



آمار و مدل سازی

آمار:

آمار علم جمع‌آوری اطلاعات، سازمان‌دهی آن‌ها و تفصیل و تجزیه و تحلیل 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. هرچقدر مفاهیم ریاضی به کار برده شده ساده‌تر و ابتدایی‌تر و نتیجه‌ی کار به پدیده مورد نظر نزدیک‌تر باشد، مدل‌سازی با ارزش‌تر است.

اندازه‌گیری عبارتست 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

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

مشکلات سرشماری:

۲- وقت‌گیر بودن دسترسی به تمام اعضای جامعه

۱- در دسترس نبودن تمام اعضای جامعه

۴- از بین رفتن جامعه در برخی از مطالعات

۳- گران تمام شدن بررسی تمام اعضای جامعه

نمونه زیر مجموعه‌ای از جامعه 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 جامعه 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 گونه ای باشد که:

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

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

بدیهی است نمونه 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 اندازه نمونه** می‌نامند.

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 آنرا در حجم جامعه‌ی آماری ضرب کرده و اولین عدد پاسخ بزرگتر از یا مساوی 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 آمده است. چه شماره‌هایی متناظر با این اعداد به ترتیب باید انتخاب شوند؟
 کحل:

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

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

روش‌های جمع‌آوری داده‌ها:

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 سوال، 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 تک کلمه‌ای یا تک رقمی

- ۶- در سؤالاتی که ممکن است پاسخ دهنده نخواهد به آن جواب دقیق بدهد از سؤالات با پاسخ از پیش آماده شده استفاده شود مانند تعیین محدوده
- ۷- عدم استفاده از سؤالات هدایت کننده یعنی سؤالاتی که جواب را به پاسخ دهنده القا می کند.
- ۸- حتی الامکان از سؤالات با پاسخ چند گزینه ای استفاده شود.
- ۹- برای پرسش عقیده راجع به موضوع یا محصول، پاسخ ها به صورت گزینه های کمی یا کیفی سطح بندی شده انتخاب شود. مانند بسیار خوب، خوب، متوسط، ضعیف، بسیار ضعیف.
- ۱۰- ضمیمه کردن دستورالعمل پاسخگویی (واضح و کامل)
- ۱۱- تشکر در پایان

فصل سوم: متغیرهای تصادفی

متغیر تصادفی: مشخصه یا صفت ویژه ای از افراد جامعه که روی آن مطالعه انجام می دهیم. از دیدگاهی دیگر متغیرها بر دو نوع اند:

(۱) **کمی:** متغیری که اندازه و مقدار دارد و دارای واحد اندازه گیری می باشد، مانند: قد، وزن، عمر

که خود بر دو نوع است **پیوسته:** قابل اندازه گیری مانند: قد، وزن

گسسته: قابل شمارش مانند: تعداد تصادفات، تعداد کالای خراب، تعداد روزهای بارانی

(۲) **کیفی:** متغیری که قابل اندازه گیری نیست و عموماً با مقایسه بررسی می شود، مانند: هوش، استعداد، درس خوان بودن و...

که خود بر دو نوع است **ترتیبی:** دارای ترتیب ذاتی هستند مانند حروف الفبای فارسی یا مراحل رشد

اسمی: دارای ترتیب ذاتی نیستند مانند رنگ چشم افراد، گروه های خونی

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

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

(۲) **صفت متغیر:** صفتی که از هر عضو به عضو دیگر تغییر می کند.

مثلاً در جامعه آماری ایرانی ها متغیر تصادفی وزن افراد صفت متغیر و ایرانی بودن صفت ثابت است.

فصل چهارم: دسته بندی داده ها و جدول فراوانی

فراوانی یک داده:

در داده های آماری دسته بندی نشده به تعداد دفعاتی که هر داده تکرار می شود، فراوانی مطلق آن داده گفته می شود و آن را با f_i نمایش می دهند و نسبت فراوانی مطلق هر داده به تعداد داده ها، فراوانی نسبی نامیده می شود.

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

$F_i \times 100 =$ درصد فراوانی نسبی

وقتی فراوانی داده های آماری که تعداد و پراکندگی شان زیاد است مورد مطالعه قرار می گیرند استفاده از جدول توزیع فراوانی بسیار دشوار است، لذا داده ها را دسته بندی می کنیم.

۴- 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 - دامنه داده‌ها:

تفاضل کمترین از بیشترین داده در یک جامعه آماری، دامنه جامعه 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 دسته 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 بدست 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 رابطه زیر به دست می‌آید.

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

که اگر عدد فوق اعشاری بود حتماً از رابطه‌ی $+1$ استفاده می‌شود.

$$\left[\frac{\text{دامنه‌ی تغییرات}}{\text{طول دسته‌ها}} \right] + 1$$

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؟

حل:

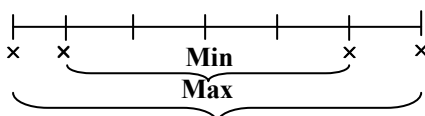
$$R = x_{\max} - x_{\min} = 37 \rightarrow 71 - x_{\min} = 37 \rightarrow x_{\min} = 34$$

پس اولین داده ۳۴ است. حال باید به اندازه $n = \left[\frac{37}{5} \right] + 1 = 8$ 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.

حل:



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

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:

$$F_i = \frac{f_i}{\sum_{i=1}^n f_i} \quad \text{فراوانی نسبی دسته } i \text{ام در داده‌های دسته‌بندی شده برابر است با:}$$

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: $f_{c_i} = f_1 + f_2 + \dots + f_i$

$$F_{c_i} = \frac{f_{c_i}}{\sum_{i=1}^n f_i} \quad \text{و فراوانی تجمعی نسبی دسته } i \text{ام در داده‌های دسته‌بندی شده برابر است با:}$$

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

نکات:

$$f_{i+1} = f_{c_{i+1}} - f_{c_i} \quad (۱)$$

$$\frac{f_1}{\sum f_i} + \frac{f_2}{\sum f_i} + \dots + \frac{f_n}{\sum f_i} = 1 \quad (۲) \quad \text{در هر جدول توزیع فراوانی، مجموع فراوانی‌های نسبی برابر ۱ است.}$$

(۳) فراوانی تجمعی دسته آخر برابر است با مجموع فراوانی‌ها

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

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

x_i	۱	۲	۳	۴	۵	۶
f_i	۴	۶	۸	۷	۳	۲

کحل:

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

☞ 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؟

کحل:

پس ۱۵ درصد داده‌ها بین ۴۰ تا ۴۳ هستند. چون دامنه ۲۱ و تعداد دسته‌ها ۷ است پس طول دسته‌ها برابر ۳ است. لذا دسته

$$\text{چهارم همان دسته } ۴۳ - ۴۰ \text{ خواهد بود. پس فراوانی دسته وسط برابر است با: } ۱۵ \times \frac{۱۵}{۱۰۰} = ۲$$

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

$$x_i^* = \frac{\text{کران بالا} + \text{کران پایین}}{۲}$$

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

مثال: در جدول توزیع فراوانی تعداد طبقات ۸ و طول هر دسته ۳ می باشد، در صورتی که نشان دسته‌ی طبقه سوم ۱۸ باشد، نشان دسته طبقه اول و آخر به ترتیب کدامند؟

کحل:

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

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

مثال: در جدول فراوانی داده‌های دسته‌بندی شده، اگر درصد فراوانی نسبی دسته وسط ۲۴ باشد، فراوانی مطلق دسته چهارم کدامست؟

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

کحل:

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

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

شافص‌های آماری و تحلیل داده‌ها:

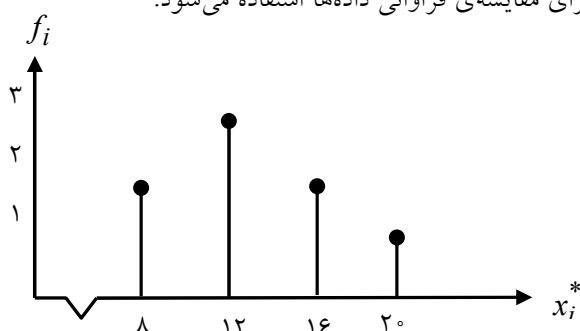
آنچه تا کنون بر روی داده‌ها انجام دادیم، به نوعی سازمان‌دهی آنها برای بهره‌برداری ساده‌تر بود که عمدتاً بر مبنای خلاصه کردن داده‌ها بنا شده بود، اما با تلفیق و به هم آمیختن داده‌ها، سرانجام مقادیر عددی به دست خواهد آمد که به نحو شایسته‌ای می‌تواند جامعه را تحلیل کند. این مقادیر عددی را شاخص‌های عددی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

فصل پنجم: نمودار و تحلیل داده‌ها (شاخص‌های هندسی)

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.

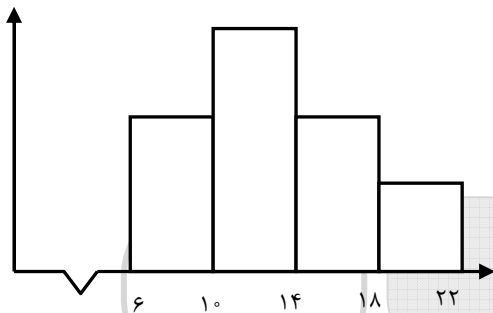


قرار داد: در مواردی که مقیاس بزرگ است و ممکن است تمام نمودار در صفحه قرار نگیرد، نزدیک ترین داده به مبدأ را با علامت به مبدأ نزدیک می کنیم.

۲) نمودار مستطیلی (هیستوگرام):

هنگامی که جدول توزیع فراوانی در دست است از این نمودار استفاده می کنیم. در این نمودار مستطیل هایی رسم می کنیم که یک ضلع آن منطبق بر دسته ها و ضلع دیگر، فراوانی (مطلق یا نسبی) دسته متناظر باشد. نمودار مستطیلی برای داده های کمی پیوسته مناسب است.

اگر طول دسته ها با هم برابر باشند، فراوانی دسته ای که مساحت آن بزرگ تر است، بیشتر خواهد بود. در واقع فراوانی متناسب با مساحت مستطیل هاست.



نمودار مستطیلی = هیستوگرام، بافت نگار

مثال: جدول زیر جدول فراوانی تجمعی داده های آماری دسته بندی شده است اگر فراوانی نسبی دسته سوم برابر ۰/۴۵ باشد آنگاه مساحت مستطیل مربوط به دسته چهارم در نمودار مستطیلی این داده های آماری کدام است؟

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

حل:

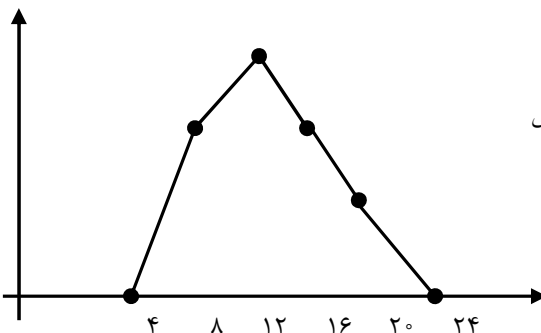
$$\text{طول دسته} = x_2^* - x_1^* = 20 - 17 = 3$$

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

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

۳) چندبر فراوانی:

در این نمودار نقاطی از صفحه را مشخص می کنیم که طول آنها برابر مرکز دسته ها و عرض آنها، فراوانی (مطلق یا نسبی) همان دسته است. با به هم وصل کردن این نقاط چندبر فراوانی به دست می آید. در چندبر فراوانی، دو دسته با فراوانی صفر به ابتدا و انتهای داده ها اضافه می کنیم که به این وسیله مساحت چندبر فراوانی با مساحت نمودار مستطیلی متناظرش برابر می شود.



مساحت چندبر = مساحت نمودار مستطیلی

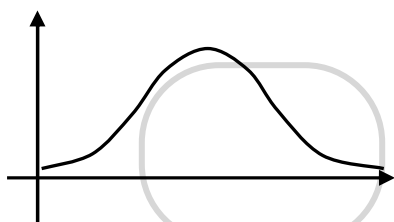
نمودار چندبر فراوانی برای داده های کمی پیوسته مناسب است.

نمودار چند بر فراوانی را می توان با داشتن فراوانی نسبی نیز رسم کرد که آن را چندبر فراوانی نسبی می گوئیم. در این صورت اطلاعات منسجم تری در اختیار ما قرار می گیرد، چون می توان فراوانی را با کل جامعه مقایسه کرد.

اگر بخواهیم تغییرات متغیر را در فاصله ی بین دو دسته یا در خود دسته بهتر نشان دهیم از نمودار چندبر فراوانی استفاده می کنیم.

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 یک دیگر مقایسه کنیم، مشکل بتوانیم با انطباق مستطیل‌ها بر یکدیگر این مقایسه را انجام دهیم. در چنین مواردی رسم چندبر فراوانی انتخاب مناسب‌تری است زیرا مقایسه توزیع فراوانی را آسان‌تر 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.

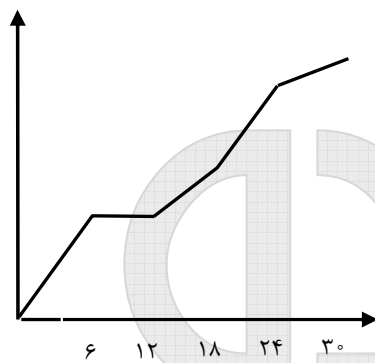
منحنی نرمال: منحنی ایست که در اکثر پدیده‌های طبیعی ظاهر می‌شود و شکلی شبیه زنگوله دارد و سطح زیر منحنی آن برابر فراوانی مطلق داده‌هاست. (اگر منحنی را با فراوانی نسبی رسم کنیم سطح زیر منحنی برابر ۱ است)



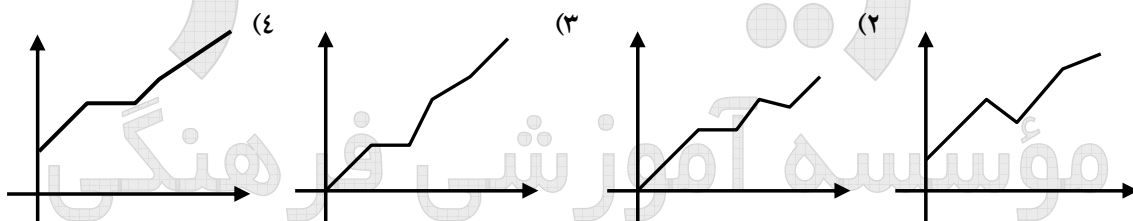
این منحنی دارای ماکسیمم است و فراوانی در دو طرف این ماکسیمم به طور یکنواخت به سمت صفر میل می‌کند. این منحنی متقارن است.

۴) نمودار تجمعی:

هنگامی که جدول توزیع فراوانی در دست است از این نمودار استفاده می‌کنیم. در این نمودار حدود دسته‌ها منطبق بر محور Xهاست 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 طبقه بندی شده دو نقطه‌ی $(21, 42)$ و $(24, 51)$ دو نقطه‌ی متوالی از نمودار

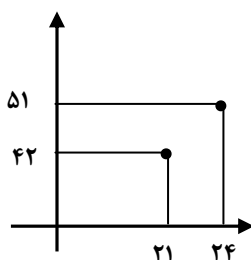
فراوانی تجمعی باشند، کدام نقطه‌ی زیر روی چندبر فراوانی قرار دارد؟

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

کحل:

دو نقطه داده شده مشخص می‌کند فراوانی مطلق دسته $24 - 21$ برابر ۹ است، پس نقطه

$(22/5, 9)$ روی چند بر فراوانی قرار می‌گیرد.



۵) نمودار دایره‌ای:

برای رسم نمودار دایره‌ای، مساحت دایره را به وسیله قطاع‌هایی به نسبت فراوانی‌ها تقسیم می‌کنیم. اگر فراوانی معلوم باشد، زاویه هر قطاع از رابطه $\theta_i = \frac{f_i}{N} \times 360^\circ$ 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 شرکت کنندگان در یک گروه سنی با زاویه قطاع ۶۳ درجه نشان داده می‌شود. تعداد اعضای این گروه کدام است؟
 کحل:

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

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

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

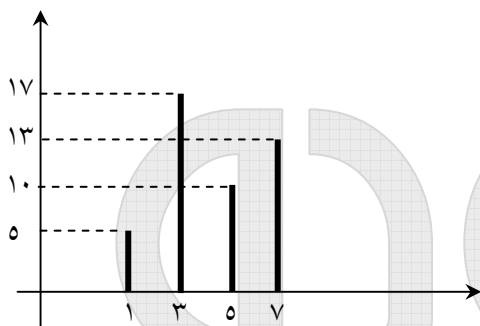
خونی O چند درجه است؟

کحل:

$$\alpha = 17/5$$

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

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



کحل:

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

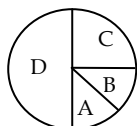
$$\theta_3 = \frac{17}{45} \times 360 = 136^\circ$$

$$\theta_5 = \frac{10}{45} \times 360 = 80^\circ$$

$$\theta_7 = \frac{13}{45} \times 360 = 104^\circ$$



مثال: در نمودار زیر تعداد فراوانی افرادی که در دسته‌های A, B, C, D قرار دارند، به ترتیب ۲، ۳ و ۶ برابر تعداد افرادی است که در دسته‌ای A قرار دارند. کدام است؟



کحل:

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

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 زاویه نمودار تغییر نمی‌کند.

مثال: داده‌های آماری در ۹ طبقه دسته بندی شده اند. فراوانی تجمعی نسبی در دسته‌ی چهارم و پنجم به ترتیب $۰/۲۸$ و $۰/۴۰$ است. در نمودار دایره ای زاویه‌ی مربوط به دسته‌ی پنجم چند درجه است؟
 کحل:

$$\text{فراوانی نسبی دسته پنجم} = ۰/۴ - ۰/۲۸ = ۰/۱۲ \rightarrow \theta = ۰/۱۲ \times ۳۶۰ = ۴۳/۲^\circ$$

۶) نمودار ساقه و برگ:

برای آنکه اعداد را به گونه ای مرتب کنیم که چیزی شبیه نمودار میله ای 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 آنرا می‌نویسیم. برای فهم آنکه 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: ۲۳ ۹

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

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

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

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

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» درست اعداد $۱۴/۵$ بوده‌اند؟
 کحل:

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

پس $\frac{۴}{۵}$ یا ۸۰٪ داده‌ها عدد $۱۴/۵$ بوده است.

۲- شاخص‌های 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 نشان دهنده مرکز تجمع داده‌ها می‌باشند و برای منظور دوم از شاخص‌های پراکندگی استفاده می‌کنیم که نشان دهنده میزان تجمع داده‌ها حول مرکز است.

فصل ششم: شاخص‌های مرکزی

شاخص‌های مرکزی عبارتند از: میانگین، میانه، مد

۱) میانگین:

میانگین n داده‌ی آماری x_1, x_2, \dots, x_n برابر است با:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \Rightarrow \sum_{i=1}^n x_i = n\bar{X}$$

مثال: اگر میانگین داده‌های x_1, x_2, x_3, x_4 برابر \bar{X} باشد، میانگین داده‌های $2x_1 + x_2, 2x_2 + x_3, 2x_3 + x_4, 2x_4 + x_1$

کدام است؟

حل:

$$\bar{X} = \frac{2x_1 + x_2 + 2x_2 + x_3 + 2x_3 + x_4 + 2x_4 + x_1}{4} = \frac{3(x_1 + x_2 + x_3 + x_4)}{4} = 3\bar{X}_{\text{قدیم}}$$

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 تک‌تک داده‌ها، میانگین آن را قرار دهیم، مجموع و میانگین داده‌ها تغییر نخواهد کرد.

$$\sum_{i=1}^n x_i = n\bar{X} \quad \sum_{i=1}^n \bar{X} = n\bar{X}$$

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 داده آماری \bar{X}_1 و میانگین m_2 داده آماری \bar{X}_2 و ... و میانگین m_n داده آماری

$$\bar{X} = \frac{\sum_{i=1}^n m_i \bar{X}_i}{\sum_{i=1}^n m_i} \quad \bar{X}_n \text{ باشد، میانگین کل داده‌ها برابر است با:}$$

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

حل:

$$\frac{\sum_{i=1}^{10} x_i}{10} = 5 \rightarrow \sum_{i=1}^{10} x_i = 50 \rightarrow \bar{x} = \frac{50 + 16}{11} = \frac{66}{11} = 6$$

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

حل:

$$\bar{x} = \frac{30 \times 12 + 20 \times 16}{30 + 20} = \frac{360 + 320}{50} = 13.6$$

(۵) اگر میانگین داده‌های x_1, x_2, \dots, x_n باشد، میانگین داده‌های $ax_1 + b, ax_2 + b, \dots, ax_n + b$ برابر است با: $a\bar{x} + b$

(۶) مجموع تفاضل داده‌ها از میانگین برابر صفر است. $\sum_{i=1}^n (x_i - \bar{x}) = 0$

محاسبه میانگین در جدول فراوانی: (میانگین وزن دار)

میانگین n داده‌ی آماری که در آن داده‌ها دارای فراوانی باشند برابر است با

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

مجموع داده‌ها
تعداد داده‌ها

که در این صورت: $\sum f_i (x_i - \bar{x}) = 0$

اگر جدول طبقه‌بندی شده داده‌ها در دسترس باشد، برای محاسبه میانگین به جای داده‌ها در رابطه فوق از نشان دسته بهره می‌بریم.

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

که در این صورت: $\sum f_i (x_i^* - \bar{x}) = 0$

مثال: میانگین داده‌های جدول زیر را بدست آورید.

دسته	۲-۴	۴-۶	۶-۸	۸-۱۰
فراوانی	۲	۴	۳	۱

حل:

$$\bar{x} = \frac{3 \times 2 + 5 \times 4 + 7 \times 3 + 9 \times 1}{2 + 4 + 3 + 1} = \frac{6 + 20 + 21 + 9}{10} = 5.6$$

(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 اگر تعداد داده‌ها زوج باشد، نصف مجموع دو عددی که در وسط قرار گرفته، 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، میانه 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، ۱۹

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

برای n داده فرد، میانه $\frac{n+1}{2}$ مین 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 $\frac{n+1}{2}$ 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	۳	۶	۴
f_{c_i}	۲	۴	۷	۱۲	۱۸	۲۰

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

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

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

چارک‌ها: مؤسسه آموزشی فرهنگی

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 پیدا می‌کنیم.

مثال: چارک‌های اول و 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؟

حل:

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

چون ۱۸ داده داریم، میانه ۹ داده اول داده ۵ ام و ۹ داده دوم، داده ۱۴ ام 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، دسته بندی شده اند. اگر ۸ داده بین چارک اول و سوم به آن اضافه شود و

یک واحد از طول دسته کم کنیم، در دسته بندی جدید تعداد دسته‌ها کدام است؟

حل:

دامنه داده‌ها $36 = 9 \times 4$ است. حال اگر طول دسته‌ها برابر ۳ باشد، تعداد دسته‌ها $12 = \frac{36}{3}$ می‌شود. اضافه شدن تعدادی داده

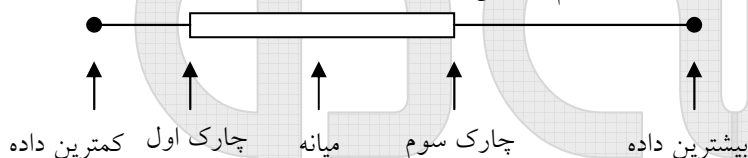
داخل دامنه، تأثیری بر حدود دسته نمی‌گذارد.

نمودار جعبه‌ای:

برای نشان دادن بهتر پراکندگی از این نمودار استفاده می‌کنیم. برای تهیه 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

حل:

۳، ۷، ۱۰، ۱۱، ۱۳، ۲۰، ۲۵



(۳) مد (نما):

مقدار یا مقادیری از متغیر که در آنها فراوانی ماکزیمم مطلق 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؟
 که حل:

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

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

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

مثال: چرا از دامنه‌ی تغییرات به عنوان معیار پراکندگی که مناسب نیست استفاده شود؟

(۲) مقدار آن اغلب بزرگ است

(۱) با ازدیاد داده‌ها فقط می‌تواند کم شود

(۳) مقدار آن از یک نمونه به نمونه دیگری تغییر می‌کند

(۴) در محاسبه آن فقط از دو اندازه استفاده می‌شود

که حل: 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.

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

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

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

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

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

اگر جدول داده‌های دسته‌بندی شده در دسترس باشد از نشان هر دسته به عنوان نماینده آن استفاده می‌کنیم تا مسأله به حالت قبلی تبدیل شود.

	x_1^*	x_2^*	...	$x_{n-1} - x_n$
حدود	$x_1 - x_2$	$x_2 - x_3$...	
فراوانی	f_1	f_2	...	f_n

$$\sigma^2 = \frac{\sum_{i=1}^n f_i(x_i^* - \bar{X})^2}{\sum_{i=1}^n f_i}$$

مثال: واریانس داده‌های جدول زیر کدام است؟

x_i	۲	۳	۴	۵	۶	۷
f_i	۱	۱	۳	۵	۴	۲

حل:

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

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

x_i	۰-۲	۲-۴	۴-۶	۶-۸
f_i	۱	۲	۹	۴

مثال: واریانس داده‌های آماری جدول زیر کدام است؟

حل:

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

$$\rightarrow \sigma^2 = \frac{(1-5)^2 + 2(3-5)^2 + 9(5-5)^2 + 4(7-5)^2}{16} = 1/75$$

مثال: در بین داده‌های آماری زیر در کدام گزینه واریانس بزرگ‌تر است؟

۱۰، ۱۰، ۱۰، ۱۰، ۱۰، ۴

۱۳، ۱۲، ۸، ۷، ۳

۱۰، ۱۰، ۱۶، ۶، ۲

۱۳، ۱۳، ۷، ۷، ۱

حل: تقریباً میانگین همه گزینه‌ها برابر ۱۰ است. پس گزینه 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

نفر اول	۷	۹	۸	۹	۷
نفر دوم	۱۰	۸	۶	۷	۹

است. دقت کاری کدام بیشتر است؟

(۲) نفر دوم

(۱) نفر اول

(۴) نیاز به اطلاعات بیشتر

(۳) یکسان

حل:

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

نکات:

(۱) واریانس را می‌توان از دستور زیر نیز محاسبه کرد:

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

نکته: واریانس را می‌توان در جدول فراوانی طبقه‌بندی نشده از رابطه زیر نیز به دست آورد.

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

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

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i} \right)^2$$

مثال: اگر $\sum x_i^2 = 100$ و $\sum x_i = 20$ و تعداد داده‌ها برابر ۱۰ باشد، واریانس داده‌ها کدامست؟

✓حل:

$$\sigma^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \frac{100}{10} - \left(\frac{20}{10}\right)^2 = 6$$

(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 $a^2 \sigma^2$.

$$\sigma^2 = \frac{\sum ((ax_i + b) - (a\bar{x} + b))^2}{n} = \frac{a^2 \sum (x_i - \bar{x})^2}{n} = a^2 \underbrace{\frac{\sum (x_i - \bar{x})^2}{n}}_{\sigma^2}$$

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

چقدر است؟

✓حل:

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

$$\sigma_{\text{جدید}}^2 = \left| \frac{11}{10} \right|^2 \sigma_{\text{قدیم}}^2 = \left(\frac{11}{10} \right)^2 \times 1000 = 1210$$

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؟

✓حل:

اگر یکی از پارامترهای پراکندگی برابر صفر باشد. یعنی همه داده‌ها برابرند. پس: $x_1 = \dots = x_n$

پس داده‌ها عبارتند از: $x_n - 2x_n, x_n - 2x_n, \dots, x_n - 2x_n$ میانگین همان است. $-x_n$

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؟

✓حل:

$$x_1 = x_2 = \dots = x_5$$

$$\rightarrow x_1, x_1 + 1, x_1 + 2, x_1 + 3, x_1 + 4 \rightarrow \bar{x} = x_1 + 2 \rightarrow \sigma^2 = \frac{2^2 + 1^2 + 1^2 + 2^2 + 2^2}{5} = 2$$

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

✓حل: ابتدا داده‌ها را بر ۸ تقسیم می‌کنیم:

$$\rightarrow \bar{x} = \frac{4 \times 2 + 5 \times 3 + 7 \times 1 + 9 \times 2}{2 + 3 + 1 + 2} = \frac{48}{8} = 6 \rightarrow \sigma^2 = \frac{2(2)^2 + 3(1)^2 + 1(1)^2 + 2(3)^2}{2 + 3 + 1 + 2} = \frac{30}{8} = \frac{15}{4} \rightarrow \sigma_{\text{قدیم}}^2 = \left(8^2 \times \frac{15}{4} \right) = 240$$

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

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

↓
قدر نسبت

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

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

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

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

حل:

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

۱، ۳، ۵، ۷

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

(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.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

در جدول داده‌های دسته‌بندی نشده:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - \bar{X})^2}{n}}$$

و در جدول داده‌های دسته‌بندی شده:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n f_i (x_i^* - \bar{X})^2}{n}}$$

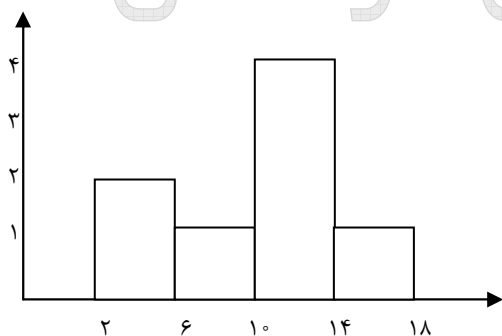
همچنین از رابطه زیر نیز می‌توان استفاده کرد:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - (\bar{X})^2} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$$

$$\text{یا } \sigma = \sqrt{\frac{\sum_{i=1}^n f_i (x_i)^2}{\sum_{i=1}^n f_i} - \left(\frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}\right)^2}$$

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

حل:



$$\bar{X} = \frac{4 \times 2 + 1 \times 6 + 12 \times 10 + 1 \times 14}{2 + 1 + 4 + 1} = \frac{140}{8} = 17.5$$

$$\sigma = \sqrt{\frac{2 \times 4^2 + 1 \times 6^2 + 4 \times 12^2 + 1 \times 14^2}{8} - (17.5)^2} = \sqrt{16} = 4$$

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

حل:

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

مثال: کدام مقدار را نمی توان از روی نمودار میله ای فراوانی دسته بندی شده به صورت قطعی به دست آورد؟

(۱) میانگین دسته ها (۲) انحراف معیار دسته ها (۳) فراوانی نسبی دسته ها (۴) دامنه ی تغییرات

✓حل:

میانگین و انحراف معیار و فراوانی از روی خود داده ها و یا از روی نشان دسته ها قابل حصول است، اما دامنه ی تغییرات دقیقاً با خود داده ها سر و کار دارد، لذا نشان دسته نمی توان پی به دامنه برد.

مثال: اگر انحراف معیار داده های $x_1, x_2, \dots, x_n, 16$ برابر صفر باشد، میانگین داده ها کدامست؟

✓حل:

چون یکی از شاخص های پراکندگی صفر است پس:

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

پس میانگین نیز برابر است با یکی از داده ها:

$$\bar{x} = 16$$

نکات:

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

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

مثال: اگر انحراف معیار داده های ۱، ۳، ۵، ۷، σ باشد، انحراف معیار داده های ۳، ۷، ۱۱، ۱۵ کدامست؟

✓حل:

چون داده ها در ۲ ضرب شده و با ۱ جمع شده اند، انحراف معیار در ۲ ضرب می شود.

مثال: اگر انحراف معیار x_1, \dots, x_n برابر ۴ باشد، واریانس داده های $3 - 2x_1, \dots, 3 - 2x_n$ کدامست؟

✓حل:

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

مثال: اگر داده های آماری مفروضی را در اختیار داشته باشیم و \bar{X} میانگین این داده ها باشد با توجه به جدول مقابل، انحراف

معیار داده ها، تقریباً برابر کدام است؟

✓حل:

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

$$\sum f_i (x_i - \bar{x}) = 0 \rightarrow 3(5) + 1(2) + 2(1) + 5(-2) + 11(t) + 5(-4) = 0 \rightarrow t = 1$$

$$\rightarrow \sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{3 \times 5^2 + 1 \times 2^2 + 2 \times 1^2 + 5 \times (-2)^2 + 11(1)^2 + 5(4)^2}{27}} = \sqrt{\frac{64}{9}} = \frac{8}{3}$$

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

↓
قدر نسبت

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

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

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

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

حل:

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.

(۴) 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؟

حل:

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

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

حل:

$$\sigma = \sqrt{\frac{272}{50} - \left(\frac{100}{50}\right)^2} = \sqrt{\frac{72}{50}} = \sqrt{\frac{36}{25}} = \frac{6}{5} \rightarrow C.V = \frac{\sigma}{\bar{X}} = \frac{\frac{6}{5}}{\frac{100}{50}} = \frac{3}{10}$$

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

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

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

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

حل:

هم میانگین، هم ضریب تغییرات دو برابر می‌شود، پس ضریب تغییرات، تغییر نمی‌کند.

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

✓ حل:

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

مثال: در داده‌های آماری با میانگین \bar{X} و انحراف معیار σ اگر به هریک از داده‌ها، مقدار \bar{X} را اضافه کنیم تا داده‌های جدید حاصل شود، ضریب تغییرات داده‌های جدید چند برابر ضریب تغییرات داده‌های قبلی است؟

✓ حل:

$$\left. \begin{array}{l} \text{میانگین جدید} = \bar{X} + \bar{X} = 2\bar{X} \\ \text{انحراف معیار جدید} = \sigma \end{array} \right\} \rightarrow C.V = \frac{\sigma}{2\bar{X}} = \frac{1}{2} C.V \text{ قدیم}$$

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

✓ حل:

$$\left. \begin{array}{l} \bar{X}_{\text{جدید}} = 2\bar{X}_{\text{قدیم}} - 7 \\ \sigma_{\text{جدید}} = 2\sigma_{\text{قدیم}} \end{array} \right\} \rightarrow C.V_{\text{جدید}} = \frac{2\sigma}{2\bar{X} - 7} = \frac{2}{2} \frac{\sigma}{\bar{X}} \rightarrow 4\sigma\bar{X} = 6\sigma\bar{X} - 21\sigma \rightarrow 2\sigma\bar{X} = 21\sigma \rightarrow \bar{X} = \frac{21}{2}$$

$$\bar{X} = \frac{\sum x_i}{20} = \frac{21}{2} \rightarrow \sum x_i = 210$$

دو نکته کلی در مورد پارامترهای پراکندگی:

(۱) شرط لازم و کافی برای آن که در یک نمونه 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 خود 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