالات کوتاه و واضح باشد. نباید از سؤالات چند برداشت شود. حتی الامکان جواب‌های تک‌کلمه‌ای یا تکرریمی
- ۶- در سؤالاتی که ممکن است پاسخ‌دهنده نخواهد به آن جواب دقیق بدهد از سؤالات با پاسخ از پیش آماده شده استفاده شود؛ مانند تعیین محدوده
- ۷- عدم استفاده از سؤالات هدایت‌کننده یعنی سؤالاتی که جواب را به پاسخ‌دهنده القا می‌کند.
- ۸- حتی الامکان از سؤالات با پاسخ چندگزینه‌ای استفاده شود.
- ۹- برای پرسش عقیده راجع به موضوع یا محصول، پاسخ‌ها به صورت گزینه‌های کمی یا کیفی سطح‌بندی شده انتخاب شوند؛ مانند بسیار خوب، خوب، متوسط، ضعیف و بسیار ضعیف.
- ۱۰- ضمیمه کردن دستورالعمل پاسخگویی (واضح و کامل)
- ۱۱- تشکر در پایان

مثال: جمع‌آوری داده‌ها به کدام طریق مورد قبول نیست؟

- (۱) مصاحبه (۲) مشاهده (۳) انجام آزمایش (۴) پرسش هدایت‌کننده

✓ حل: گزینه ۴ پاسخ است.

مثال: کدام طریق برای جمع‌آوری داده‌ها مناسب نیست؟

- (۱) مصاحبه (۲) الگوی خاص (۳) مشاهده (۴) آزمایش

✓ حل: گزینه ۲ پاسخ است.

فصل سوم: متغیرهای تصادفی

متغیر تصادفی: مشخصه یا صفت ویژه‌ای از افراد جامعه که روی آن مطالعه انجام می‌دهیم.

از دیدگاهی دیگر متغیرها بر دو نوع‌اند:

(۱) کمی: متغیری که اندازه و مقدار دارد و دارای واحد اندازه‌گیری می‌باشد، مانند: قد، وزن، عمر و ...

که خود بر دو نوع است $\left\{ \begin{array}{l} \text{پیوسته: غیرقابل شمارش} \\ \text{مانند: قد، وزن و ...} \end{array} \right.$

(۲) کیفی: متغیری که قابل اندازه‌گیری نیست و عموماً با مقایسه بررسی می‌شود، مانند: هوش، استعداد، درس‌خوان بودن و ...

که خود بر دو نوع است $\left\{ \begin{array}{l} \text{ترتیبی: دارای ترتیب ذاتی هستند مانند حروف الفبای فارسی، مراحل رشد و ...} \\ \text{اسمی: دارای ترتیب ذاتی نیستند مانند رنگ چشم افراد، گروه‌های خونی و ...} \end{array} \right.$

صفت: کمیت یا کیفیتی که متعلق به عناصر جامعه آماری است، صفت نامیده می‌شود که بر دو نوع است:

(۱) صفت ثابت: صفتی که در بین عناصر جامعه‌ی آماری مشترک است.

(۲) صفت متغیر: صفتی که از هر عضو به عضو دیگر تغییر می‌کند.

مثلاً در جامعه‌ی آماری ایرانی‌ها متغیر تصادفی وزن افراد صفت متغیر و ایرانی بودن صفت ثابت است.

مثال: گروه خونی افراد کدام نوع متغیر است؟

- (۱) کیفی - اسمی (۲) کیفی - ترتیبی (۳) کمی - پیوسته (۴) کمی - گسسته

✓ حل: گزینه ۱ پاسخ است.

مثال: مراحل تحصیلی، متغیر تصادفی است. نوع آن کدام است؟

(۱) کمی گسسته (۲) کمی پیوسته (۳) کیفی اسمی (۴) کیفی ترتیبی

بهر حل: گزینه ۴ پاسخ است.

فصل چهارم: دسته‌بندی داده‌ها و جدول فراوانی

فراوانی یک داده:

در داده‌های آماری دسته‌بندی نشده به تعداد دفعاتی که هر داده تکرار می‌شود، فراوانی مطلق آن داده گفته می‌شود و آن را با f_i نمایش می‌دهند هم‌چنین نسبت فراوانی مطلق هر داده به تعداد داده‌ها، فراوانی نسبی نامیده می‌شود.

$$F_i = \frac{f_i}{N} \quad 0 \leq F_i \leq 1$$

$F_i \times 100 =$ درصد فراوانی نسبی

وقتی فراوانی داده‌های آماری که تعداد و پراکندگی‌شان زیاد است مورد مطالعه قرار می‌گیرند استفاده از جدول توزیع فراوانی بسیار دشوار است، لذا داده‌ها را دسته‌بندی می‌کنیم.

دسته‌بندی داده‌ها:

اگر تعداد داده‌ها کم باشد یا داده‌ها خیلی پراکنده و دارای توزیع وسیع نباشند، جدولی متشکل از داده‌ها و تعداد تکرار آن‌ها رسم می‌کنیم که به آن جدول توزیع فراوانی گفته می‌شود.

تعاریف مرتبط:

الف - دامنه‌ی تغییرات:

تفاضل کمترین از بیشترین داده در یک جامعه‌ی آماری، دامنه‌ی تغییرات جامعه‌ی آماری نامیده می‌شود. $R = X_{\max} - X_{\min}$

ب- محدود دسته‌ها:

به اعدادی که دو طرف یک دسته قرار می‌گیرند، حدود آن دسته و یا کران‌های پایین و بالا گفته می‌شود و معمولاً به صورت قراردادی به استثنای دسته‌ی آخر حد بالای هر دسته متعلق به آن دسته نیست و جزو دسته‌ی بعد حساب می‌شود.

ج- تعداد دسته‌ها:

دستور خاصی برای انتخاب تعداد دسته‌ها وجود ندارد، اما معمولاً تعداد دسته‌ها را به گونه‌ای انتخاب می‌کنند که طول دسته‌ها دچار تعارض نگردد. به این معنا که فراوانی هیچ دسته‌ای صفر نباشد.

د- طول دسته:

تفاضل کران پایین هر دسته از کران بالای آن است که در درس ما معمولاً طول تمام دسته‌ها برابر فرض می‌شود. اگر در مورد تعداد دسته‌ها تصمیم‌گیری کردیم، طول دسته‌ها از رابطه‌ی زیر به دست می‌آید:

$$\text{طول دسته} = \frac{\text{دامنه‌ی تغییرات}}{\text{تعداد دسته‌ها}}$$

اگر بر اثر گرد کردن تعداد دسته‌ها، حاصل ضرب طول دسته در تعداد دسته از دامنه‌ی تغییرات بزرگ‌تر شد، این مقدار اضافی را بین دسته‌ی اول و دسته‌ی آخر به صورت مساوی تقسیم می‌کنیم.

اگر در مورد طول دسته‌ها تصمیم‌گیری کرده باشیم، در این صورت تعداد دسته‌ها از رابطی زیر به دست می‌آید.

$$\text{تعداد دسته‌ها} = \frac{\text{دامنه‌ی تغییرات}}{\text{طول دسته‌ها}}$$

که اگر عدد فوق اعشاری بود حتماً از رابطی «+1» $\left[\frac{\text{دامنه‌ی تغییرات}}{\text{طول دسته‌ها}} \right] + 1$ = تعداد دسته‌ها استفاده می‌شود.

مثال: در یک آمارگیری بیشترین داده ۷۱ و دامنه‌ی تغییرات ۳۷ می‌باشد. اگر طول دسته‌ها ۵ انتخاب شده باشد و حد پایین اولین دسته را برابر کوچک‌ترین داده انتخاب کنیم، حد بالای آخرین دسته کدام است؟
 کحل:

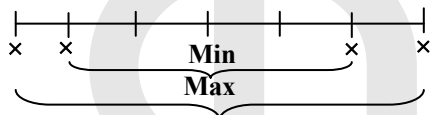
$$R = x_{\max} - x_{\min} = 37 \rightarrow 71 - x_{\min} = 37 \rightarrow x_{\min} = 34$$

پس اولین داده ۳۴ است. حال باید به اندازه‌ی $n = \left[\frac{37}{5} \right] + 1 = 8$ دسته با طول ۵ جلو برویم.

$$\text{حد بالای آخرین دسته} = 34 + 8 \times 5 = 74$$

مثال: در یک آمارگیری پس از دسته‌بندی داده‌ها، ۶ دسته با طول ۴ به وجود آمده است، به فرض آن که هیچ دسته‌ای خالی نباشد، حدود تغییرات دامنه این داده‌ها را به دست آورید.

کحل:



$$4 \times 6 < d \leq 6 \times 4 \rightarrow 16 < d \leq 24$$

ه- فراوانی یک دسته در داده‌های دسته‌بندی شده:

تعداد داده‌هایی که در هر دسته قرار می‌گیرند را فراوانی مطلق آن دسته می‌گویند.

و- فراوانی نسبی و تجمعی یک دسته در داده‌های دسته‌بندی شده:

فراوانی نسبی دسته i ام در داده‌های دسته‌بندی شده برابر است با:
$$F_i = \frac{f_i}{\text{مجموع کل فراوانی‌ها}}$$

اگر فراوانی مربوط به هر دسته را با مجموع فراوانی‌های دسته‌های قبل از آن دسته جمع کنیم، فراوانی تجمعی مربوط به آن دسته به دست می‌آید:
$$f_{ci} = f_1 + f_2 + \dots + f_i$$

و فراوانی تجمعی نسبی دسته i ام در داده‌های دسته‌بندی شده برابر است با:
$$F_{ci} = \frac{f_{ci}}{\text{مجموع کل فراوانی‌ها}}$$

و اگر اعداد فوق را در ۱۰۰ ضرب کنیم درصد فراوانی نسبی و فراوانی تجمعی نسبی به دست می‌آید.

نکات:

$$f_{i+1} = f_{ci+1} - f_{ci} \quad (1)$$

(۲) در هر جدول توزیع فراوانی، مجموع فراوانی‌های نسبی برابر ۱ است.

$$\frac{f_1}{\text{مجموع کل فراوانی‌ها}} + \frac{f_2}{\text{مجموع کل فراوانی‌ها}} + \dots + \frac{f_n}{\text{مجموع کل فراوانی‌ها}} = 1$$

(۳) فراوانی تجمعی دسته‌ی آخر برابر است با مجموع فراوانی‌ها، یعنی تعداد کل داده‌ها

(۴) در هر جدول توزیع فراوانی، فراوانی تجمعی نسبی دسته‌ی آخر برابر ۱ است.

مثال: توزیع زیر را در نظر می‌گیریم، درصد فراوانی نسبی متناظر با $x_j = 5$ کدام است؟

x_i	۱	۲	۳	۴	۵	۶
f_i	۴	۶	۸	۷	۳	۲

حل:

$$\text{درصد فراوانی نسبی} = \frac{۳}{۴+۶+۸+۷+۳+۲} \times ۱۰۰ = ۱۰\%$$

مثال: دانش‌آموزان یک مدرسه با سال تولد یکسان را وزن‌کشی کرده و عدد صحیح وزن آنان را یادداشت کرده‌ایم. چند درصد آنان وزن کم‌تر از ۵۰ دارند؟

وزن	۴۶	۴۷	۴۸	۴۹	۵۰	۵۱
تعداد	۸	۹	۱۲	۱۵	۶	۵

(۱) ۷۲

(۲) ۷۵

(۳) ۷۸

(۴) ۸۰

حل:

$$\text{درصد فراوانی تجمعی دسته ی ۴} = \frac{۱۵+۱۲+۹+۸}{۵+۶+۱۵+۱۲+۹+۸} \times ۱۰۰ = \frac{۴۴}{۵۵} \times ۱۰۰ = \frac{۴}{۵} \times ۱۰۰ = ۸۰\%$$

مثال: کوچک‌ترین و بزرگ‌ترین داده‌های آماری ۳۱ و ۵۲ می‌باشند. این داده‌ها در ۷ دسته، دسته‌بندی شده‌اند. ۳۷ درصد داده‌ها کمتر از ۴۰ و ۴۸ درصد آن‌ها بیشتر یا مساوی ۴۳ می‌باشد. اگر فراوانی کل ۸۰ باشد، فراوانی دسته وسط کدام است؟

حل: ۱۵ درصد داده‌ها بین ۴۰ و ۴۳ هستند. چون دامنه ۲۱ و تعداد دسته‌ها ۷ است پس طول دسته‌ها برابر ۳ می‌شود. لذا دسته چهارم همان دسته ۴۳-۴۰ خواهد بود. پس فراوانی دسته وسط برابر است با: $۸۰ \times \frac{۱۵}{۱۰۰} = ۱۲$

مثال: هشتاد داده‌ی آماری در ۷ طبقه دسته‌بندی شده‌اند. اگر ۲۰ داده‌ی جدید به آن‌ها افزوده شود، فراوانی نسبی دسته وسط تغییر نمی‌کند. نسبت افزایش داده‌های دسته‌ی مذکور به فراوانی مطلق قبلی آن کدام است؟

(۱) $\frac{۳}{۸}$ (۲) $\frac{۱}{۵}$ (۳) $\frac{۱}{۴}$ (۴) $\frac{۱}{۸}$

حل: فرض کنیم x واحد از داده‌ها به دسته‌ی وسط افزوده شوند:

$$\frac{f_f + x}{۸۰ + ۲۰} = \frac{f_f}{۸۰} \Rightarrow ۸۰f_f + ۸۰x = ۸۰f_f + ۲۰f_f \Rightarrow \frac{x}{f_f} = \frac{۲۰}{۸۰} = \frac{۱}{۴}$$

مثال: در دسته‌بندی ۱۲۰ داده‌ی آماری در ۹ طبقه، دسته‌ی اول به صورت ۲۵-۲۲ می‌باشد. می‌دانیم ۴۵ درصد داده‌ها کم‌تر از ۳۴ و فراوانی نسبی دسته‌ی وسط $۰/۲$ است. تعداد داده‌های کم‌تر از ۳۷ کدام است؟

(۴) ۸۷

(۳) ۷۸

(۲) ۷۶

(۱) ۶۷

حل:

۲۲-۲۵	۲۵-۲۸	۲۸-۳۱	۳۱-۳۴	۳۴-۳۷	...
-------	-------	-------	-------	-------	-----

تعداد داده‌های کم‌تر از ۳۴:

$$0.45 \times 120 = 54$$

داده‌های بین ۳۴ تا ۳۷:

$$\frac{f_5}{120} = 0.2 \Rightarrow f_5 = 24$$

لذا داده‌های کم‌تر از ۳۷ برابر است با: $54 + 24 = 78$

ز- نشان دسته (مرکز یا نماینده دسته):

فلسفه دسته‌بندی کردن داده‌ها در آمار، دادن ارزش یکسان به داده‌هایی است که در یک دسته قرار می‌گیرند، در داده‌های دسته‌بندی شده، نشان دسته (نماینده دسته) را به صورت زیر در نظر می‌گیریم:

$$x_i^* = \frac{\text{کران بالا} + \text{کران پایین}}{2}$$

$$\text{طول دسته} = x_{i+1}^* - x_i^*$$

مثال: در جدول توزیع فراوانی تعداد طبقات ۸ و طول هر دسته ۳ می‌باشد، در صورتی که نشان دسته‌ی طبقه سوم ۱۸ باشد، نشان دسته‌ی طبقه اول و آخر را بیابید.

حل:

$$x_1^* = x_3^* - 2c \rightarrow x_1^* = 18 - 2 \times 3 = 12$$

$$x_8^* = x_3^* + 5c \rightarrow x_8^* = 18 + 5 \times 3 = 33$$

مثال: کوچک‌ترین و بزرگ‌ترین داده‌های آماری ۱۷/۲ و ۲۲/۶ هستند. اگر کران پایین دسته‌ی دوم ۱۷/۸ باشد، مرکز دسته‌ی

آخر کدام است؟

$$22/4 \text{ (۴)}$$

$$22/3 \text{ (۳)}$$

$$21/8 \text{ (۲)}$$

$$21/7 \text{ (۱)}$$

حل:

$$\text{طول دسته} = 17/8 - 17/2 = 0/6$$

$$22/3 = 0/3 - 0/6 = 22/6 \rightarrow \text{نشان دسته‌ی آخر} = 22/6 - 0/6 = 22/3 \rightarrow \text{تعداد دسته‌ها} = \frac{5/6}{0/6} = 9$$

مثال: در جدول فراوانی داده‌های دسته‌بندی شده، اگر درصد فراوانی نسبی دسته‌ی وسط ۲۴ باشد، فراوانی مطلق دسته‌ی

چهارم کدام است؟

نشان دسته	۱۳	۱۵	۱۷	۱۹	۲۱
فراوانی تجمعی	۵	۱۴	a	۴۱	۵۰

حل:

نشان دسته	۱۳	۱۵	۱۷	۱۹	۲۱
فراوانی تجمعی	۵	۱۴	a	۴۱	۵۰
فراوانی مطلق	۵	۹	a-14	41-a	9

$$\frac{a-14}{50} = \frac{24}{100} = \frac{12}{50} \rightarrow a = 26 \rightarrow f_4 = 41 - 26 = 15$$

مثال: داده‌های جدول مقابل، داده‌های آماری پیوسته است. چند درصد داده‌ها، در فاصله‌ی (۱۸/۵-۲۱/۵) قرار دارند؟

مرکز دسته	۱۴	۱۷	۲۰	۲۳	۲۶
فراوانی تجمعی	۵	۱۳	۲۵	۳۴	۴۰

$$40 \text{ (۴)}$$

$$30 \text{ (۳)}$$

$$25 \text{ (۲)}$$

$$20 \text{ (۱)}$$

حل:

مرکز دسته	۱۴	۱۷	۲۰	۲۳	۲۶
فراوانی تجمعی	۵	۱۳	۲۵	۳۴	۴۰
فراوانی مطلق	۵	۸	۱۲	۹	۶

$$30\% = \frac{12}{40} \times 100\%$$

درصد فراوانی دسته‌ی سوم

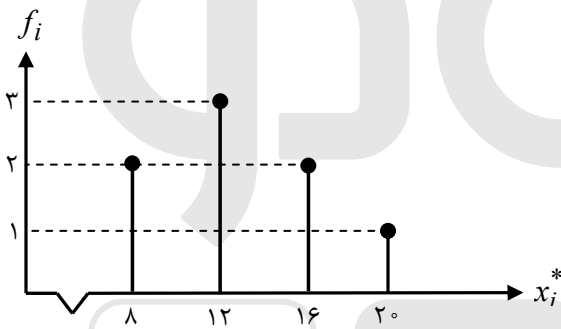
شکاف‌های آماری و تحلیل داده‌ها:

آنچه تاکنون بر روی داده‌ها انجام دادیم، به نوعی سازمان‌دهی آن‌ها برای بهره‌برداری ساده‌تر بود که عمدتاً بر مبنای خلاصه کردن داده‌ها بنا شده بود، اما با تلفیق و به هم آمیختن داده‌ها، سرانجام مقادیر عددی به دست خواهد آمد که به نحو شایسته‌ای می‌تواند جامعه را تحلیل کند. این مقادیر عددی را شاخص‌های عددی می‌گوییم، در مقابل برای تجسم جامعه آماری از شاخص‌های هندسی بهره می‌گیریم.

فصل پنجم: نمودارها و تحلیل داده‌ها (شاخص‌های آماری)

انواع مختلف نمودارهای آماری مفید به قرار زیرند:

۱) **نمودار میله‌ای:** اگر نقاط متناظر با متغیرها (یا نشان دسته‌ها) را روی محور X ها مشخص کنیم و روی هر نقطه پاره‌خطی به ارتفاع فراوانی (مطلق یا نسبی) نظیر آن رسم کنیم، شکل حاصل، نمودار میله‌ای داده‌ها می‌باشد. این نمودار برای متغیرهای کیفی و کمی گسسته مناسب است و بیشتر برای مقایسه‌ی فراوانی داده‌ها استفاده می‌شود.

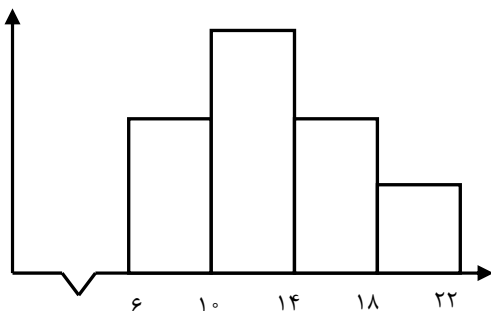


قرارداد: در مواردی که مقیاس بزرگ است و ممکن است تمام نمودار در صفحه قرار نگیرد، نزدیک‌ترین داده به مبدأ را با علامت \sim به مبدأ نزدیک می‌کنیم.

۲) نمودار مستطیلی (هیستوگرام):

هنگامی که جدول توزیع فراوانی در دست است از این نمودار استفاده می‌کنیم. در این نمودار مستطیل‌هایی رسم می‌کنیم که یک ضلع آن منطبق بر دسته‌ها و ضلع دیگر آن، فراوانی (مطلق یا نسبی) دسته متناظر باشد. نمودار مستطیلی برای داده‌های کمی پیوسته مناسب است.

اگر طول دسته‌ها با هم برابر باشند، فراوانی دسته‌ای که مساحت آن بزرگ‌تر است، بیشتر خواهد بود. در این حالت فراوانی متناسب با مساحت مستطیل‌هاست.



نمودار مستطیلی = هیستوگرام، بافت نگار

مثال: جدول زیر جدول فراوانی تجمعی داده‌های آماری دسته‌بندی شده است اگر فراوانی نسبی دسته سوم برابر $0/45$ باشد آن‌گاه مساحت مستطیل مربوط به دسته‌ی چهارم در نمودار مستطیلی این داده‌های آماری کدام است؟

مرکز دسته	۱۷	۲۰	۲۳	۲۶	۲۹	۳۲
فراوانی تجمعی	۴	۱۳	a	۴۹	۵۸	۶۰

که حل:

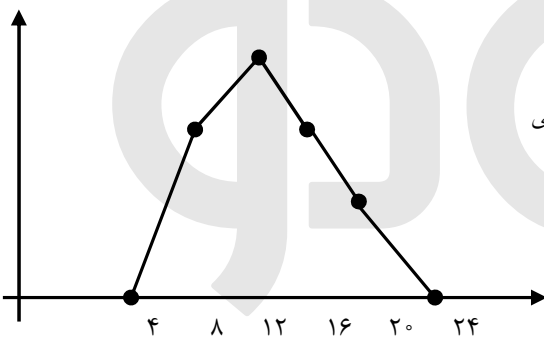
$$\text{طول دسته} = x_2^* - x_1^* = 20 - 17 = 3$$

$$\text{فراوانی نسبی دسته‌ی سوم} = \frac{a-13}{60} = \frac{45}{100} \rightarrow a-13=27 \rightarrow a=40$$

$$\text{مساحت مستطیل دسته‌ی چهارم} = 9 \times 3 = 27 \rightarrow \text{فراوانی مطلق دسته‌ی چهارم} = 49 - a = 9$$

۳) چندبر فراوانی:

در این نمودار نقاطی از صفحه را مشخص می‌کنیم که طول آن‌ها برابر مرکز دسته‌ها و عرض آن‌ها، فراوانی (مطلق یا نسبی) همان دسته است. با به هم وصل کردن این نقاط چندبر فراوانی به دست می‌آید. در چندبر فراوانی، دو دسته با فراوانی صفر به ابتدا و انتهای داده‌ها اضافه می‌کنیم که به این وسیله مساحت چندبر فراوانی با مساحت نمودار مستطیلی متناظرش برابر می‌شود.



مساحت چندبر = مساحت نمودار مستطیلی

نمودار چندبر فراوانی برای داده‌های کمی پیوسته مناسب است.

نمودار چند بر فراوانی را می‌توان با داشتن فراوانی نسبی نیز رسم کرد که آن را چندبر فراوانی نسبی می‌گوییم. در این صورت اطلاعات منسجم‌تری در اختیار ما قرار می‌گیرد، چون می‌توان فراوانی را با کل جامعه مقایسه کرد.

اگر بخواهیم تغییرات متغیر را در فاصله‌ی بین دو دسته یا در خود دسته بهتر نشان دهیم از نمودار چندبر فراوانی استفاده می‌کنیم.

اگر بخواهیم دو جدول توزیع فراوانی را از لحاظ هندسی با هم مقایسه کنیم بهتر است از نمودار چندبر فراوانی استفاده کنیم.

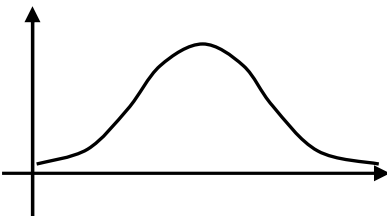
نمودارهای چندبر فراوانی و مستطیلی به منظور نمایش هندسی جدول‌های فراوانی مورد استفاده قرار می‌گیرند، از این رو می‌توان به دلخواه یکی از آن دو را انتخاب کرد. ولی هنگامی که قصد داشته باشیم دو یا چند توزیع را به صورت هندسی با یکدیگر مقایسه کنیم، مشکل می‌توانیم با انطباق مستطیل‌ها بر یکدیگر این مقایسه را انجام دهیم. در چنین مواردی رسم چندبر فراوانی انتخاب مناسب‌تری است زیرا مقایسه‌ی توزیع فراوانی را آسان تر می‌کند.

با افزایش داده‌ها یا کوچکتر شدن حدود دسته‌ها چندبر به یک منحنی شبیه می‌شود از این رو این چندبر فراوانی به منحنی فراوانی تغییر نام می‌دهد.

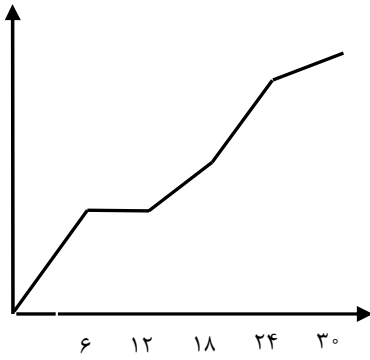
منحنی نرمال: منحنی‌ای است که در اکثر پدیده‌های طبیعی ظاهر می‌شود و شکلی شبیه زنگوله دارد و سطح زیر منحنی آن برابر فراوانی مطلق داده‌هاست. (اگر منحنی را با فراوانی نسبی رسم کنیم سطح زیر منحنی برابر ۱ است)

این منحنی دارای ماکسیمم است و فراوانی در دو طرف این ماکسیمم به طور

یکنواخت به سمت صفر میل می‌کند. این منحنی متقارن است.

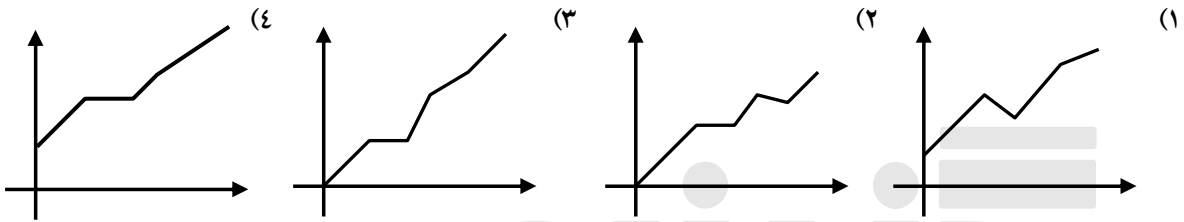


۴) نمودار تجمعی:



هنگامی که جدول توزیع فراوانی در دست است از این نمودار استفاده می‌کنیم. در این نمودار حدود دسته‌ها منطبق بر محور X هاست و در انتهای هر دسته فراوانی تجمعی (مطلق یا نسبی) دسته متناظر قرار می‌گیرد. این نمودار همواره به صورت صعودی است.

مثال: کدام شکل نمودار فراوانی تجمعی است؟



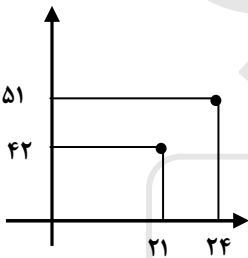
کحل: نمودار تجمعی صعودی است و حتماً از مبدأ شروع می‌شود، لذا فقط گزینه‌ی ۳ می‌تواند جواب صحیح باشد.

مثال: اگر در جدول فراوانی داده‌های پیوسته و طبقه‌بندی شده دو نقطه‌ی $(21, 42)$ و $(24, 51)$ دو نقطه‌ی متوالی از نمودار فراوانی تجمعی باشند، کدام نقطه‌ی زیر روی چندبر فراوانی قرار دارد؟

- (۱) $(21, 51)$ (۲) $(22/5, 9)$ (۳) $(24, 9)$ (۴) $(22/5, 42)$

کحل:

دو نقطه‌ی داده شده مشخص می‌کنند که فراوانی مطلق دسته‌ی $24 - 21$ برابر ۹ است، پس نقطه‌ی $(22/5, 9)$ روی چندبر فراوانی قرار می‌گیرد.



مثال: در رسم نمودار درصد فراوانی تجمعی داده‌های پیوسته‌ی دسته‌بندی شده، دو نقطه‌ی متوالی $(44, 55)$ و $(47, 67)$ از روی جدول رسم شده‌اند. اگر فراوانی کل ۷۵ باشد، چند داده بین ۴۴ و ۴۷ قرار دارد؟

- (۱) ۸ (۲) ۹ (۳) ۱۰ (۴) ۱۲

کحل:

درصد فراوانی مطلق دسته $47 - 44 = 3$ برابر $3/75 = 4\%$ است یعنی 12% داده‌ها در این محدوده‌اند. $12\% \times 75 = 9$

۵) نمودار دایره‌ای:

برای رسم نمودار دایره‌ای، مساحت دایره را به وسیله‌ی قطاع‌هایی به نسبت فراوانی‌ها تقسیم می‌کنیم. اگر فراوانی معلوم باشد، زاویه‌ی هر قطاع از رابطه‌ی $\theta_i = \frac{f_i}{N} \times 360^\circ$ به دست می‌آید. نمودار دایره‌ای امکان مقایسه بین فراوانی‌ها را با سرعت بیشتر فراهم می‌کند.

مثال: 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 زاویه‌ی قطاع ۶۳ درجه نشان داده می‌شود. تعداد اعضای این گروه کدام است؟

کحل:

$$\theta_i = \frac{f_i}{N} \times 360^\circ = 63^\circ \Rightarrow f_i = \frac{63}{360} \times N = \frac{7}{40} \times 540000 = 94500$$

مثال: جدول مقابل درصد فراوانی نسبی گروه خونی افراد یک جامعه است. در نمودار دایره‌ای، زاویه‌ی سطح مربوط به گروه

گروه خونی	A	B	AB	O
درصد فراوانی نسبی	۲۴	۲۲/۵	۳۶	α

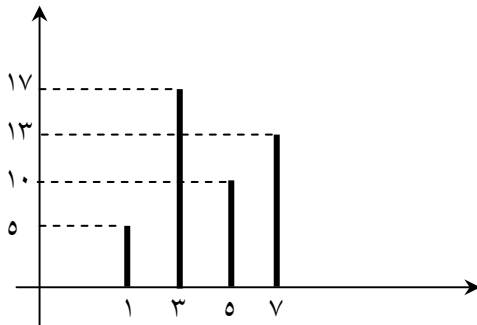
خونی O چند درجه است؟

کحل:

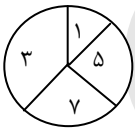
$$\alpha = 17/5$$

$$\theta = \frac{17/5}{100} \times 360^\circ = 63^\circ$$

مثال: نمودار دایره‌ای متناظر با نمودار میله‌ای روبه‌رو را رسم کنید.



کحل:

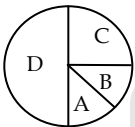


$$\theta_1 = \frac{5}{45} \times 360^\circ = 40^\circ$$

$$\theta_2 = \frac{17}{45} \times 360^\circ = 136^\circ$$

$$\theta_3 = \frac{10}{45} \times 360^\circ = 80^\circ$$

$$\theta_4 = \frac{13}{45} \times 360^\circ = 104^\circ$$



مثال: در نمودار زیر تعداد افرادی که در دسته‌های B, C و D قرار دارند، به ترتیب ۲، ۳ و ۶ برابر تعداد

افرادی است که در دسته‌ی A قرار دارند. زاویه‌ی A کدام است؟

کحل:

$$\theta_A + \theta_B + \theta_C + \theta_D = \theta_A + 2\theta_A + 3\theta_A + 6\theta_A = 12\theta_A = 360^\circ \Rightarrow \theta_A = 30^\circ$$

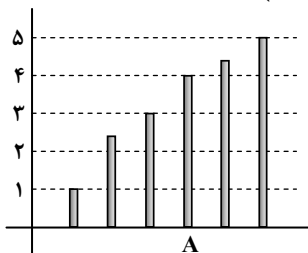
مثال: نمودار دایره‌ای داده‌های x_1, \dots, x_n را با نمودار دایره‌ای داده‌های $x_1, x_1, \dots, x_n, x_n$ مقایسه کنید.

کحل:

چون فراوانی نسبی داده‌ها تغییر نکرده پس نمودار دایره‌ای تغییر نمی‌کند.

مثال: در مقایسه‌ی سطح زیر کشت غله‌ای در شش استان، نمودار میله‌ای مقابل رسم شده است. در نمودار دایره‌ای، زاویه‌ی

مرکزی متناظر استان A چند درجه است؟ (قسمت غیر صحیح دو میله‌ی دوم و پنجم ۰/۵ است).



۷۲ (۲)

۶۴ (۱)

۹۶ (۴)

۸۰ (۳)

کحل:

$$n = 1 + 2/5 + 3 + 4 + 4/5 + 5 = 20 \Rightarrow \theta_A = \frac{4}{20} \times 360 = 72$$

مثال: داده‌های آماری در ۹ طبقه دسته‌بندی شده‌اند. فراوانی تجمعی نسبی در دسته‌ی چهارم و پنجم به ترتیب $0/40$ و $0/28$ است. در نمودار دایره‌ای زاویه‌ی مربوط به دسته‌ی پنجم چند درجه است؟
 کحل:

$$\text{فراوانی نسبی دسته پنجم} = 0/4 - 0/28 = 0/12 \Rightarrow \theta = 0/12 \times 360^\circ = 43/2^\circ$$

مثال: در جدول فراوانی تجمعی داده‌های دسته‌بندی شده به شکل زیر، زاویه‌ی مرکزی متناسب با فراوانی مطلق دسته‌ی وسط در نمودار دایره‌ای، 90 درجه است. فراوانی مطلق دسته‌ی چهارم کدام است؟

حدود دسته	۱۲-۱۴	۱۴-۱۶	۱۶-۱۸	۱۸-۲۰	۲۰-۲۲
فراوانی تجمعی	۶	۱۷	x	۴۸	۶۰

(۱) ۱۴

(۲) ۱۵

(۳) ۱۶

(۴) ۱۸

کحل:

حدود	۱۲-۱۴	۱۴-۱۶	۱۶-۱۸	۱۸-۲۰	۲۰-۲۲
فراوانی تجمعی	۶	۱۷	x	۴۸	۶۰
فراوانی مطلق	۶	۱۱	x-۱۷	۴۸-x	۱۲

$$\theta_3 = \frac{x-17}{60} \times 360 = 6(x-17) = 90 \Rightarrow x-17=15 \Rightarrow x=32$$

$$f_4 = 48 - 32 = 16$$

مثال: در جدول مقابل مرکز دسته با درصد فراوانی نسبی داده شده است. در نمودار دایره‌ای زاویه‌ی مربوط به بازه‌ی $[25, 28)$ چند درجه است؟

(۱) ۷۲

(۲) ۸۱

(۳) ۸۴

(۴) ۹۰

کحل:

مرکز دسته	۱۷/۵	۲۰/۵	۲۳/۵	۲۶/۵	۲۹/۵
درصد فراوانی نسبی	۱۷	۲۰/۵	۲۲	x	۱۸

$$17 + 20/5 + 22 + x + 18 = 100 \Rightarrow x = 22/5$$

$$\theta = \frac{22/5}{100} \times 360 = 81^\circ$$

مثال: شرکتی ۱۶۰ کارمند دارد که مدارک تحصیلی آنان با ۶ کد متمایز مشخص شده‌اند. در نمودار دایره‌ای، زاویه‌ی مرکزی هر گروه با واحد درجه مطابق جدول روبه‌رو است. تعداد کارکنان با کد ۴ کدام است؟

(۱) ۵۲

(۲) ۵۴

(۳) ۵۶

(۴) ۵۸

کحل:

کد	۱	۲	۳	۴	۵	۶
زاویه مرکزی	۲۷	۴۵	۹۹	α	۵۴	۱۸

$$27 + 45 + 99 + \alpha + 54 + 18 = 360$$

$$\Rightarrow \alpha = 117 \Rightarrow \theta_i = \frac{f_i}{n} \times 360 = 117 \Rightarrow \theta_4 = \frac{f_4}{160} \times 360 = 117 \Rightarrow f_4 \times \frac{9}{4} = 117 \Rightarrow f_4 = \frac{117 \times 4}{9} = 52$$

۶) نمودار ساقه و برگ:

برای آن که اعداد را به گونه‌ای مرتب کنیم که چیزی شبیه نمودار میله‌ای به دست آید، از نمودار ساقه و برگ استفاده می‌کنیم. برای تهیه نمودار ساقه و برگ، داده‌ها را به دو بخش تقسیم می‌کنیم. ساقه شامل یک یا چند رقم اولیه و برگ که با فاصله‌ی کمی از ساقه جلوی آن به صورت صعودی نوشته می‌شود، شامل ارقام باقی‌مانده است. اگر هم عددی تکرار شود به دفعات تکرار آن را می‌نویسیم. برای فهم آن که چند رقم ساقه و چند رقم برگ است گاهی کلید نمودار را به صورت زیر مشخص می‌کنند:

ساقه	برگ
۳۲	۱ ۴ ۵ ۷ ۸ ۸ ۹
۳۳	۰ ۰ ۴ ۵ ۵ ۶
۳۴	۲ ۳ ۶ ۶ ۷

کلید: ۹ : ۲۳ : ۳۳۹

این نمودار به خوبی فاصله‌ی کمترین و بیشترین داده را نشان می‌دهد.

مثال: در نمودار ساقه و برگ زیر چه اعدادی به جای ؟ می‌توانند قرار گیرند؟

ساقه	برگ
۱	۱ ۴ ؟ ۷ ۸ ۸ ۹
۲	۰ ۰ ۴ ۵ ؟ ۶
۳	۲ ۳ ۶ ۶ ۷

حل:

$$\left. \begin{array}{l} 4 \leq ? \leq 7 \\ 5 \leq ? \leq 6 \end{array} \right\} \Rightarrow ? = 5 \text{ یا } 6$$

مثال: 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 به صورت ۰۳۵۵۵۵۶۷۷۹، ۱۴ نمایش داده شده «چند درصد داده‌ها» برابر ۱۴/۵ بوده‌اند؟

حل:

این داده‌ها عبارتند از: ۱۴، ۱۴/۳، ۱۴/۵، ۱۴/۵، ۱۴/۵، ۱۴/۵، ۱۴/۶، ۱۴/۷، ۱۴/۷، ۱۴/۹

پس $\frac{4}{50}$ یا ۸٪ داده‌ها عدد ۱۴/۵ بوده‌اند.

۱- شافص‌های عددی:

معمولاً علاقمندیم به کمک مقادیری تمرکز داده‌ها را نشان دهیم. برای این منظور باید بدانیم، داده‌ها حول چه نقطه‌ای تجمع پیدا کرده‌اند و ضمناً چگونه تجمع پیدا کرده‌اند. آیا داده‌ها به هم نزدیکند یا از هم دورند. برای منظور اول از شاخص‌های مرکزی استفاده می‌کنیم که نشان‌دهنده‌ی مرکز تجمع داده‌ها می‌باشند و برای منظور دوم از شاخص‌های پراکندگی استفاده می‌کنیم که نشان‌دهنده‌ی میزان تجمع داده‌ها حول مرکز است.

فصل ششم: شاخص‌های مرکزی

شاخص‌های مرکزی عبارتند از: میانگین، میانه و مد

۱) میانگین:

میانگین n داده‌ی آماری X_1, X_2, \dots, X_n برابر است با:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

مثال: داده‌های آماری با یک رقم اعشار با نمودار ساقه و برگ داده شده‌اند، میانگین آن‌ها کدام است؟

ساقه	برگ
۸	۰ ۰ ۱ ۲ ۲ ۵ ۶ ۷
۹	۰ ۱ ۲ ۳ ۳ ۴ ۵ ۵
۱۰	۱ ۱ ۲ ۲

حل:

$$\bar{x} = \frac{8 \times 8 + 8 \times 9 + 4 \times 10 + 0 / (1 + 2 + 2 + 5 + 6 + 7 + 1 + 2 + 3 + 3 + 4 + 5 + 5 + 1 + 1 + 2 + 2)}{20} = 9 / 0.6$$

مثال: اگر میانگین داده‌های x_1, x_2, x_3, x_4 برابر \bar{X} باشد، میانگین داده‌های $x_1, 2x_2, 3x_3, 4x_4$ کدام است؟

حل:

$$\bar{x} = \frac{2x_1 + x_2 + 2x_2 + x_3 + 2x_3 + x_4 + 2x_4 + x_1}{4} = \frac{2(x_1 + x_2 + x_3 + x_4)}{4} = 2\bar{x} \text{ قدیم}$$

نکات:

(۱) در یک جامعه آماری، میانگین عددی منحصر به فرد است که همواره بین بزرگ‌ترین و کوچک‌ترین داده قرار دارد.

(۲) اگر داده‌های آماری تشکیل تصاعد حسابی بدهند، میانگین آن‌ها عبارت است از:

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

اولین داده ← ← آخرین داده

(۳) مفیدترین خاصیت میانگین آن است که اگر به جای تک تک داده‌ها، میانگین آن‌ها را قرار دهیم، مجموع و میانگین داده‌ها تغییر نخواهد کرد.

میانگین \times تعداد کل داده‌ها = مجموع کل داده‌ها

(۴) اگر میانگین m داده‌ی آماری \bar{X} و میانگین n داده‌ی آماری \bar{Y} باشد، میانگین کلیه داده‌ها برابر است با: $\frac{m\bar{X} + n\bar{Y}}{m + n}$ و از

همین جا می‌توان نتیجه گرفت اگر میانگین m_1 داده‌ی آماری \bar{X}_1 ، میانگین m_2 داده‌ی آماری \bar{X}_2 ، ... و میانگین m_n داده‌ی آماری

$$\bar{X} = \frac{m_1\bar{X}_1 + m_2\bar{X}_2 + \dots + m_n\bar{X}_n}{m_1 + m_2 + \dots + m_n}$$

باشد، میانگین کل داده‌ها برابر است با:

مثال: میانگین داده‌های x_1, x_2, \dots, x_{10} برابر ۵ است. میانگین مقادیر x_1, x_2, \dots, x_{16} کدام است؟

حل:

$$\frac{x_1 + x_2 + \dots + x_{10}}{10} = 5 \Rightarrow x_1 + x_2 + \dots + x_{10} = 50 \Rightarrow \bar{x} = \frac{50 + 16}{11} = \frac{66}{11} = 6$$

مثال: اگر میانگین نمرات یک کلاس ۳۰ نفری ۱۲ و میانگین یک کلاس ۲۰ نفری ۱۶ باشد، میانگین نمرات این ۵۰

دانش آموز چقدر است؟

حل:

$$\bar{x} = \frac{20 \times 12 + 30 \times 16}{20 + 30} = \frac{360 + 480}{50} = 13.2$$

۵) اگر میانگین داده‌های x_1, x_2, \dots, x_n ، \bar{X} باشد، میانگین داده‌های $ax_1 + b, ax_2 + b, \dots, ax_n + b$ برابر است با: $a\bar{X} + b$
 مثال: میانگین چند داده برابر ۵۷ است. ابتدا از هر داده ۱۲ واحد کم و سپس داده‌های حاصل را سه برابر کرده‌ایم. میانگین داده‌های نهایی کدام است؟
 کحل:

$$\bar{X} = 57$$

$$y_i = 3(x_i - 12) = 3x_i - 36$$

$$\bar{Y} = 3\bar{X} - 36 = 135 \text{ پس: } \bar{Y} = 3\bar{X} - 36$$

$$(6) \text{ مجموع تفاضل داده‌ها از میانگین برابر صفر است. } (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$$

مماسبهی میانگین در جدول فراوانی: (میانگین وزن دار)

میانگین f_1 تا x_1 ، f_2 تا x_2 ، ...، f_n تا x_n برابر است با:

$$\bar{X} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_n}$$

$$\text{که در این صورت: } f_1(x_1 - \bar{x}) + f_2(x_2 - \bar{x}) + \dots + f_n(x_n - \bar{x}) = 0$$

اگر جدول دسته‌بندی شده داده‌ها در دسترس باشد، برای محاسبه‌ی میانگین به‌جای داده‌ها در رابطه‌ی فوق از مرکز دسته بهره می‌بریم.
 مثال: جدول زیر مقادیر انحراف از میانگین داده‌های آماری دسته‌بندی شده را مشخص می‌کند. فراوانی مطلق در دسته‌ی ششم چقدر است؟

انحراف از میانگین	-۴	-۲	-۱	۰	۱	۲	۳
فراوانی مطلق	۵	۱۱	۹	۴	۸	x	۳

کحل:

$$(-4 \times 5) + (-2 \times 11) + (-1 \times 9) + 1 \times 8 + 2x + 3 \times 3 = 0 \Rightarrow 2x = 34 \Rightarrow x = 17$$

مثال: میانگین داده‌های جدول زیر را به دست آورید.

دسته	۲-۴	۴-۶	۶-۸	۸-۱۰
فراوانی	۲	۴	۳	۱

کحل:

$$\bar{X} = \frac{3 \times 2 + 5 \times 4 + 7 \times 3 + 9 \times 1}{2 + 4 + 3 + 1} = \frac{6 + 20 + 21 + 9}{10} = 5/6$$

مثال: در جدول فراوانی مقابل، میانگین به صورت $\bar{X} = 12 + 2\alpha$ محاسبه شده است. α کدام است؟

x	۸	۱۰	۱۲	۱۴	۱۶
f	۲	۵	۵	۹	۳

کحل:

$$\bar{\alpha} = \frac{\bar{X} - 12}{2}$$

یعنی اگر از همه‌ی داده‌ها ۱۲ واحد کم کرده و آن‌ها را بر ۲ تقسیم کنیم، میانگین جدید چیست؟

$$\Rightarrow \bar{X} = \frac{1 \times 9 + 2 \times 3 + (-1) \times 5 + (-2) \times 2}{2 + 5 + 5 + 9 + 3} = \frac{6}{24} = \frac{1}{4}$$

مثال: در جدول فراوانی داده‌های دسته‌بندی شده زیر اگر به تمام داده‌ها ۱/۵ واحد اضافه شود، میانگین داده‌های جدید، برابر ۱۰ می‌شود. فراوانی دسته سوم کدام است؟

حدود دسته	۱-۵	۵-۹	۹-۱۳	۱۳-۱۷
فراوانی	۴	۵	a	۳

حل:

$$\bar{Y} = \bar{X} + 1/5 = 10 \Rightarrow \bar{X} = 8/5$$

$$\Rightarrow \bar{X} = \frac{3 \times 4 + 5 \times 5 + 11 \times a + 15 \times 3}{4 + 5 + a + 3} = 8/5 \Rightarrow \frac{11a + 92}{12 + a} = 8/5 \Rightarrow a = 4$$

مثال: میانگین ۵۰ داده‌ی دسته‌بندی شده‌ی زیر با روش سریع کدام است؟

x	۱۱۰	۱۱۶	۱۲۲	۱۲۸	۱۳۴
f	۵	۸	۱۵	۱۲	۱۰

حل:

چون داده‌ها دنباله‌ی حسابی می‌سازند، ابتدا همه‌ی داده‌ها را از ۱۲۲ کم می‌کنیم و سپس داده‌های حاصل را بر ۶ تقسیم می‌کنیم:

$y = \frac{x-122}{6}$	-۲	-۱	۰	۱	۲
f	۵	۸	۱۵	۱۲	۱۰

$$\Rightarrow \bar{Y} = \frac{1 \times 12 + 2 \times 10 - 1 \times 8 - 2 \times 5}{10 + 12 + 15 + 8 + 5} = \frac{14}{50} = 0.28$$

حال میانگین داده‌های قبلی عبارتست از:

$$\bar{X} = 6\bar{Y} + 122 = 6 \times 0.28 + 122 = 123.68$$

مثال: در جدول فراوانی تجمعی داده‌های آماری زیر اگر میانگین جامعه ۴۱ باشد، در نمودار دایره‌ای زاویه‌ی مربوط به دسته‌ی (۳۹, ۴۲] چند درجه است؟

حل:

ابتدا از همه‌ی داده‌ها ۴۱ واحد کم کرده و آن‌ها را بر ۴ تقسیم می‌کنیم، سپس فراوانی مطلق دسته‌ی سوم را به دست می‌آوریم:

نماینده	-۲	-۱	۰	۱	۲
فراوانی مطلق	۷	۱۰	۱۵	۱۲	a-۴۴

$$\Rightarrow \bar{X} = \frac{1 \times 12 + 2(a-44) - 1 \times 10 - 2 \times 7}{a} = \frac{2a-100}{a}$$

$$\Rightarrow \bar{Y} = 4\bar{X} + 41 = 4\left(2 - \frac{100}{a}\right) + 41 = 41 \Rightarrow a = 50 \Rightarrow \theta_i = \frac{15}{50} \times 360^\circ = 108^\circ$$

(۷) میانه:

اگر داده‌های آماری را به صورت صعودی یا نزولی مرتب کنیم، در صورتی که تعداد داده‌ها فرد باشد، عددی که در وسط قرار می‌گیرد، و در صورتی که تعداد داده‌ها زوج باشد، نصف مجموع دو عددی که در وسط قرار می‌گیرند، میانه نام دارد. معمولاً میانه مقداری است که تعداد اعضایی از جامعه که از آن بیشترند برابر تعداد اعضایی است که از آن کمترند. (نماد میانه: M و \tilde{x})

نکته: اگر میانه داده‌های X_1, X_2, \dots, X_n برابر M باشد، میانه‌ی داده‌های $ax_1 + b, \dots, ax_n + b$ برابر است با: $aM + b$

مثال: در یک امتحان ریاضی نمرات ۱۵ نفر به صورت زیر است:

۴, ۷, ۷, ۳, ۱۲, ۱۱, ۱۷, ۱۵, ۱۴, ۱۷, ۱۹, ۱۴, ۱۰, ۹, ۵

میانه این نمرات کدام است؟

حل:

۳, ۴, ۵, ۷, ۷, ۹, ۱۰, ۱۱, ۱۲, ۱۴, ۱۴, ۱۵, ۱۷, ۱۷, ۱۹

چون ۱۵ داده داریم، داده‌ی ۸ام میانه است. پس میانه ۱۱ است.

برای n داده‌ی فرد، میانه $\frac{n+1}{2}$ امین داده است.

نکته: اگر جدول توزیع فراوانی برای داده‌های دسته‌بندی نشده در دسترس باشد، برای پیدا کردن میانه، ابتدا فراوانی تجمعی را

به دست آورده و سپس اولین ردیفی که فراوانی تجمعی آن بزرگ‌تر از یا مساوی با $\frac{n+1}{2}$ می‌باشد، ردیف میانه خواهد بود.

مثال: در جدول زیر میانه را به دست آورید.

x_i	۲	۳	۴	۵	۶	۷
f_i	۲	۲	۳	۵	۴	۴
x_i	۲	۳	۴	۵	۶	۷
f_i	۲	۲	۳	۵	۴	۴
f_{c_i}	۲	۴	۷	۱۲	۱۶	۲۰

حل: ابتدا در جدول فراوانی تجمعی را به دست می‌آوریم:

چون ۲۰ داده داریم، میانه جمع داده‌ی ۱۰ام و ۱۱ام است که هر ۲ در دسته‌ی $x_i = 5$ قرار دارند (داده‌های ۷ تا ۱۲ همگی ۵ اند)

پس میانه برابر $\frac{5+5}{2} = 5$ است.

مثال: در ۸۰ داده‌ی آماری دسته‌بندی شده، فراوانی نسبی دسته‌ی اول $\frac{10}{1125}$ می‌باشد. اگر ۱۰ داده‌ی دیگر بزرگ‌تر از میانه

به آن‌ها افزوده شود، درصد فراوانی نسبی جدید در دسته‌ی اول کدام است؟

حل: چون میانه‌ی داده‌ها میانگین داده‌ی ۴۰ام و ۴۱ام است، لذا ۱۰ داده‌ی اضافه شده قطعاً داده‌ای به دسته‌ی اول اضافه نخواهد کرد، پس:

$$F_1 = \frac{9}{80+10} = \frac{9}{90} = \frac{1}{10} \Rightarrow \text{درصد فراوانی نسبی دسته‌ی اول} = 10\%$$

پارک‌ها:

چارک اول: در واقع عددی است که از $\frac{1}{4}$ داده‌ها بزرگ‌تر و از $\frac{3}{4}$ داده‌ها کوچک‌تر است و با Q_1 نمایش داده می‌شود.

چارک دوم: که با Q_2 نمایش داده می‌شود همان میانه است.

چارک سوم: عددی است که از $\frac{3}{4}$ داده‌ها بزرگ‌تر و از $\frac{1}{4}$ داده‌ها کوچک‌تر است و با Q_3 نمایش داده می‌شود.

برای یافتن چارک اول و چارک سوم، پس از مرتب کردن داده‌ها، میانه را پیدا می‌کنیم و سپس مجدداً برای داده‌های قبل و بعد از میانه، میانه را پیدا می‌کنیم.

مثال: چارک‌های اول، دوم و سوم داده‌های زیر را بیابید.

۱۰, ۵, ۷, ۱۲, ۳, ۶, ۹

حل:

۳, ۵, ۶, ۷, ۹, ۱۰, ۱۲

پس میانه یا چارک دوم ۷ است، چارک اول ۵ و چارک سوم ۱۰ است.

مثال: در داده‌های آماری با نمودار ساقه و برگ، داده‌های کمتر از چارک اول و بیشتر از چارک سوم را حذف می‌کنیم.

میانگین داده‌های باقی‌مانده کدام است؟

ساقه	برگ
۳	۱ ۴ ۵ ۷ ۸ ۸ ۹
۴	۰ ۰ ۴ ۵ ۵ ۶
۵	۲ ۳ ۶ ۶ ۷

کحل:

چون ۱۸ داده داریم، میانه‌ی ۹ داده‌ی اول داده‌ی ۱۵ و ۹ داده‌ی دوم، داده‌ی ۱۴ام است، یعنی ۳۸ چارک اول و ۵۲ چارک سوم است. پس میانگین داده‌های باقی‌مانده عبارت است از: $42/7$

$$\bar{X} = \frac{(38 + 38 + 39 + 40 + 40 + 44 + 45 + 45 + 46 + 52)}{10} = 42/7$$

مثال: داده‌های آماری در ۹ طبقه با طول دسته‌ی ۴، دسته‌بندی شده‌اند. اگر ۸ داده بین چارک اول و سوم به داده‌ها اضافه و یک واحد از طول دسته‌ها کم کنیم، در دسته‌بندی جدید تعداد دسته‌ها کدام است؟

کحل:

دامنه‌ی داده‌ها $9 \times 4 = 36$ است. حال اگر طول دسته‌ها برابر ۳ باشد، تعداد دسته‌ها $36/3 = 12$ می‌شود. اضافه شدن تعدادی داده داخل دامنه، تأثیری بر حدود دسته نمی‌گذارد.

نمودار جعبه‌ای:

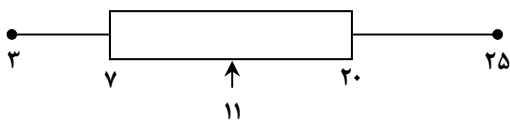
برای نشان دادن بهتر پراکندگی از این نمودار استفاده می‌کنیم. برای تهیه‌ی این نمودار کوچکترین و بزرگترین داده را پیدا می‌کنیم. چارک اول و سوم را می‌یابیم سپس با رسم جعبه‌ای بین چارک اول و سوم و وصل کردن دو سر جعبه به بزرگترین و کوچکترین داده، نمودار را رسم می‌کنیم.



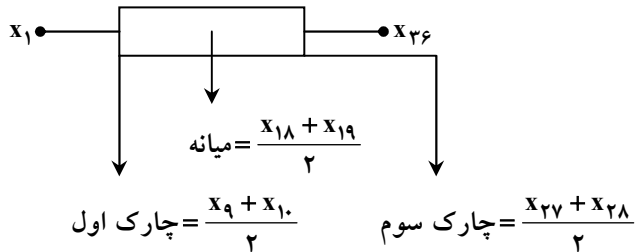
مثال: نمودار جعبه‌ای داده‌های زیر را رسم کنید: ۱۰, ۳, ۲۰, ۱۳, ۷, ۱۱, ۲۵

کحل:

۳, ۷, ۱۰, ۱۱, ۱۳, ۲۰, ۲۵



مثال: در نمودار جعبه‌ای ۳۶ داده‌ی آماری، میانگین داده‌های دو طرف جعبه جداگانه به ترتیب ۲۲ و ۳۰ می‌باشد. اگر میانگین تمام داده‌ها ۲۷/۵ باشد، آن‌گاه میانگین داده‌های داخل جعبه کدام است؟
 کحل:



یعنی از هر طرف جعبه ۹ داده بیرون است و ۱۸ داده درون جعبه قرار دارد.

$$\frac{9 \times 22 + 9 \times 30 + 18 \bar{X}}{36} = 27/5 \Rightarrow \frac{22 + 30 + 2\bar{X}}{4} = 27/5 \Rightarrow \bar{X} = 29$$

(۳) مد (نما):

مقدار یا مقادیری از متغیر که در آن‌ها فراوانی ماکزیمم باشد، مد نامیده می‌شود (\hat{x}). در واقع مد داده‌ای است که بیشترین تکرار را در میان داده‌ها داشته باشد و توجه کنید که مد منحصر به فرد نیست.

معمولاً مد برای تحلیل متغیرهای کیفی به کار می‌رود.

اگر \hat{x} مد داده‌های x_1, \dots, x_n باشد، مد داده‌های $ax_1 + b, \dots, ax_n + b$ برابر است با $a\hat{x} + b$. اگر جدول داده‌های دسته‌بندی شده در دسترس باشد، نشان دسته طبقه‌ای که بیشترین فراوانی را دارد، مد می‌باشد.

فصل هفتم: شاخص‌های پراکندگی

شاخص‌های مرکزی در مورد پراکندگی داده‌ها اطلاعاتی به ما نمی‌دهند لذا به تعریف شاخص‌هایی برای پراکندگی می‌پردازیم.

(۱) دامنه‌ی تغییرات:

با تعریف دامنه‌ی تغییرات آشنا شدیم. $R = x_{\max} - x_{\min}$ دامنه‌ی تغییرات برای بیان پراکندگی اطلاعات زیادی به ما نمی‌دهد زیرا فقط بزرگ‌ترین و کوچک‌ترین داده در آن مؤثرند.

مثال: اگر کلیه‌ی داده‌های آماری در عدد ثابتی ضرب و با عدد ثابتی جمع شوند، دامنه‌ی آن چه تغییری می‌کند؟

کحل:

$$R = |x_{\max} - x_{\min}|$$

$$R' = |(ax_{\max} + b) - (ax_{\min} + b)| = |a|R$$

اگر $a < 0$ باشد، جای داده ماکزیمم و مینیمم برعکس می‌شود.

مثال: چرا دامنه‌ی تغییرات به‌عنوان معیار پراکندگی مناسب نیست؟

(۱) با ازدیاد داده‌ها فقط می‌تواند کم شود. (۲) مقدار آن اغلب بزرگ است.

(۳) مقدار آن از یک نمونه به نمونه دیگری تغییر می‌کند. (۴) در محاسبه‌ی آن فقط از دو اندازه استفاده می‌شود.

کحل: دامنه فقط تابع بزرگ‌ترین و کوچک‌ترین داده است.

۲) واریانس (پراش):

بهترین پارامتر پراکندگی، واریانس است.

برای نشان دادن پراکندگی باید از مفهوم انحراف از میانگین بهره بگیریم، اما چون مجموع انحراف از میانگین‌ها در یک جامعه آماری صفر می‌باشد، مجموع مربع‌های آن‌ها را در نظر می‌گیریم. اگر این مجموع را بر تعداد داده‌ها تقسیم کنیم واریانس (پراش) داده‌ها حاصل می‌شود.

برای داده‌های x_1, x_2, \dots, x_n با میانگین \bar{x} داریم:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

اگر جدول داده‌های طبقه‌بندی نشده در دسترس باشد، از رابطه‌ی زیر استفاده می‌کنیم.

x_i	x_1	x_2	...	x_n
f_i	f_1	f_2	...	f_n

$$\sigma^2 = \frac{f_1(x_1 - \bar{X})^2 + f_2(x_2 - \bar{X})^2 + \dots + f_n(x_n - \bar{X})^2}{f_1 + f_2 + \dots + f_n}$$

اگر جدول داده‌های دسته‌بندی شده در دسترس باشد از مرکز هر دسته به‌عنوان نماینده‌ی آن استفاده می‌کنیم تا مسأله به حالت قبلی تبدیل شود.

مثال: در داده‌های آماری با نمودار ساقه و برگ مقابل، واریانس داده‌های کم‌تر از مد و بیشتر از میانه، کدام است؟

ساقه	برگ					
۲	۰	۲	۳	۵	۶	۸
۳	۲	۴	۶	۷	۹	
۴	۴	۵	۵	۶		

حلول: چون ۴۵ دو بار تکرار شده است، پس مد جامعه است. چون تعداد داده‌ها ۱۵ است، پس داده‌ی $\frac{15+1}{2} = 8$ ام میانه است.

است، که داده‌ی هشتم ۳۴ است، پس داده‌های مورد نظر عبارتند از: ۳۶، ۳۷، ۳۹، ۴۴.

$$\bar{X} = \frac{36+37+39+44}{4} = 39 \Rightarrow \sigma^2 = \frac{3^2+2^2+5^2}{4} = 9/5$$

مثال: واریانس داده‌های جدول زیر کدام است؟

x_i	۲	۳	۴	۵	۶	۷
f_i	۱	۱	۳	۵	۴	۲

حلول:

$$\bar{X} = \frac{2 \times 1 + 3 \times 1 + 4 \times 3 + 5 \times 5 + 6 \times 4 + 7 \times 2}{1+1+3+5+4+2} = 5$$

$$\sigma^2 = \frac{1(5-2)^2 + 1(3-5)^2 + 3(4-5)^2 + 5(5-5)^2 + 4(6-5)^2 + 2(7-5)^2}{16} = 1/75$$

مثال: واریانس داده‌های آماری جدول زیر را بیابید.

x_i	۰-۲	۲-۴	۴-۶	۶-۸
f_i	۱	۲	۹	۴

حلول:

$$\bar{X} = \frac{1 \times 1 + 2 \times 3 + 9 \times 5 + 4 \times 7}{1+2+9+4} = 5$$

$$\sigma^2 = \frac{(1-5)^2 + 2(3-5)^2 + 9(5-5)^2 + 4(7-5)^2}{16} = 2/5$$

مثال: در داده‌هایی با جدول فراوانی زیر اگر واریانس برابر ۶ باشد، فراوانی دسته‌ی سوم کدام است؟

حدود دسته	۵-۷	۷-۹	۹-۱۱	۱۱-۱۳	۱۳-۱۵
فراوانی	۳	۲	A	۶	۱

کحل:

ابتدا میانگین را می‌یابیم:

$$\bar{X} = \frac{6 \times 3 + 8 \times 2 + 10A + 12 \times 6 + 14 \times 1}{3 + 2 + A + 6 + 1} = \frac{10A + 120}{A + 12} = 10$$

$$\sigma^2 = \frac{3(2)^2 + 2(2)^2 + 6(2)^2 + 1(4)^2}{12 + A} = 6 \Rightarrow \frac{96}{12 + A} = 6 \Rightarrow 6A = 96 - 72 \Rightarrow A = 4$$

مثال: داده‌های آماری کدام گزینه واریانس بزرگ‌تری دارند؟

۱۰, ۱۰, ۱۰, ۱۰ (۴)

۱۳, ۱۲, ۸, ۷ (۳)

۱۰, ۱۰, ۱۶, ۶ (۲)

۱۳, ۱۳, ۷, ۷ (۱)

کحل: تقریباً میانگین همه‌ی گزینه‌ها برابر ۱۰ است. پس گزینه‌ی ۲ دارای پراکندگی بیش‌تری است، چون فاصله‌ی داده‌ها از میانگین بزرگ‌تر است.

مثال: دو نفر در یک آزمایشگاه در ۵ روز متوالی همزمان شروع به کار کرده‌اند. امتیازات دقت کاری آن‌ها مطابق جدول زیر است. دقت کاری کدام بیشتر است؟

نفر اول	۷	۹	۸	۹	۷
نفر دوم	۱۰	۸	۶	۷	۹

(۱) نفر اول

(۲) نفر دوم

(۳) یکسان

(۴) نیاز به اطلاعات بیشتر

کحل:

دقت کاری کسی بیش‌تر است که پراکندگی امتیازش، کم‌تر باشد نفر اول پراکندگی‌اش کم‌تر است، پس امتیاز دقت کاریش بیش‌تر است. (هر ۲ نفر میانگین یکسانی دارند)

نکات:

(۱) برای داده‌های x_1, x_2, \dots, x_n با میانگین \bar{x} ، واریانس را می‌توان از دستور زیر نیز محاسبه کرد:

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - (\bar{x})^2$$

نکته: واریانس را می‌توان در جدول فراوانی طبقه‌بندی نشده از رابطه‌ی زیر نیز به‌دست آورد.

x_i	x_1	...	x_n
f_i	f_1	...	f_n

$$\sigma^2 = \frac{f_1(x_1 - \bar{X})^2 + f_2(x_2 - \bar{X})^2 + \dots + f_n(x_n - \bar{X})^2}{f_1 + f_2 + \dots + f_n}$$

مثال: اگر $x_1^2 + x_2^2 + \dots + x_n^2 = 100$ و $x_1 + x_2 + \dots + x_n = 20$ ، واریانس داده‌ها کدام است؟

کحل:

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - (\bar{x})^2 = \frac{100}{10} - \left(\frac{20}{10}\right)^2 = 6$$

(۲) اگر واریانس داده‌های آماری x_1, \dots, x_n ، σ^2 باشد، واریانس داده‌های $ax_1 + b, \dots, ax_n + b$ برابر است با $a^2\sigma^2$.

مثال: اگر واریانس در سال گذشته ۱۰۰۰ ریال باشد و امسال ۱۰٪ به قیمت‌ها افزوده شده باشد، واریانس قیمت‌های جدید چقدر است؟
 که حل:

$$X_{\text{جدید}} = X_{\text{قدیم}} + \frac{1}{10} X_{\text{قدیم}} = \frac{11}{10} X_{\text{قدیم}}$$

$$\sigma_{\text{جدید}}^2 = |a|^2 \sigma_{\text{قدیم}}^2 = \left(\frac{11}{10}\right)^2 \times 1000 = 1310$$

مثال: در جدول فراوانی مقابل واریانس داده‌ها کدام است؟

مرکز دسته	۱۲	۱۵	۱۸	۲۱	۲۴
فراوانی	۴	۳	۹	۷	۲

که حل:

ابتدا از داده‌ها ۱۸ واحد کم کرده، سپس آن‌ها را بر ۳ تقسیم می‌کنیم:

مرکز دسته	-۲	-۱	۰	۱	۲
فراوانی	۴	۳	۹	۷	۲

$$\Rightarrow \bar{X} = \frac{1 \times 7 + 2 \times 2 - 1 \times 3 - 2 \times 4}{4 + 3 + 9 + 7 + 2} = 0$$

$$\sigma^2 = \frac{7(1)^2 + 2(2)^2 + 3(-1)^2 + 4(-2)^2}{25} = \frac{7 + 8 + 3 + 16}{25} = \frac{32}{25} = \frac{136}{100} = 1/36$$

مثال: اگر واریانس داده‌های X_1, \dots, X_n برابر صفر باشد میانگین داده‌های $X_1 - 2X_n, X_2 - 2X_n, \dots, X_n - 2X_n$ چقدر است؟
 که حل:

اگر یکی از پارامترهای پراکندگی برابر صفر باشد. یعنی همه‌ی داده‌ها برابرند. پس: $X_1 = \dots = X_n$
 پس داده‌ها عبارتند از: $X_n - 2X_n, X_n - 2X_n, \dots, X_n - 2X_n$ همان $-X_n$ است.

مثال: اگر واریانس X_1, X_2, X_3, X_4, X_5 برابر صفر باشد، واریانس داده‌های $X_1, X_2 + 1, X_3 + 2, X_4 + 3, X_5 + 4$ چقدر است؟
 که حل:

$$X_1 = X_2 = \dots = X_5$$

$$\Rightarrow X_1, X_1 + 1, X_1 + 2, X_1 + 3, X_1 + 4 \rightarrow \bar{x} = X_1 + 2 \rightarrow \sigma^2 = \frac{2^2 + 1^2 + 1^2 + 2^2}{5} = 2$$

مثال: واریانس داده‌های جدول زیر را به دست آورید.

X_i	۳۲	۴۰	۵۶	۷۲
f_i	۲	۳	۱	۲

x_i	۴	۵	۷	۹
f_i	۲	۳	۱	۲

که حل: ابتدا داده‌ها را بر ۸ تقسیم می‌کنیم:

$$\rightarrow \bar{x} = \frac{4 \times 2 + 5 \times 3 + 7 \times 1 + 9 \times 2}{2 + 3 + 1 + 2} = \frac{48}{8} \rightarrow \sigma^2 = \frac{2(2)^2 + 3(1)^2 + 1(1) + 2(3)^2}{2 + 3 + 1 + 2} = \frac{30}{8} = \frac{15}{4} \rightarrow \sigma_{\text{قدیم}}^2 = \left(8^2 \times \frac{15}{4}\right) = 240$$

(۳) در حالت کلی واریانس داده‌هایی که دنباله‌ی حسابی با قدر نسبت d تشکیل می‌دهند برابر است با:

$$\sigma^2 = d^2 \frac{n^3 - 1}{12}$$
 قدر نسبت

واریانس داده‌های $n, 2, \dots, 1$ برابر است با:

$$\sigma^2 = \frac{n^3 - 1}{12}$$

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - (\bar{x})^2 = \frac{1^2 + \dots + n^2}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}$$

مثال: واریانس داده‌های آماری ۴۰۷, ۴۰۵, ۴۰۳, ۴۰۱ را به دست آورید.

کحل: ابتدا از هر داده ۴۰۰ واحد کم می‌کنیم و می‌دانیم واریانس آن‌ها تغییر نمی‌کند. حال داده‌های جدید تصاعد حسابی با قدر نسبت ۲ می‌سازند:

$$\sigma^2 = 2^2 \times \frac{4^2 - 1}{12} = 5$$

(۳) انحراف معیار:

از جذر مثبت واریانس، انحراف معیار به دست می‌آید که تفاوت آن با واریانس در این است که انحراف معیار از جنس داده‌ها می‌باشد (هم‌واحد با داده‌هاست) ولی واریانس از جنس مربع داده‌ها می‌باشد. مثلاً اگر داده‌های ما طول‌هایی بر حسب سانتی‌متر باشند، انحراف معیار دارای واحد سانتی‌متر و واریانس سانتی‌متر مربع می‌باشد. برای داده‌های x_1, \dots, x_n انحراف معیار از دستور زیر به دست می‌آید.

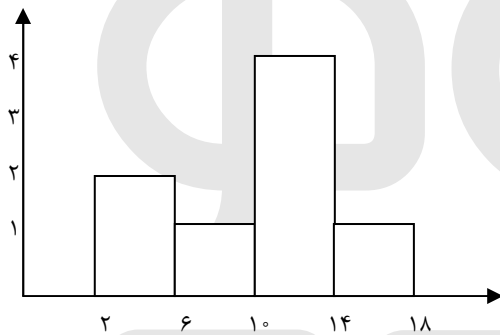
$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

در جدول داده‌های دسته‌بندی نشده:

$$\sigma = \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_n(x_n - \bar{x})^2}{f_1 + f_2 + \dots + f_n}$$

و در جدول داده‌های دسته‌بندی شده از مرکز دسته دسته استفاده می‌کنیم.

مثال: با توجه به نمودار زیر انحراف معیار داده‌ها را به دست آورید.



کحل:

$$\bar{x} = \frac{4 \times 2 + 1 \times 6 + 4 \times 10 + 1 \times 14}{2 + 1 + 4 + 1} = \frac{80}{8} = 10$$

$$\sigma = \sqrt{\frac{2 \times 4^2 + 1 \times 6^2 + 4 \times 10^2 + 1 \times 14^2}{8} - (10)^2} = \sqrt{16} = 4$$

مثال: مجموع مربعات ده داده برابر ۲۲۵۰ و میانگین این ده داده برابر $10\sqrt{2}$ می‌باشد. انحراف معیار این داده‌ها کدام است؟

کحل:

$$\sigma = \sqrt{\frac{2250}{10} - (10\sqrt{2})^2} = \sqrt{225 - 200} = 5$$

مثال: پانزده داده آماری با واریانس ۱۲ و ده داده آماری دیگر با واریانس ۷/۶ را با هم ترکیب می‌کنیم. اگر میانگین هر دو گروه یکسان باشد، انحراف معیار ۲۵ داده حاصل کدام است؟

کحل:

$$\begin{cases} 12 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{15} - \bar{x})^2}{15} \Rightarrow (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{15} - \bar{x})^2 = 180 \\ 7/6 = \frac{(y_1 - \bar{x})^2 + (y_2 - \bar{x})^2 + \dots + (y_{10} - \bar{x})^2}{10} \Rightarrow (y_1 - \bar{x})^2 + (y_2 - \bar{x})^2 + \dots + (y_{10} - \bar{x})^2 = 76 \end{cases}$$

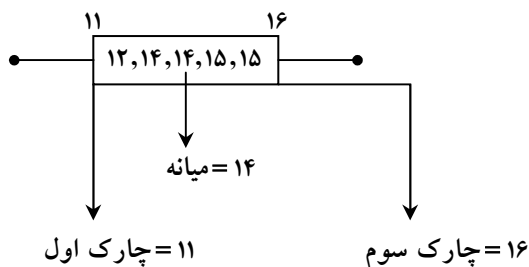
$$\Rightarrow \sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{15} - \bar{x})^2 + (y_1 - \bar{x})^2 + (y_2 - \bar{x})^2 + \dots + (y_{10} - \bar{x})^2}{25}}$$

$$\Rightarrow \sigma = \sqrt{\frac{180 + 76}{25}} = \sqrt{\frac{256}{25}} = \frac{16}{5} = 3.2$$

مثال: اگر داده‌های آماری ۱۱، ۱۵، ۱۷، ۱۶، ۱۴، ۹، ۱۱، ۱۲، ۱۵، ۱۸، ۱۴ را با نمودار جعبه‌ای نشان دهیم، انحراف معیار داده‌های داخل جعبه کدام است؟

کحل:

۹، ۱۱، ۱۱، ۱۲، ۱۴، ۱۴، ۱۵، ۱۵، ۱۶، ۱۷، ۱۸



$$\Rightarrow \bar{x} = \frac{12 + 14 + 14 + 15 + 15}{5} = 14$$

$$\sigma = \sqrt{\frac{2^2 + 1^2 + 1^2}{4}} = \sqrt{\frac{3}{2}}$$

مثال: انحراف معیار ۲۶ داده‌ی آماری برابر ۲ می‌باشد. اگر یکی از داده‌ها که با میانگین برابر است از بین آنان حذف شود، واریانس ۲۵ داده‌ی دیگر چقدر می‌شود؟

کحل:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{26} - \bar{x})^2}{26}} = 2 \Rightarrow (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{26} - \bar{x})^2 = 4 \times 26 = 104$$

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{26} - \bar{x})^2 + 0^2 = 104 \Rightarrow \sigma^2 = \frac{104}{25} = 4/16$$

مثال: کدام مقدار را نمی‌توان از روی نمودار میله‌ای فراوانی دسته‌بندی شده به صورت قطعی به دست آورد؟

- (۱) میانگین دسته‌ها (۲) انحراف معیار دسته‌ها (۳) فراوانی نسبی دسته‌ها (۴) دامنه‌ی تغییرات

کحل:

میانگین و انحراف معیار و فراوانی از روی خود داده‌ها و یا از روی نشان دسته‌ها قابل حصول است، اما دامنه‌ی تغییرات دقیقاً با خود داده‌ها سر و کار دارد، لذا از نشان دسته نمی‌توان پی به دامنه برد.

مثال: اگر انحراف معیار داده‌های $x_1, x_2, \dots, x_n, 16$ برابر صفر باشد، میانگین داده‌ها کدام است؟

کحل:

چون یکی از شاخص‌های پراکندگی صفر است پس:

$$x_1 = x_2 = \dots = x_n = 16$$

پس میانگین نیز برابر است با یکی از داده‌ها:

$$\bar{x} = 16$$

مثال: واریانس ۱۱ داده‌ی آماری صفر است. اگر داده‌های ۲۴، ۱۶ و ۲۶ به آن‌ها اضافه شود، میانگین داده‌ها تغییر نمی‌کند.

انحراف معیار ۱۴ داده‌ی حاصل کدام است؟

کحل: اگر یکی از شاخص‌های پراکندگی صفر باشد، تمام داده‌ها با هم برابرند.

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_{11}}{11} = \frac{x_1 + x_2 + \dots + x_{11} + 24 + 16 + 26}{14} \Rightarrow 3(x_1 + x_2 + \dots + x_{11}) = 11(66) \Rightarrow x_1 + x_2 + \dots + x_{11} = 11^2 \times 2$$

$$\Rightarrow \bar{X} = \frac{11^2 \times 2}{11} = 22$$

در واقع از ابتدا می‌توان گفت اگر با اضافه شدن تعدادی داده میانگین تغییر نکند، باید میانگین داده‌های جدید با میانگین داده‌های قدیم برابر باشد.

$$\bar{X} = \frac{m\bar{X} + n\bar{Y}}{m+n} \Rightarrow \bar{X} = \bar{Y}$$

پس از ابتدا می‌توان گفت که میانگین ۱۱ داده‌ی قبلی برابر ۲۲ است. پس هر ۱۱ داده ۲۲ اند.

$$\sigma = \sqrt{\frac{(22-22)^2 + (26-22)^2 + (24-22)^2 + (16-22)^2}{14}} = \sqrt{\frac{0+16+4+36}{14}} = \sqrt{\frac{56}{14}} = \sqrt{4} = 2$$

نکات:

(۱) اگر انحراف معیار داده‌های x_1, \dots, x_n ، σ باشد، انحراف معیار داده‌های $ax_1 + b, \dots, ax_n + b$ برابر است با:

$$\sigma_{\text{جدید}} = |a|\sigma \Rightarrow \sigma^2_{\text{جدید}} = a^2\sigma^2$$

مثال: اگر انحراف معیار داده‌های ۱، ۳، ۵ و ۷، σ باشد، انحراف معیار داده‌های ۳، ۷، ۱۱ و ۱۵ کدام است؟

کحل: چون داده‌ها در ۲ ضرب شده و با ۱ جمع شده‌اند، انحراف معیار در ۲ ضرب می‌شود.

مثال: اگر انحراف معیار x_1, \dots, x_n برابر ۴ باشد، واریانس داده‌های $3 - 2x_1 + 3, \dots, -2x_n + 3$ کدام است؟

کحل:

$$\sigma^2 = (-2)^2 \times 16 = 64$$

مثال: اگر داده‌های آماری مفروضی را در اختیار داشته باشیم و \bar{X} میانگین این داده‌ها باشد با توجه به جدول زیر، انحراف معیار داده‌ها را بیابید.

f_i	۳	۱	۲	۵	۱۱	۵
$x_i - \bar{X}$	۵	۲	۱	-۲	t	-۴

$$3(5) + 1(2) + 2(1) + 5(-2) + 11(t) + 5(-4) = 0 \rightarrow t = 1$$

$$\sqrt{\frac{3 \times 5^2 + 1 \times 2^2 + 2 \times 1^2 + 5 \times (-2)^2 + 11(1)^2 + 5(-4)^2}{-4}} = \sqrt{\frac{64}{9}} = \frac{8}{3}$$

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

↓
قدر نسبت

(۲) در حالت کلی انحراف معیار داده‌هایی که تصاعد حسابی با قدر نسبت d برابر است با:

انحراف معیار داده‌های $1, 2, \dots, n$ برابر است با:

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

مثال: انحراف معیار ۴۰۵، ۴۰۴، ۴۰۳، ۴۰۲، ۴۰۱ را به دست بیاورید.

کحل: ابتدا از تمام داده‌ها ۴۰۰ واحد کم می‌کنیم، بدون آن‌که انحراف معیار تغییر کند انحراف معیار جدید برابر است با:

$$\sigma = \sqrt{\frac{5^2 - 1}{12}} = \sqrt{2}$$

۳) اگر در داده‌های آماری X_1, \dots, X_n میانگین \bar{X} و انحراف معیار σ باشد، در داده‌های $Y_i = \frac{X_i - \bar{X}}{\sigma}$ میانگین صفر و واریانس ۱ است.

$$\sigma^2 \text{ جدید} = a^2 \sigma^2 \text{ قدیم} = \left(\frac{1}{\sigma}\right)^2 \sigma^2 = 1$$

این کار را استانداردسازی داده‌ها می‌نامند و داده‌های جدید، داده‌های استاندارد نام دارند.

۱۴) ضریب تغییرات:

اگر بخواهیم معیاری را معرفی کنیم که در آن پراکندگی به نسبت میانگین (که با بزرگی داده‌ها در ارتباط است) سنجیده شود، از معیار جدیدی به نام ضریب تغییرات استفاده می‌کنیم.

$$CV = \frac{\sigma}{\bar{X}}$$

ضریب تغییرات بدون واحد است که آن را معمولاً به صورت درصد بیان می‌کنند.

ضریب تغییرات از آن جهت حائز اهمیت است که امکان مقایسه داده‌های غیر هم‌واحد را فراهم می‌سازد.

مثال: ضریب تغییرات داده‌های ۸، ۱۰، ۱۲، ۱۴ کدام است؟

کحل:

$$\left. \begin{aligned} \bar{X} &= \frac{14+12+10+8}{4} = \frac{44}{4} = 11 \\ \rightarrow \sigma &= \sqrt{\frac{3^2+1^2+1^2+3^2}{4}} = \sqrt{5} \end{aligned} \right\} \Rightarrow C.V = \frac{\sqrt{5}}{11}$$

مثال: اگر X متغیر کمی باشد، از اطلاعات جدول زیر، ضریب تغییرات این داده‌ها کدام است؟

$x_i - 12$	-3	-2	-1	0	1	2
f_i	1	3	1	3	6	2

کحل: چون $0 = 4 + 6 - 1 - 6 - 3$ است، لذا میانگین داده‌های فوق، ۱۲ است.

$$\sqrt{\frac{9+3 \times 4+1 \times 1+6 \times 1+2 \times 4}{6}} = \sqrt{6} \Rightarrow C.V = \frac{\sigma}{\bar{X}} = \frac{\sqrt{6}}{12}$$

مثال: ضریب تغییرات داده‌ها در جدول فراوانی مقابل، کدام است؟

x_i	8	9	10	11	12
f_i	3	2	12	6	1

کحل: برای سهولت در محاسبات و جلوگیری از بزرگ شدن اعداد ابتدا از تمام داده‌ها ۱۰ واحد کم می‌کنیم، سپس میانگین و انحراف معیار را محاسبه می‌کنیم.

$$\bar{X} = \frac{1 \times 6 + 2 \times 1 + (-1) \times 2 + (-2) \times 3}{3 + 2 + 12 + 6 + 1} = 0$$

$$\sigma = \sqrt{\frac{6 \times 1^2 + 1 \times 2^2 + 2 \times 1^2 + 3 \times 2^2}{24}} = 1$$

$$\bar{X}_{\text{قدیم}} = \bar{X}_{\text{جدید}} + 10 = 0 + 10 = 10$$

چون انحراف معیار در اثر جمع و تفریق تغییر نمی‌کند، پس:

$$\sigma_{\text{قدیم}} = \sigma_{\text{جدید}} \Rightarrow C.V = \frac{\sigma}{\bar{X}} = \frac{1}{10}$$

مثال: در ۵۰ داده‌ی آماری مجموع تمام داده‌ها ۱۰۰ و مجموع مربعات داده‌ها ۲۷۲ می‌باشد. ضریب تغییرات کدام است؟

کحل:

$$\sigma = \sqrt{\frac{272}{50} - \left(\frac{100}{50}\right)^2} = \sqrt{\frac{72}{50}} = \sqrt{\frac{36}{25}} = \frac{6}{5} \rightarrow CV = \frac{\sigma}{\bar{X}} = \frac{\frac{6}{5}}{2} = \frac{3}{5}$$

مثال: مجموع ۸ داده‌ی آماری برابر ۴۸ و ضریب تغییرات آن‌ها ۰/۵ می‌باشد. مجموع مربعات این داده‌ها کدام است؟

کحل:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{48}{8} = 6$$

$$C.V = \frac{\sigma}{\bar{X}} = \frac{\sigma}{6} = 0.5 \Rightarrow \sigma = 3$$

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - (\bar{X})^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{8} - (6)^2 = 9 \Rightarrow x_1^2 + x_2^2 + \dots + x_n^2 = 360$$

نکته: اگر همه‌ی داده‌های آماری در عدد ثابت مثبتی ضرب شوند، ضریب تغییرات ثابت می‌ماند.

اما اگر همه‌ی داده‌های آماری با عدد ثابت مثبتی جمع شوند، ضریب تغییرات جدید از ضریب تغییرات اولیه کوچک‌تر است. و اگر همه‌ی داده‌های آماری با عدد ثابت منفی‌ای جمع شوند، ضریب تغییرات جدید از ضریب تغییرات اولیه بزرگتر است. (به شرط آن‌که علامت ضریب تغییرات تغییر نکند.)

مثال: اگر ضریب تغییرات داده‌هایی برابر ۴ باشد و داده‌ها را دو برابر کنیم، ضریب تغییرات جدید کدام است؟

کحل:

$$CV = \frac{\sigma}{\bar{X}}$$

میانگین و انحراف معیار دو برابر می‌شوند، پس ضریب تغییرات، تغییر نمی‌کند.

مثال: در ۶۰ داده‌ی آماری میانگین ۳ و انحراف معیار ۱/۲ محاسبه شده است. اگر به تمام داده‌ها ۹ واحد اضافه شود، ضریب تغییرات داده‌های جدید کدام است؟

کحل:

$$\left. \begin{array}{l} \text{میانگین جدید} = 3 + 9 = 12 \\ \text{انحراف معیار جدید} = 1/2 \end{array} \right\} \rightarrow C.V = \frac{1/2}{12} = \frac{1}{10}$$

مثال: در n داده‌ی آماری $x_i: i=1, 2, 3, \dots, n$ ضریب تغییرات برابر ۱/۲ محاسبه شده است. میانگین داده‌های مفروض را به هر یک از آنان اضافه می‌کنیم. ضریب تغییرات در داده‌های جدید کدام است؟

کحل:

$$C.V = 1/2 = \frac{\sigma}{\bar{X}}$$

اگر به هر یک از داده‌ها \bar{X} را بیافزاییم، σ تغییر نمی‌کند، اما میانگین جدید $\bar{X} + \bar{X} = 2\bar{X}$ خواهد بود.

$$C.V_{\text{جدید}} = \frac{\sigma}{2\bar{X}} = \frac{\bar{X}}{2} = \frac{C.V_{\text{قدیم}}}{2} = 0.6$$

مثال: اگر ۲۰ داده‌ی آماری را دو برابر کرده و سپس ۷ واحد از هر کدام کم کنیم، ضریب تغییرات داده‌های جدید، ۱/۵ برابر ضریب تغییرات داده‌های قبلی می‌شود. مجموع داده‌های قبلی کدام است؟
 که حل:

$$\left. \begin{aligned} \bar{x}_{\text{جدید}} &= 2\bar{x}_{\text{قدیم}} - 7 \\ \sigma_{\text{جدید}} &= 2\sigma_{\text{قدیم}} \end{aligned} \right\} \rightarrow CV_{\text{جدید}} = \frac{2\sigma}{2\bar{x} - 7} = \frac{2}{2} \frac{\sigma}{\bar{x}} \rightarrow 4\sigma\bar{x} = 6\sigma\bar{x} - 21\sigma \rightarrow 2\sigma\bar{x} = 21\sigma \rightarrow \bar{x} = \frac{21}{2}$$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_{20}}{20} = \frac{21}{2} \rightarrow x_1 + x_2 + \dots + x_{20} = 210$$

• دو نکته‌ی کلی در مورد پارامترهای پراکندگی:

(۱) شرط لازم و کافی برای آن‌که در یک نمونه‌ی آماری تمام اعضا با هم برابر باشند آن است که تمام پارامترهای پراکندگی برابر صفر باشد.

$$R = \sigma^2 = \sigma = CV = 0$$

که در این صورت میانگین داده‌ها با هر یک داده‌ها برابر است.

(۲) دقت پارامترهای واریانس، انحراف معیار و دامنه‌ی تغییرات به صورت زیر است:

(۱) واریانس (۲) انحراف معیار (۳) دامنه‌ی تغییرات

هرچه دامنه‌ی تغییرات بزرگ‌تر باشد، باعث بزرگ‌تر شدن انحراف معیار و واریانس می‌شود و برعکس.



مؤسسه آموزشی فرهنگی