به نام خداوند جان و خرد   کز این برتر اندیشه برنگذرد

آمار و احتمال مهندسی

شهرام منصوری

۲۹۹۰۵۵۵۷

Sh_mansouri@sbu.ac.ir
mansoury956@gmail.com

فصل اول: آمار توصیفی

تعریف جامعه آماری:

جامعه آماری به تمام افرادی که مورد مطالعه و بررسی ما هستند اطلاق می گردد.

این افراد صرفاً انسان نمی‌باشند، مثلاً می‌توانند

۱- کامیون‌های شهر تهران   ۲- کارگاه‌های اصفهان   ۳- اطاق‌های یک بیمارستان باشند.

جامعه می تواند محدود یا نامحدود باشد.

۱- جامعه محدود جامعه‌ای است که افراد آن قابل شمارش بوده و پایان‌پذیر باشند.

۲- جامعه نامحدود جامعه‌ای است که در آن افراد قابل شمارش بوده و پایان پذیر نیستند مثل ستارگان آسمان یا افرادی که به دنیا آمده و خواهند آمد.

تعریف پارامتر:

یک مشخصه یا واقعیتی از جامعه است، و آن را با Θ نشان میدهیم.

انواع پارامترها:

۱- **پارامترهای مرکزی:** مانند الف) میانگین جامعه (μ)

ب) میانه

ج) مد و ....

۲- پارامترهای پراکندگی: مانند الف) واریانس (σ^2)

ب) انحراف معیار (σ) و ....

تعریف علم آمار: آمار مجموعه روشهای علمی شامل

۱- جمع آوری و گردآوری اطلاعات

۲- سازماندهی و تنظیم مشاهدات

۳- تجزیه و تحلیل داده‌ها

۴- استنباط و نتیجه‌گیری

است، و در مواردی است که با عدم حتمیت یا عدم قطعیت و تغییرپذیری رو به رو هستیم.

لذا موضوع آمار را می‌توان به دو قسمت عمده تقسیم کرد.

۱- آمار توصیفی

۲- آمار استنباطی

تعریف آمار توصیفی:

این بخش شامل گردآوری، سازماندهی، تلخیص و تجزیه و تحلیل داده‌هاست به عبارت دیگر خلاصه کردن و توضیح خصوصیات مهم مجموعه داده‌ها را آمار توصیفی می‌نامند. این بحث مشتمل است بر فشرده کردن داده‌ها در قالب جداول، نمایش آنها به وسیله نمودارها و محاسبه شاخص‌های عددی گرایش به مرکز و پراکندگی می‌باشد.

تعریف آمار استنباطی:

- برآورد (الف- برآورد نقطه‌ای ب- برآورد فاصله‌ای)
- آزمون فرض
- تعیین رابطه بین متغیرها

جمع آوری و گردآوری اطلاعات:

الف) چه چیزی راجمع‌آوری کنیم؟

مشخصه، متغیر، صفت، $DATA$ می‌باشد که همان اطلاعات پایه‌ای مورد بررسی ما هستند

معمولاً آنها را با X، Y یا Z نشان می‌دهند. متغیر می‌تواند کمی یا کیفی باشد

متغیر کمی قابل اندازه‌گیری است. مثل قد، زن، طول، حجم و ...،

متغیر کیفی قابل اندازه‌گیری با اعداد و ارقام نیست.

انواع متغیرهای کیفی:

۱) اسمی: مانند گروه خونی ، نژاد، ...

۲) ترتیبی: مثل زیبایی، مرغوبیت یک پارچه که قابل اندازه‌گیری نبوده ولی قابل رتبه‌بندی هستند.

،متغیرهای تصادفی کمی دو دسته هستند:

۱- پیوسته: متغیرهای پیوسته معمولاً مقادیر روی یک بازه را می‌گیرند. مانند وزن و قد

۲- گسسته: متغیرهای گسسته معمولاً اعداد صحیح را به خود می‌گیرند و شمارشی هستند مانند تعداد افراد یک خانواده.

مثال ۱.۳ جدول زیر چند سطر از اطلاعات کارکنان یک بخش را نشان می‌دهد

ردیف	نام	سن	جنس	میزان تحصیلات	شغل
۱	اکبری، حسن	۴۲	مرد	دیپلم	تکنسین
۲	بهادری، پروین	۳۷	زن	دکتری	پزشک
۳	جواهری، لاله	۴۰	زن	کارشناسی‌ارشد	سرپرستار
۴	دایی، فاطمه	۳۵	زن	کارشناسی	پرستار

هر سطر داده‌های یک فرد را نشان می‌دهد. هر ستون مقدارهای یک متغیر را بیان می‌کند. به غیر از نام کارکنان، چهار متغیر سن، جنس، میزان تحصیلات، و شغل ثبت شده‌اند. جنس و شغل متغیرهای رسته‌ای‌اند. سن و میزان تحصیلات (تعداد سال‌هایی که درس خوانده‌اند) متغیرهای کمّی‌اند. اغلب نرم‌افزارهای آماری از این قالب برای واردکردن داده‌ها استفاده می‌کنند. ■

ب) از کجا جمع‌آوری کنیم؟ از جامعه

ج) چگونه جمع‌آوری کنیم. (روش تحقیق)؟ ۱- بررسی کامل (سرشماری) ۲- بررسی نمونه‌ای (نمونه‌گیری)

تعریف سرشماری: در این نوع بررسی تک تک افراد جامعه مورد مطالعه، مشاهده و اندازه‌گیری قرار می گردند.

در موارد زیر سرشماری توصیه می‌شود:

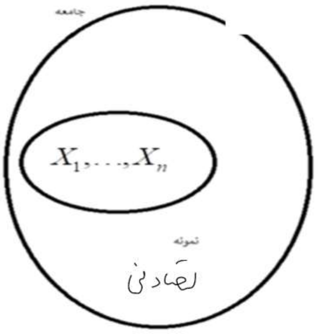
- جامعه کوچک باشد.
- اطلاعات تک تک افراد خواسته شده باشد.
- به تک تک افراد دسترسی داشته باشیم.
- از نظر امکانات و تجهیزات و نیروی انسانی متخصص، محدودیت نداشته باشیم.
- از نظر زمانی محدودیت نداشته باشیم.
- تخریب واحد وجود نداشته باشد.
- عملی یا امکانپذیر باشد.

تعریف نمونه‌گیری:

در این روش بخشی از جامعه مورد بررسی قرار می‌گیرد به طوری که بتوانیم نتیجه به دست آمده را برای کل جامعه تعمیم دهیم.

برای آنکه نمونه بیانگر جامعه باشد، لازم است بصورت تصادفی انتخاب گرددتا بتوان به مصداق "مشت نمونه خروار است" نتیجه حاصل از مطالعه نمونه به تمام جامعه تعمیم داد.

یک نمونه تصادفی با حجم معین نمونه‌ای است که به گونه‌ای انتخاب شده باشد که هرگروه با همان حجم شانس برابر برای انتخاب شدن داشته باشد. این نمونه‌گیری، نمونه‌گیری تصادفی نامیده می‌شود.



$$\Theta: \mu,\ \sigma^2,\ \sigma,\ p,\ ...$$

$$\widehat{\Theta}: \bar{X},\ S^2,\ S,\ \widehat{p},\ ...$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,\quad \widehat{p} = \frac{X}{n}$$

تعریف آماره:

یک آماره مشخصه‌ای از نمونه یا حقیقتی درباره آن است که برای تخمین پارامتر های جامعه بکار میروند،

مانند میانگین نمونه‌ای $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ که برای تخمین میانگین جامعه بکار میرود

واریانس نمونه‌ای $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ که برای تخمین واریانس جامعه بکار میرود

و نسبت نمونه‌ای $\widehat{p} = \frac{X}{n}$ که برای تخمین نسبت در جامعه بکار میرود.

روش های نمونه گیری:

- نمونه‌گیری تصادفی ساده
- نمونه گیری تصادفی طبقه‌بندی شده
- نمونه گیری تصادفی خوشه‌ای
- نمونه گیری تصادفی سیستماتیک،
- نمونه گیری چند مرحله‌ای

در موارد زیر نمونه‌گیری پیشنهاد می گردد:

- – جامعه بزرگ و یا نامحدود باشد.
- –به اطلاعات تک تک افراد نیاز نداشته باشیم
- –تخریب واحد وجود داشته باشد.
- – عملی و امکانپذیر باشد.
- – به تک تک افراد جامعه دسترسی نداشته باشیم.
- – از نظر نیروی انسانی متخصص محدودیت داشته باشید
- – محدودیت زمانی داشته باشیم.

در آمار به علل زیر تاکید بر نمونه‌گیری می باشد:

- سرعت بیشتر
- دقت بیشتر
- عملی‌تر است.
- هزینه کمتر
- عدم تخریب واحد

محصول یک بررسی آماری چیست؟ *DATA*

تنظیم مشاهدات: گردآوری اطلاعات اغلب توده‌های بزرگی از مشاهدات را ایجاد می کند.

برای قابل درک بودن یا به طور مؤثر عرضه شدن باید به نحوی خلاصه شوند.

ارائه روشن و دقیق مشاهدات کمک مهمی به فهم و تفسیر صحیح آنها می کند

برای تنظیم مشاهدات روش‌های زیر وجود دارد:

- مرتب کردن داده‌ها بصورت صعودی یا نزولی : امتیاز آن این است که کوچکترین و بزرگترین داده‌ها به راحتی قابل تشخیص است.

فرض کنید داده‌های زیر، طول عمر ۲۵ لامپ را برحسب ساعت نشان می‌دهد

۱۰۵ ۱۰۴ ۱۱۰ ۱۰۳ ۱۰۵ ۱۰۹ ۹۸ ۹۹ ۱۰۴ ۱۰۳ ۱۰۹ ۱۰۰ ۱۰۵ ۹۸ ۱۰۳ ۱۰۴ ۱۱۰ ۱۰۶ ۱۰۳ ۱۱۰ ۱۰۲ ۱۰۴ ۹۷ ۱۰۱ ۱۰۰

حال داده‌ها را از کوچک به بزرگ مرتب می‌کنیم.

۱۱۰ ۱۱۰ ۱۱۰ ۱۰۹ ۱۰۹ ۱۰۶ ۱۰۵ ۱۰۵ ۱۰۵ ۱۰۴ ۱۰۴ ۱۰۴ ۱۰۴ ۱۰۳ ۱۰۳ ۱۰۳ ۱۰۳ ۱۰۲ ۱۰۱ ۱۰۰ ۱۰۰ ۹۹ ۹۸ ۹۸ ۹۷

اکنون با یک نگاه جمالی می‌تون اطلاعاتی را کسب نمود ین اطلاعات ر از داده‌های اولیه به سختی می‌توانستیم دریابیم. مثلاًکمترین طول عمر، ۹۷ و بیشترین آن ۱۱۰ می‌باشد و یا طول عمر بیش ز ۵۰٪ لامپ‌ها بین ۱۰۰ تا ۱۰۶ است و یا طول عمر ۲۰٪ لامپ‌ها بیش از ۱۰۸ ساعت می‌باشد و ... هر چند ین اطلاعات، کافی نیست ولی برای شروع مفید می‌باشد.

- استفاده از جداول
- استفاده از نمودارها

استفاده از جدول:

۱- **تنظیم داده‌ها بصورت جدول فراوانی:** معمولا برای داده های کیفی و یا گسسته بکار می‌رود

تعریف فراوانی: فراوانی یک مقدار داده برابر تعداد دفعاتی است که این مقدار در مجموعه داده‌ها یافت می‌شود. فراوانی را معمولاً با f_i نشان می‌دهیم.

مثال ۱: گروه خونی ۲۵ نفر در زیر داده شده است .با استفاده از توزیع فراوانی داده ها را تشکیل دهید

$AB\ B\ A\ O\ B\ O\ B\ O\ A\ O\ B\ O\ B\ B\ B\ A\ O\ AB\ AB\ O\ B\ AB\ O\ A\ A$

حل: جدول فراوانی یا توزیع فراوانی داده ها در زیر نوشته شده است:

گروه خونی	A	B	O	AB	Σ
f_i	5	8	8	4	25

صفت کیفی اسمی

لازم به ذکر است که تنظیم داده‌ها به صورت جدول فراوانی علاوه بر تنظیم یک تلخیص است، زیرا نشان می‌دهد کدام مقدار دارای بیشترین تکرار (**مد**) و کدام یک کمترین تکرار را دارد.

اگر در همان کارخانه، مهارت کارگران را مورد تحقیق آماری قرار دهیم و آن‌ها را برحسب درجه مهارت طبقه‌بندی کنیم، فرض کنید جدول زیر به دست آید

گروه	تعداد
ضعیف	۲۰
متوسط	۱۵
خوب	۳۵
عالی	۳۰
Σ	100

نوع صفت کیفی ترتیبی

در ین مثال، طبقات دارای ترتیب می‌باشند ولی اگر به گروه‌ها کد بدهیم مثلاً برای ضعیف عدد ۱ و بری متوسط عدد ۲ و برای خوب عدد ۳ و برای عالی عدد ۴ را در نظر بگیریم، در اینجا ۲ بهتر از ۱ است ولی دو برابر آن نیست، همچنین ۳ بهتر از ۲ می‌باشد ولی ۲ + ۱ برابر ۳ نمی‌شود (زیر اصولاً معنی ندارد که ضعیف + متوسط، بتوند برابر خوب باشد) داده‌های فوق از نوع صفت کیفی ترتیبی هستند

مثال ۱۲. اگر در یک نظرخواهی ز ۲۰ نفر در مورد موافقت یا مخالفت با یک لایحه سؤال شود و مثلاً ۱۲ نفر موافق و ۸ نفر مخالف باشند، جدول توزیع فراونی به صورت زیر بیان می‌شود

E_i	f_i
موفق	۱۲
مخالف	۸
Σ	۲۰

صفت کیفی اسمی

مثال ۱۳. کمیته پزشکی متشکل است از ۵ پزشک، ۳ پرستار و ۲ بهیار. جدول توزیع فراوانی عبارت است از

E_i	f_i
پزشک	۵
پرستار	۳
بهیار	۲
$\sum$	۱۰

مثال ۱۴. فرض کنید میزان تحصیلات ۸۰ نفر از پرسنل یک اد ره به صورت زیر باشد:

میزان تحصیلات	توزیع فروانی
زیر دیپلم	۶
دیپلم	۱۰
فوق دیپلم	۱۴
لیسانس	۲۲
فوق لیسانس	۱۷
دکتری	۱۱
$\sum$	۸۰

مثال ۱۶. جدول زیر، جدول توزیع فراوانی نمرات درس ریاضی ۱۲ د نش‌آموز ر نشان می‌دهد

x_i	f_i
۱۳	۲
۱۴	۳
۱۵	۰
۱۶	۴
۱۷	۱
۱۸	۲
$\sum$	۱۲

در تمرین‌های زیر، معلوم کنید کد م صفت کیفی و کدام کمی است.

۱) حجم نوشابه‌های پر شده درون شیشه‌های نوشابه.

۲) فاصله بین شهرها

۳) قیمت یک کیلو روغن حیوانی در شهرهای مختلف یک استان

۴) رده‌بندی تیم‌های فوتبال شرکت‌کننده در یک مسابقه

۵) قیمت یک کیلو تخم‌مرغ در ماه‌های مختلف یک سال

۶) میزن تحصیلات کارکنان یک اد ره (بر حسب دیپلم، فوق دیپلم، لیسانس، فوق لیسانس)

۷) رنگ نمونه‌ای ز ماشین‌ها

۸) وزن یک کشتی ۱۴۷۲ تن ست.

۹) تعداد محصلان غایب یک کلاس در روزگذشته ۱۴ نفر بوده ست.

مفهوم و کاربرد سیگما

فرض کنید می‌خواهیم اعداد ۱، ۲، ۳، ۴ و ۵ را با هم جمع کنیم، می‌نویسیم

$$۱+۲+۳+۴+۵$$

حال می‌خواهیم، نشانه ی ر معرفی کنیم که معرف جمع بالا باشد، این نشان را با نماد $\sum$ نمایش می‌دهیم و آن را می‌خوانیم **سیگما** و آن ر به صورت زیر تعریف می‌کنیم

$$\sum_{i=۱}^{۵} i = ۱+۲+۳+۴+۵$$

حال فرض کنید، می‌خواهیم مقادیر $x_۱$، $x_۲$، ...، x_n را با هم جمع کنیم، با توجه به نماد معرفی شده، می‌نویسیم

$$\sum_{i=۱}^{n} x_i = x_۱ + x_۲ + \cdots + x_n$$

و آن را می‌خونیم **سیگمای x_i، i از یک تا n**. i را اندیس می‌نامیم و عمل جمع روی اندیس i، نجام می‌گیرد. منظور از $x_۱$ یعنی عدد اول و $x_۲$ یعنی عدد دوم و ...

مثال ۱. می‌خواهیم اعد د ۲، ۴، ۶، ۷ و ۹ را جمع کنیم، با توجه به نماد معرفی شده داریم

$$\begin{aligned}\sum_{i=۱}^{۵} x_i &= x_۱+x_۲+x_۳+x_۴+x_۵\\ &= ۲+۴+۶+۷+۹ = ۲۸\end{aligned}$$

مثال ۲. اگر x مقادیر $x_۱, x_۲, x_۳, x_۴, x_۵$ را اختیار کند، مطلوب است $\sum_{i=۱}^{۵}(x_i+۳)$

$$\begin{aligned}\sum_{i=۱}^{۵}(x_i+۳) &= (x_۱+۳)+(x_۲+۳)+(x_۳+۳)+(x_۴+۳)+(x_۵+۳)\\ &= x_۱+x_۲+x_۳+x_۴+x_۵+۳+۳+۳+۳+۳\\ &= \sum_{i=۱}^{۵} x_i + ۵\times ۳\end{aligned}$$

خواص $\sum$:

$$(۱)\quad \sum_{i=۱}^{n}(x_i+c) = \sum_{i=۱}^{n} x_i + nc \qquad (۲)\quad \sum_{i=۱}^{n} Cx_i = C\sum_{i=۱}^{n} x_i$$

$$(۳)\quad \sum_{i=۱}^{n}(ax_i+b) = a\sum_{i=۱}^{n} x_i + nb$$

مثال ۴. اگر x مقادیر ۳، ۵، ۶ و ۷ را اختیار کند، مطلوب است محاسبه $\sum_{i=۱}^{۴}(۳x_i-۲)$.

$$\begin{aligned}\sum_{i=۱}^{۴}(۳x_i-۲) &= ۳\sum_{i=۱}^{۴} x_i - ۴\times ۲ = ۳(x_۱+x_۲+x_۳+x_۴) - ۸\\ &= ۳(۳+۵+۶+۷) - ۸ = ۵۵\end{aligned}$$

مثال ۵. اگر x مقادیر ۴، ۵، ۶ و ۹ را اختیار کند، مطلوب است محاسبه

الف. $\frac{۱}{۴}\sum_{i=۱}^{۴} x_i$ ب. $\sum_{i=۱}^{۴}(x_i-۶)^۲$ پ. $\sum_{i=۱}^{۴}|x_i-۶|$

حل: الف.

$$\frac{۱}{۴}\sum_{i=۱}^{۴} x_i = \frac{۱}{۴}(x_۱+x_۲+x_۳+x_۴) = \frac{۱}{۴}(۴+۵+۶+۹) = \frac{۲۴}{۴} = ۶$$

ب.

$$\begin{aligned}\sum_{i=۱}^{۴}(x_i-۶)^۲ &= (x_۱-۶)^۲+(x_۲-۶)^۲+(x_۳-۶)^۲+(x_۴-۶)^۲\\ &= (۴-۶)^۲+(۵-۶)^۲+(۶-۶)^۲+(۹-۶)^۲\\ &= (-۲)^۲+(-۱)^۲+۰^۲+۳^۲ = ۴+۱+۰+۹ = ۱۴\end{aligned}$$

پ.

$$\begin{aligned}\sum_{i=۱}^{۴}|x_i-۶| &= |x_۱-۶|+|x_۲-۶|+|x_۳-۶|+|x_۴-۶|\\ &= |۴-۶|+|۵-۶|+|۶-۶|+|۹-۶|\\ &= |-۲|+|-۱|+|۰|+|۳| = ۲+۱+۰+۳ = ۶\end{aligned}$$

مثال ۶. اگر x مقادیر $x_۱$، $x_۲$، $x_۳$ و $x_۴$ و y مقادیر $y_۱$، $y_۲$، $y_۳$ و $y_۴$ را اختیار کند، مطلوب است محاسبه $\sum_{i=۱}^{۴}(x_i+y_i)$.

حل:

$$\begin{aligned}\sum_{i=۱}^{۴}(x_i+y_i) &= (x_۱+y_۱)+(x_۲+y_۲)+(x_۳+y_۳)+(x_۴+y_۴)\\ &= x_۱+x_۲+x_۳+x_۴+y_۱+y_۲+y_۳+y_۴\\ &= \sum_{i=۱}^{۴} x_i + \sum_{i=۱}^{۴} y_i\end{aligned}$$

خواص $\sum$:

$$(۴)\quad \sum_{i=۱}^{n}(x_i+y_i) = \sum_{i=۱}^{n} x_i + \sum_{i=۱}^{n} y_i$$

مثال ۷. اگر x مقادیر ۱، ۳، ۴، ۷ و y مقادیر ۲، ۴، ۵ و ۶ را اختیار کنند. مطلوب است محاسبه $\sum_{i=۱}^{۴} x_i y_i$.

حل:

$$\begin{aligned}\sum_{i=۱}^{۴} x_i y_i &= x_۱y_۱+x_۲y_۲+x_۳y_۳+x_۴y_۴\\ &= ۱\times ۲+۳\times ۴+۴\times ۵+۷\times ۶ = ۷۶\end{aligned}$$

مثال ۸. اگر x مقادیر ۲، ۴، ۶ و ۸ و y مقادیر ۱، ۳، ۵ و ۷ را اختیار کنند. مطلوب است محاسبه $\sum_{i=1}^{4}(x_i-5)(y_i-4)$

$$\sum_{i=1}^{4}(x_i-5)(y_i-4)=(x_1-5)(y_1-4)+(x_2-5)(y_2-4)$$
$$+(x_3-5)(y_3-4)+(x_4-5)(y_4-4)$$
$$=(2-5)(1-4)+(4-5)(3-4)$$
$$+(6-5)(5-4)+(8-5)(7-4)$$
$$=9+1+1+9=20$$

مثال ۹. با توجه به آمار داده شده زیر برای x و y مقادیر Σx، Σy، $\Sigma(x-4)$، Σx^2، Σxy، $\Sigma x\Sigma y$ و $(\Sigma x)^2$ را محاسبه کنید.

x	۲	۳	۵
y	۴	۵	۸

x	y	$(x-4)$	x^2	xy
۲	۴	−۲	۴	۸
۳	۵	−۱	۹	۱۵
۵	۸	۱	۲۵	۴۰
$\Sigma x=10$	$\Sigma y=17$	$\Sigma(x-4)=-2$	$\Sigma x^2=38$	$\Sigma xy=63$

$$\Sigma X\Sigma Y=10\times 17=170$$
$$(\Sigma X)^2=(10)^2=100$$

مثال ۱۰. اگر x مقادیر ۴، ۵، ۶ و ۹ را اختیار کند مطلوب ست محاسبه $\sum_{i=1}^{4}(x_i-6)^2$

$$\sum_{i=1}^{4}(x_i-6)^2=(x_1-6)^2+(x_2-6)^2+(x_3-6)^2+(x_4-6)^2$$
$$=(4-6)^2+(5-6)^2+(6-6)^2+(9-6)^2=14$$

مثال ۱۱. حاصل $\sum_{i=1}^{n}(i+3)$ را به دست آورید.

$$\sum_{i=1}^{n}(i+3)=\sum_{i=1}^{n}i+\sum_{i=1}^{n}3$$
$$=1+2+3+\cdots+n+3n$$
$$=\frac{n(n+1)}{2}+3n$$
$$=n\left[\frac{n+1}{2}+3\right]=n\left[\frac{n+7}{2}\right]=\frac{n\left(n+7\right)}{2}$$

مجموعه مسائل

۱) مطلوب ست محاسبه مجموع مقادیر زیر

الف. $\sum_{i=1}^{n}i$ ب. $\sum_{i=1}^{n}i^2$ پ. $\sum_{i=1}^{n}i^3$

۲) اگر x مقادیر ۲ و ۳ و ۴ و ۵ را اختیار کند، مطلوب است محاسبه

الف. $\sum_{i=1}^{4}x_i$ ب. $\sum_{i=1}^{4}x_i^2$ پ. $\sum_{i=1}^{4}(x_i-3.5)^2$ ت. $\sum_{i=1}^{4}|x_i-3.5|$

۳) اگر x مقادیر ۲ و ۵ و ۶ و ۷ و y مقادیر ۱ و ۲ و ۳ و ۴ را اختیار کند، مطلوب است محاسبه

الف. $\sum_{i=1}^{4}x_iy_i$ ب. $\sum_{i=1}^{4}(x_i-5)(y_i-2.5)$

پ. $\sum_{i=1}^{4}(x_iy_i-3x_i+2y_i)$ ت. $\sum_{i=1}^{4}(7x_i-6y_i)$

۴) اگر x مقادیر ۱، ۲، ۴ و ۹ ر ختیار کند مطلوب است محاسبه

الف. $\sum_{i=1}^{4}\frac{1}{x_i}$ ب. $\sum_{i=1}^{4}|x_i-1|$ ج. $\sum_{i=1}^{4}\sqrt{x_i}$

۴) اگر x مقادیر ۱، ۲، ۴ و ۹ ر ختیار کند مطلوب است محاسبه

الف. $\sum_{i=1}^{4}\frac{1}{x_i}$ ب. $\sum_{i=1}^{4}|x_i-1|$ ج. $\sum_{i=1}^{4}\sqrt{x_i}$

۵) فرض کنیم: $\sum_{i=1}^{3}X_i=6$ و $\sum_{i=1}^{3}X_i^2=14$ آنگاه $\sum_{i=1}^{3}(X_i-2)^2$ را محاسبه کنید.

۲ -تنظیم داده‌ها به صورت دسته‌های رده بندی شده: برای داده های گسسته و پیوسته بکار میرود

روشی برای دسته بندی داده‌ها:

۱- تعیین تعداد رده‌ها:

تعداد رده‌ها یا k نباید خیلی کم و نباید خیلی زیاد باشد معمولاً تعداد آنها بین ۲ و ۱۸ پیشنهاد شده است.

مقدار آن بسته به نوع داده‌ها و فرد استفاده کننده از داده‌ها متفاوت است.

اگر تعداد رده‌ها خیلی کم باشد دقت کار ما کاهش می‌یابد

اگر تعداد رده‌ها خیلی زیاد باشد با خود داده‌ها چندان فرقی نمی‌کند.

پیشنهادات خاصی از طریق آمارشناسان داده شده است که این نوع پیشنهادات نمی‌تواند همه جا مورد استفاده قرار گیرد و به عنوان الگوی اولیه مورد استفاده قرار می‌گیرد.

یک قاعده استفاده از دستور استورگ است که درآن k از رابطه $k = \lceil 1 + 3.322 \log n \rceil$ بدست می‌آید که در آن log لگاریتم اعشاری (پایه ۱۰) و $\lceil x \rceil$ اولین عدد صحیح و بزرگتر از x است.

۲- تعیین میزان تغییر پذیری داده‌ها:

$\frac{\text{واحد گرد شده}}{2} = s$=میزان تغییرپذیری داده‌ها

۳- تعیین کران‌های داده‌ها:

s- کوچکترین داده‌ها = کران پایین داده‌ها $= L$

s+بزرگترین داده‌ها = کران بالای داده‌ها $= U$

٤-تعیین دامنه داده‌ها:

کران پایین داده‌ها - کران بالای داده‌ها $= R$

٥-تعیین فاصله دسته ها:

$$I = \frac{R}{k}$$

٦-دسته بندی داده‌ها:

یک روش برای رده بندی داده‌های گسسته و پیوسته بصورت زیر پیشنهاد شده است:

رده های گسسته	رده های پیوسته
$(L, L+I-1)$	$(L, L+I)$
$(L+I, L+2I-1)$	$(L+I, L+2I)$
$\vdots$	$\vdots$

۱+ مرز پایین − مرز بالا $= I$ مرز پایین − مرز بالا $= I$

نکته: اعداد سمت چپ و سمت راست رده‌ها را مرزهای رده می نامیم که اعداد سمت چپ مزر پایینی و اعداد سمت راست مرز بالایی رده را مشخص می‌کند.

نکته: طول رده‌ها می‌تواند متفاوت باشد مخصوصاً برای رده اول و رده آخر. متفاوت بودن طول رده برای رده بندی داده‌ها عیب محسوب نمی‌شود چرا که رده بندی داده‌ها ما را به شناخت تغییرات داده‌ها به عنوان یک الگو یا مدل، راهنمایی می‌کند.

نکته: برای داده‌های پیوسته نوع رده‌بندی طوری است که در بین رده‌ها هیچ گونه گسستگی وجود ندارد. اما برای داده‌های گسسته بین دو رده این گسستگی مشهود است.

نکته: برای منظور کردن تعداد تکرار هر رده این ابهام وجود دارد که اگر عددی برابر مرز رده وجود داشت این عدد در کدام رده منظور شود. پاسخ این سؤال بدین صورت از ابهام خارج می‌شود که اگر عددی برابر مرز رده باشد آن عدد در رده‌ای منظور شود که برابر مرز پایینی آن رده است. برای تحقق این پاسخ گاهی اوقات مرزهای رده‌ها را با یک رقم اعشار اضافی منظور می‌کنیم به عنوان مثال رده‌ها به صورت زیر تعریف می‌کنیم:

۰ − ۳/۵ و ۳/۵ − ۷ و ...

مثال ۱ : وزن ۴۰ قالب کره که به بزرگترین واحد گرد شده‌اند به قرار زیر است:

۳۳	۴۶	۳۸	۳۴	۵۱	۳۶	۴۷	۲۴	۳۵	۵۲
۴۰	۴۲	۴۰	۴۱	۳۴	۴۷	۵۰	۳۸	۳۶	۴۷
۲۱	۴۱	۳۷	۳۷	۳۵	۳۰	۳۲	۳۰	۲۹	۲۶
۴۳	۳۴	۳۱	۴۳	۲۳	۴۵	۳۵	۲۶	۳۰	۳۱

جدول رده بندی این داده‌ها را تشکیل دهید.

حل:

- $k=\lceil 1+3.322\log 40\rceil=\lceil 6.31\rceil=7$
  پس داده‌ها را به ۷ رده گروه بندی می کنیم.
- چون داده‌ها به واحد گرد شده‌اند، پس
  $\frac{1}{2}=\frac{\text{واحد گرد شده}}{2}=$ میزان تغییرپذیری
- می دانیم $\min=21,\ \max=52$ است

$20.5=21-0.5=s-$ کوچکترین داده = کران پایین داده‌ها

$52.5=52+0.5=s+$ بزرگترین داده = کران بالای داده‌ها

$R=52.5-20.5=32$

$I=\frac{R}{k}=\frac{32}{7}=4.571\approx 5$

بنابراین جدول رده‌بندی داده‌ها بصورت زیر است:

فراوانی مطلق	خط و نشان	رده ها	i
3	$\|\|\|$	$20.5-25.5$	1
6	$\cancel{\|\|\|\|}\ \|$	$25.5-30.5$	2
10	$\cancel{\|\|\|\|}\ \cancel{\|\|\|\|}$	$30.5-35.5$	3
8	$\cancel{\|\|\|\|}\ \|\|\|$	$35.5-40.5$	4
6	$\cancel{\|\|\|\|}\ \|$	$40.5-45.5$	5
5	$\cancel{\|\|\|\|}$	$45.5-50.5$	6
2	$\|\|$	$50.5-55.5$	7
40	---	---	Σ

توجه کنید که رده‌ها فواصلی هستندکه از سمت چپ بسته و از سمت راست باز هستند، مثلاً رده چهارم $[35.5,40.5)$

تعریف: ابتدا و انتهای هر دسته را حدود دسته و متوسط حد بالا و حد پایین هر دسته را به نام «حد متوسط دسته» ($\bar{x}_i$) می‌نامیم. همچنین تعداد مشاهدات در داخل هر دسته را فراوانی دسته یا به طور ساده‌تر فراوانی (f_i) می گویند.

تعریف : سری فراوانی‌ها با دسته‌های مربوطه با حد وسط دسته ها را توزیع فراوانی می‌نامند.

تعریف: فراوانی تجمعی (f_{ci}) هر دسته عبارتست از فراوانی آن دسته به علاوه فراوانی دسته‌های ما قبل.

$$f_{ci}=f_1+f_2+\cdots+f_i$$

تعریف: فراوانی نسبی هر دسته عبارتست از فراوانی هر دسته تقسیم به تعداد کل داده‌ها و آن را با r_i نشان می‌دهیم.

$$r_i=\frac{f_i}{n},\quad i=1,2,\ldots$$

مثال: برای مثال ۱ جدول زیر را تکمیل و مفهوم هریک از اعداد آن را بیان کنید

i	رده ها	x_i	f_i	f_{ci}	r_i	r_{ci}	درصد فراوانی	درصد فراوانی تجمعی
1	20.5-25.5	23	3	3	0.075	0.075	7.5	7.5
2	25.5-30.5	28	6	9	0.15	0.225	15	22.5
3	30.5-35.5	33	10	19	0.25	0.475	25	47.5
4	35.5-40.5	38	8	27	020	0.675	20	67.5
5	40.5-45.5	43	6	33	0.15	0.825	15	82.5
6	45.5-50.5	48	5	38	0.125	0.95	12.5	95
7	50.5-55.5	53	2	40	0.05	1	5	100
Σ	---	---	40	---	1	---	100	---

تعریف: فراوانی نسبی تجمعی r_{ci} عبارتست از فراوانی نسبی یک دسته به علاوه فراوانی نسبی دسته‌های ما قبل.

$$r_{ci}=r_1+r_2+\cdots+r_i$$

مثال ۲۶. فرض کنید آمار مربوط به تولید کارخانه پارچه‌بافی بی‌تا در طول ۶۰ روز گذشته به صورت زیر باشد. جدول توزیع فراوانی را تنظیم نمایید.

$K = 6$

۴٫۱	۵٫۷	۶٫۵	۶٫۹	۷٫۵	۸٫۳
۴٫۵	۵٫۸	۶٫۵	۷٫۰	۷٫۵	۸٫۳
۴٫۹	۵٫۹	۶٫۵	۷٫۱	۷٫۷	۸٫۷
۵٫۱	۵٫۹	۶٫۷	۷٫۱	۷٫۷	۸٫۹
۵٫۲	۶٫۰	۶٫۷	۷٫۱	۷٫۷	۸٫۹
۵٫۳	۶٫۱	۶٫۷	۷٫۳	۷٫۷	۹٫۱
۵٫۵	۶٫۱	۶٫۷	۷٫۳	۷٫۹	۹٫۲
۵٫۶	۶٫۲	۶٫۹	۷٫۳	۸٫۱	۹٫۳
۵٫۶	۶٫۳	۶٫۹	۷٫۳	۸٫۱	۹٫۴
۵٫۷	۶٫۳	۶٫۹	۷٫۳	۸٫۱	۹٫۶

$$I = \frac{R}{K} = \frac{\text{بزرگترین عدد} - \text{کوچکترین عدد}}{\text{تعداد طبقات}}$$

$$= \frac{۹٫۶ - ۴٫۱}{۶} = \frac{۵٫۵}{۶} = ۰٫۹۲ \Rightarrow I \approx 1$$

حدود طبقات	x_i	f_i	f_{P_i}	P_i	F_{C_i}	F_{CP_i}	P_C
۴٫۰ − ۴٫۹	۴٫۴۵	۳	۰٫۰۵	۵	۳	۰٫۰۵	۵
۵٫۰ − ۵٫۹	۵٫۴۵	۱۱	۰٫۱۸	۱۸	۱۴	۰٫۲۳	۲۳
۶٫۰ − ۶٫۹	۶٫۴۵	۱۷	۰٫۲۸	۲۸	۳۱	۰٫۵۱	۵۱
۷٫۰ − ۷٫۹	۷٫۴۵	۱۶	۰٫۲۷	۲۷	۴۷	۰٫۷۸	۷۸
۸٫۰ − ۸٫۹	۸٫۴۵	۸	۰٫۱۴	۱۴	۵۵	۰٫۹۲	۹۲
۹٫۰ − ۹٫۹	۹٫۴۵	۵	۰٫۰۸	۸	۶۰	۱	۱۰۰
		۶۰	۱	۱۰۰			

نکاتی در مورد جدول فراوانی

۱. مجموع فراو نی تمام طبقات، برابر با تعداد کل داده‌ها ست.

۲. مجموع فراو نی نسبی تمام طبقات، برابر با یک ست.

۳. فراوانی تجمعی آخرین طبقه، برابر با تعداد کل داده‌ها است.

۴. فراوانی تجمعی نسبی آخرین طبقه، برابر با یک ست.

مجموعه مسائل

۱) نمرات درس ۱۲ دانشجو به صورت زیر داده شده، جدول توزیع فراوانی ر تشکیل دهید.
متوسط، عالی، عالی، خوب، بد، متوسط، عالی، متوسط، بد، خوب، بد، خوب

۲) در تمرین (۱) ستون‌های مربوط به فراوانی نسبی و درصد فراوانی را تشکیل دهید. چند درصد از دانشجویان دارای نمره عالی هستند؟ چند درصد از دانشجویان دارای نمره بد می‌باشند؟

۳) د ده‌های زیر مربوط به نمره درس ریاضی ۵۰ دانشجو می‌باشد. داده‌ها را با فاصله طبقات ۱۰، طبقه‌بندی کنید و سپس ستون‌های مربوط به فراوانی نسبی، درصد فرونی، فراوانی تجمعی و درصد فروانی تجمعی ر تشکیل دهید.

۳۹	۳۸	۳۷	۳۴	۳۳	۳۱	۳۰	۲۲	۲۰	۱۴	۱۰	۸	
۵۳	۵۱	۵۰	۴۹	۴۸	۴۵	۴۵	۴۵	۴۳	۴۲	۴۲	۴۱	۴۰
۶۳	۶۲	۶۱	۶۰	۵۹	۵۸	۵۸	۵۷	۵۷	۵۶	۵۵	۵۵	۵۴
۹۵	۸۳	۸۲	۸۰	۷۸	۷۵	۷۱	۷۰	۶۸	۶۷	۶۶	۶۵	

۶) برای د ده‌های تمرین ۳، یک جدول فراوانی با ۱۵ طبقه تشکیل دهید.

۴) در جدول زیر، فاصله طبقات، نماینده طبقه و حدود واقعی را مشخص کنید.

الف.

فرونی	حدود طبقات
۴	۸۰ − ۸۶
۵	۸۷ − ۹۳
۲	۹۴ − ۱۰۰
۳	۱۰۱ − ۱۰۷

ب.

فراوانی	حدود طبقات
۷	۰٫۰ − ۵٫۵
۱۰	۵٫۶ − ۱۱٫۱
۱۲	۱۱٫۲ − ۱۶٫۷
۳	۱۶٫۸ − ۲۲٫۳

۵) برای جدول توزیع فراو نی تمرین (۴)، ستون‌های مربوط به فراوانی نسبی، فراو نی تجمعی، درصد فرونی تجمعی را تشکیل دهید.

۷) درصد سود سهام پرداختی ۳۵ شرکت پذیرفته شده در بازار بورس ورق بهادار تهران پس از آن که با توجه به تورم تعدیل شده است در زیر آمده ست:

−۲٫۵	۱٫۵	۱۵٫۰	۵٫۸	۸٫۴	۱۲٫۵	۱۰٫۰
۳٫۷	۴٫۸	۲٫۲	۱۸٫۹	۱۲٫۷	۱۱٫۸	−۵٫۵
۵٫۰	۱۶٫۸	۱۲٫۶	−۹٫۶	۱۰٫۱	۱۳٫۴	۱۱٫۲
۴٫۳	−۱٫۱	۱۹٫۵	۵٫۱	−۴٫۳	۱۷٫۳	۴٫۳
۱۴٫۶	۵٫۹	۵٫۵	۱۴٫۵	۱۲٫۷	۱۰٫۰	۵٫۸

جدول توزیع فرونی سود سهام پرداختی ر در ۶ طبقه تهیه نمایید.

۸) برای د ده‌های جدول زیر:

جدول توزیع فراوانی ر به نحوی تشکیل دهید که حد پایین طبقه اول ۱۳/۵ و فاصله طبقه معادل ۱ باشد.

۱۵/۷	۱۸/۱	۱۷/۱	۱۷/۷	۱۶/۹	۱۵/۶	۱۷/۱	۱۷/۷	۱۶/۶	۱۸/۱	۱۶/۸	۱۷/۱	۱۸/۸	۱۶/۰	۱۵/۹	۱۷/۶
۱۸/۲	۱۶/۷	۱۴/۱	۱۶/۱	۱۶/۹	۱۷/۹	۱۷/۰	۱۶/۷	۱۷/۱	۱۴/۹	۱۶/۸	۱۷/۱	۲۰/۲	۱۳/۷	۱۷/۳	۱۶/۸
۱۶/۹	۱۶/۶	۱۷/۵	۱۶/۸	۱۶/۹	۱۷/۹	۱۷/۰	۱۴/۶	۱۶/۰	۱۷/۴	۱۷/۶	۱۸/۶	۱۶/۳	۱۷/۷	۱۶/۱	۱۸/۶
۱۵/۹	۱۷/۸	۱۷/۹	۱۵/۱	۱۸/۴	۱۵/۷	۱۷/۳	۱۶/۹	۱۸/۱	۱۷/۰	۱۹/۱	۱۶/۳	۱۹/۲	۱۷/۳	۱۶/۱	۱۵/۶
۱۷/۹	۱۶/۰	۱۷/۰	۱۵/۲	۱۷/۳	۱۶/۰	۱۷/۸	۱۶/۵								

۹) میانگین قد ۳۰ نفر به صورت زیر می‌باشد. جدول توزیع فراوانی ر در ۵ طبقه تنظیم نمایید.

۱۶۵	۱۷۵	۱۸۲	۱۸۳	۱۶۴	۱۷۰	۱۶۰	۱۶۳	۱۸۲	۱۸۹	۱۸۱	۱۷۹	۱۷۹	۱۸۱	۱۸۹	۱۸۸	۱۶۳	۱۶۴	۱۷۷	۱۷۴
۱۶۲	۱۶۳	۱۸۴	۱۸۵	۱۷۱	۱۷۲	۱۶۶	۱۸۷	۱۶۷	۱۶۸										

۱۰) نمرات درس یک کلاس ۲۰ نفره به صورت زیر است.
جدول فراوانی مربوط به نمرات ر در سه دسته تنظیم نمایید.

۱۱	۱۲	۱۶	۱۸	۱۱	۱۰	۱۲
۱۶	۱۹	۱۵	۱۳	۷	۱۷	۲۰
۵	۱۹	۱۵	۱۰	۱۷	۱۸	

۱۱) سود حاصل ز فروش کالا در یک فروشگاه در ۳۰ روز متوالی به شرح زیر است (مقادیر به هزر ریال می‌باشد).

۱۰۰/۶	۸۴/۲	۸۵/۲	۹۱/۰	۸۵/۵	۹۴/۷
۱۰۱/۷	۸۷/۲	۹۲/۷	۹۸/۳	۷۳/۶	۸۷/۶
۱۱۰/۴	۱۰۵/۸	۱۰۳/۷	۹۳/۷	۸۹/۰	۷۳/۱
۹۰/۰	۶۸/۱	۹۱/۶	۹۵/۹	۹۴/۷	۷۹/۱
۹۳/۷	۷۹/۴	۸۴/۵	۹۴/۲	۸۸/۶	۹۷/۸

جدول توزیع فروانی ر تنظیم نمایید.

۱۲) تعداد قرص‌های آسپرین که در ۵۰ خانواده در عرض ماه گذشته مصرف شده‌اند عبارتند ز:

۷	۹	۳	۱۱	۴	۵	۳	۲	۸	۳
۳	۲	۱	۴	۱۱	۶	۸	۹	۷	۴
۵	۱۱	۹	۴	۵	۲	۳	۴	۲	۷
۳	۵	۴	۹	۲	۲	۳	۴	۹	۱۱
۸	۱۱	۳	۲	۲	۶	۴	۵	۹	۸

یک جدول فراوانی کامل برای این د ده‌ها تنظیم نمایید (جدول فوق د رای ۶ طبقه باشد).

۱۳) در آزمونی یک پرسشنامه ۲۰ پرسشی برای ندازه‌گیری استعداد ریاضی، به ۴۰ نفر از دانش‌آموزن کلاس پنجم ابتدایی یکی از دبستان‌ها داده ند. نمره‌های ین آزمون عبارتند از:

۱۴	۱۲	۱۲	۱۳	۱۰	۷	۸	۱۱	۱۴	۱۳
۱۴	۱۳	۹	۱۰	۱۰	۱۴	۱۳	۱۲	۱۱	۱۰
۱۲	۱۷	۱۴	۹	۱۲	۱۲	۱۳	۱۳	۱۲	۱۴
۹	۷	۱۵	۱۵	۱۲	۱۱	۱۱	۱۲	۱۰	۹

یک جدول فراو نی کامل برای داده‌های فوق تنظیم نمایید (جدول توزیع فراوانی دارای ۸ طبقه باشد).

۱۴) وزن‌های ۴۰ بسته پسته تا نزدیک‌ترین کیلو عبارتند از:

۱۳۸	۱۶۴	۱۵۰	۱۳۲	۱۴۴	۱۲۵	۱۴۹	۱۵۷	۱۴۶	۱۵۸
۱۴۰	۱۴۷	۱۳۶	۱۴۸	۱۵۲	۱۴۴	۱۶۸	۱۲۶	۱۳۸	۱۷۶
۱۶۳	۱۱۹	۱۵۴	۱۶۵	۱۴۶	۱۷۳	۱۴۲	۱۴۷	۱۳۵	۱۵۳
۱۴۰	۱۳۵	۱۶۱	۱۳۵	۱۴۵	۱۴۲	۱۵۰	۱۵۶	۱۴۵	۱۲۸

یک جدول توزیع فراونی کامل با هشت طبقه با فواصل مساوی پیدا کنید.

۱۵) میزان هموگلوبین خون در ۵۰ بیمار سرطانی برحسب گرم در ۱۰۰ میلی‌لیتر عبارتند از:

۱۳/۶	۱۴/۸	۱۳/۷	۱۴/۲	۱۱/۵	۱۱/۹	۱۳/۸	۱۴/۶	۱۴/۲	۱۲/۷
۱۳/۴	۱۱/۵	۱۱/۹	۱۴/۸	۱۲/۷	۱۲/۴	۱۵/۳	۱۵/۲	۱۳/۵	۱۵/۰
۱۲/۴	۱۲/۰	۱۳/۸	۱۱/۷	۱۰/۰	۱۳/۲	۱۵/۵	۱۴/۰	۱۳/۵	۱۵/۰
۱۲/۷	۱۲/۹	۱۳/۷	۱۵/۱	۱۳/۵	۱۲/۷	۱۵/۷	۱۰/۹	۱۴/۰	۱۴/۸
۱۴/۰	۱۳/۸	۱۲/۷	۱۱/۹	۱۲/۰	۱۱/۴	۱۱/۱	۱۳/۷	۱۳/۲	۱۶/۲

با تشکیل ۷ طبقه با فاصله ۰/۹، جدول فرونی کاملی را تشکیل دهید.

۱۶) جدول توزیع فراوانی نمرت درس آمار یک کلاس ۸۰ نفره را برحسب ۱۰۰ نشان می‌دهد.

فروانی مطلق	حدود طبقات
۲۵	۲۰ — ۳۰
۱۵	۳۰ — ۴۰
۵	۴۰ — ۵۰
۱۵	۵۰ — ۶۰
۲۰	۶۰ — ۷۰
۱۰	۷۰ — ۸۰
۷	۸۰ — ۹۰
۳	۹۰ — ۱۰۰

جدول توزیع فروانی ر تنظیم نمایید.

۱۸) جدول زیر توزیع فرونی حقوق ماهانه ۶۰ کارمند یک شرکت را نشان می‌دهد.

فراوانی مطلق	حدود طبقات
۸	۳۰ — ۵۰
۱۵	۵۰ — ۷۰
۲۵	۷۰ — ۹۰
۸	۹۰ — ۱۱۰
۴	۱۱۰ — ۱۳۰

جدول توزیع فرونی ر تشکیل دهید.

۱۹) جدول توزیع فراوانی صفت x به صورت زیر است:

فراونی مطلق	حدود طبقات
۲	۴۰ — ۴۵
۳	۴۵ — ۵۰
۷	۵۰ — ۵۵
۱۰	۵۵ — ۶۰
۱۲	۶۰ — ۶۵
۱۵	۶۵ — ۷۰
۱۲	۷۰ — ۷۵
۱۰	۷۵ — ۸۰
۸	۸۰ — ۸۵
۳	۸۵ ۹۰

مطلوب است محاسبه فراوانی نسبی، نماینده طبقه، درصد فراونی نسبی، فرونی تجمعی، فراوانی تجمعی نسبی و درصد فراوانی تجمعی نسبی.

۱۴) وزن‌های ۴۰ بسته پسته تا نزدیک‌ترین کیلو عبارتند از:

۱۵۸	۱۴۶	۱۵۷	۱۴۹	۱۲۵	۱۴۴	۱۳۲
۱۷۶	۱۳۸	۱۲۶	۱۶۸	۱۴۴	۱۵۲	۱۴۸
۱۵۳	۱۳۵	۱۴۷	۱۴۲	۱۷۳	۱۴۶	۱۶۵
۱۲۸	۱۴۵	۱۵۶	۱۵۰	۱۴۲	۱۴۵	۱۳۵
۱۶۱	۱۵۴	۱۱۹	۱۶۳	۱۵۰	۱۶۴	۱۳۸
		۱۳۵	۱۴۰	۱۳۶	۱۴۷	۱۴۰

یک جدول توزیع فراو نی کامل با هشت طبقه با فواصل مساوی پیدا کنید.

۱۵) میزان هموگلوبین خون در ۵۰ بیمار سرطانی برحسب گرم

در ۱۰۰ میلی‌لیتر عبارتند از:

۱۲٫۷	۱۴٫۲	۱۴٫۶	۱۳٫۸	۱۱٫۹	۱۱٫۵	۱۴٫۲	۱۳٫۷	۱۴٫۸	۱۳٫۶
۱۵٫۰	۱۳٫۵	۱۵٫۲	۱۵٫۳	۱۲٫۴	۱۲٫۷	۱۴٫۸	۱۱٫۹	۱۱٫۵	۱۳٫۴
۱۵٫۰	۱۳٫۵	۱۴٫۰	۱۵٫۵	۱۳٫۲	۱۰٫۰	۱۱٫۷	۱۳٫۸	۱۲٫۰	۱۲٫۴
۱۴٫۸	۱۴٫۰	۱۰٫۹	۱۵٫۷	۱۲٫۷	۱۳٫۵	۱۵٫۱	۱۳٫۷	۱۲٫۹	۱۲٫۷
۱۶٫۲	۱۳٫۲	۱۳٫۷	۱۱٫۱	۱۱٫۴	۱۲٫۰	۱۱٫۹	۱۲٫۷	۱۳٫۸	۱۴٫۰

با تشکیل ۷ طبقه با فاصله ۰٫۹، جدول فرونی کاملی را تشکیل دهید.

«یک تصویر خوب گویاتر از هزار کلام است»

نمودارها:

روش دوم برای تنظیم مشاهدات استفاده از نمودارهاست.

در واقع نمودارها نمایش داده‌ها بصورت اشکال هندسی و غیر هندسی است.

انواع نمودارها:

- نمودار نقطه‌ای
- نمودار تصویری
- نمودار خطی یا میله‌ای
- نمودار مستطیلی یا ستونی
- نمودار خط شکسته
- نمودار دایره‌ای(کلوچه‌ای)
- نمودار بافت نگار یا هیستوگرام
- نمودار چند ضلعی یا پلیگون
- نمودار توزیع
- نمودار فراوانی تجمعی
- نمودار اجایو
- منحنی اجایو (منحنیS)

نمودار نقطه‌ای و تصویری:

نمودار نقطه‌ای نموداری است که هر مقدار در مجموعه داده‌ها را با یک نقطه بر روی محور اعداد نمایش می‌دهد.

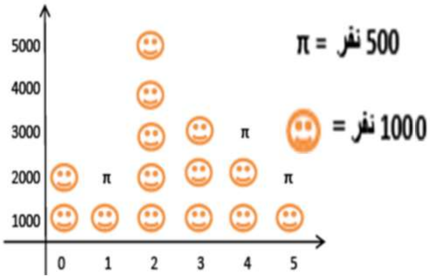
اگر مقدار مشخصی چندین بار تکرار شده باشد، آن نقطه‌ها روی یک خط قائم قرار می‌گیرند.

بدین ترتیب محور افقی به متغیر و محور عمودی به فراوانی اختصاص دارد.

این نمودار معمولاً برای داده‌های گسسته وکم رسم می‌شود.

مثال: تعداد افراد تحت تکفل پرسنل یک وزارتخانه به شرح زیر است نموار تصویری آن را رسم کنید.

x_i	0	1	2	3	4	5	Σ
f_i	2000	1500	5000	3000	2500	1500	15500



نمودار میله‌ای:

در نمودار میله‌ای برای نمایش فراوانی رده‌ها میله‌های قائم یا افقی استفاده می‌کنیم و آنها در مقابل داده ها رسم می‌شود.

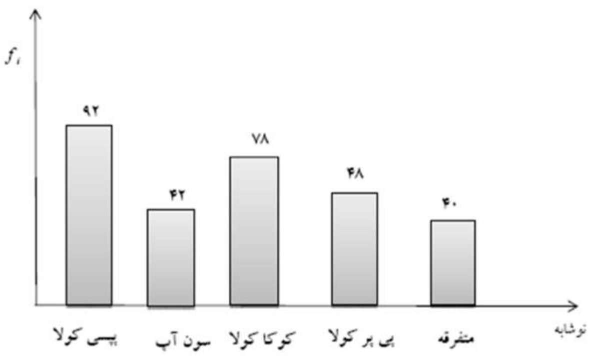
این نمودار برای داده‌های کم یا زیاد ولی گسسته کاربرد دارد.

بطور مشابه در نمودار ستونی یا مستطیلی فراوانی با مکعب مستطیل یا استوانه نشان داده می‌شود.

مثال: از یک نمونه ۳۰۰ تایی از دانش‌آموزان پیش دانشگاهی در مورد نوشابه به مورد علاقه‌شان سوال شده است نتایج نظر خواهی به شرح زیر است نمودار میله‌ای آن را نمایش دهید.

نوشابه	پپسی کولا	کوکا کولا	پی پر کولا	سون آپ	متفرقه
تعداد دانش‌آموزان	۹۲	۷۸	۴۸	۴۲	۴۰

حل:



نمودار خط شکسته و خط منحنی:

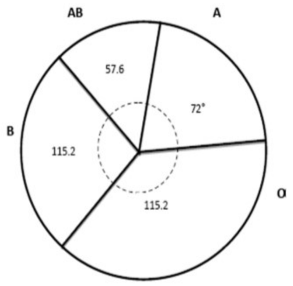
نموار خط شکسته برای نشان دادن رابطه بین دو متغیر مختلف به کار می‌رود. برای رسم آن مقادیر متناظر دو متغیر را بصورت نقاط ترسیم کرده سپس نقاط متوالی را به وسیله یک سری از خطوط مستقیم به هم وصل می کنیم.
یک نمودار خط شکسته را می‌توان از نمودار ستونی با وصل کردن نقاط وسط انتهای ستون‌های متوالی به وسیله خطوط مستقیم به دست آورد.
با اضافه شدن تعداد نقاط نمودار خط شکسته به نمودار خط منحنی نزدیک می‌شود. این نمودار معمولا برای نشان دادن روند داده‌ها استفاده می شود

نمودار دایره‌ای یا کلوچه‌ای:

این نمودار بیشتر در مورد داده‌های کمی یا کیفی به کار برده می‌شود که بصورت درصد بوده و تنوع زیادی ندارند.
معمولاً رده‌ها با رنگ‌ها و هاشورهای مختلف نمایش داده می‌شوند.

مثال: یک نمودار کلوچه‌ای برای داده‌های گروه خونی رسم کنید

گروه خونی	fi	درصد	زاویه
A	5	%20	72
B	8	%32	115/2
O	8	%32	115/2
AB	4	%16	57/6
Σ	25	%100	360°



مثال ۳۱. با ستفاده ز یک نمودار دایره‌ی نسبت کارکنان یک شرکت را که به ترتیب در جدول زیر رده‌بندی شده ست را نشان دهید.

رده	نسبت
مدیریت	۰٫۰۸
دفتری	۰٫۲۲
فروشندگان	۰٫۲۸
خدمات	۰٫۴۲

زاویه قطاع مربوط به رده مدیریت $= \frac{۸}{۱۰۰} \times ۳۶۰ = ۲۸٫۸$

زاویه قطاع مربوط به رده دفتری $= \frac{۲۲}{۱۰۰} \times ۳۶۰ = ۷۹٫۲$

زاویه قطاع مربوط به رده فروشندگان $= \frac{۲۸}{۱۰۰} \times ۳۶۰ = ۱۰۰٫۸$

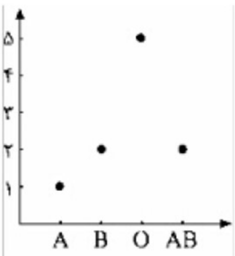
زاویه قطاع مربوط به رده خدمات $= \frac{۴۲}{۱۰۰} \times ۳۶۰ = ۱۵۱٫۲$

مثال ۲۷. فرض کنید گروه خونی ۱۰ دانشجو تعیین و نتایج زیر به دست آمده باشد.

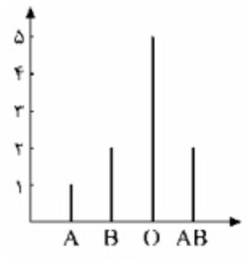
O , O , A , AB , B , B , O , O , O , AB

زاویه هر قطاع $= x_i = \frac{۳۶۰}{\text{کل فراوانی}} \times$ (فراوانی هر دسته)

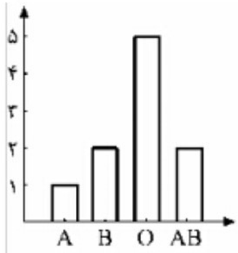
گروه خونی	f_i	x_i
A	۱	36°
B	۲	72°
O	۵	180°
AB	۲	72°
Σ	۱۰	



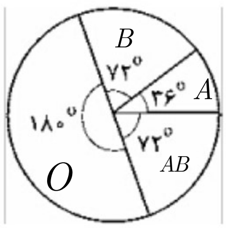
نمودار سوزنی



نمودار میله‌ای



نمودار ستونی



نمودار دایره‌ای

مثال

جدول توزیع فراوانی مهارت ۲۲ کارگر کارخانه A، به صورت زیر داده شده،

E_i	فراوانی
بد	۵
متوسط	۴
خوب	۶
عالی	۷
	۲۲

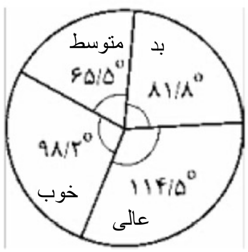
نمودارهای زیر را رسم کنید.

الف. نمودار سوزنی

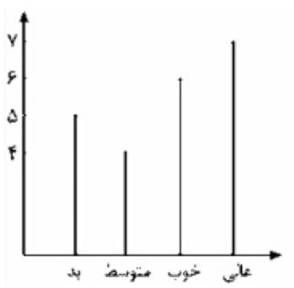
ب. نمودار میله‌ای

پ. نمودار دایره ی

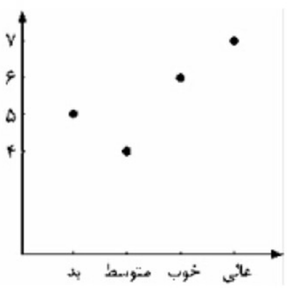
حل:



پ. نمودار دایره ی



ب. نمودار میله‌ای



الف. نمودار سوزنی

۲۲ ۶
۳۶۰ x $\Rightarrow x = \frac{۳۶۰}{۲۲} \times ۶ = ۹۸/۲$

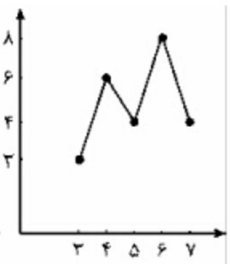
۲۲ ۷
۳۶۰ x $\Rightarrow x = \frac{۳۶۰}{۲۲} \times ۷ = ۱۱۴/۵$

۲۲ ۵
۳۶۰ x $\Rightarrow x = \frac{۳۶۰}{۲۲} \times ۵ = ۸۱/۸$

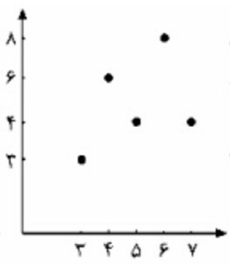
۲۲ ۴
۳۶۰ x $\Rightarrow x = \frac{۳۶۰}{۲۲} \times ۴ = ۶۵/۵$

مثال ۲۹. توزیع فراونی فراد خانوار به صورت زیر مفروض است:

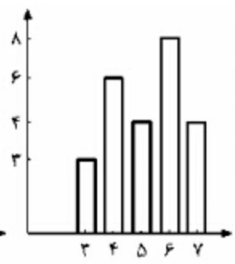
x_i	f_i
۳	۳
۴	۶
۵	۴
۶	۸
۷	۴
$\sum$	۲۵



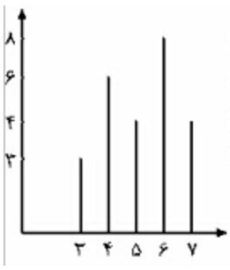
چند ضلعی فراوانی



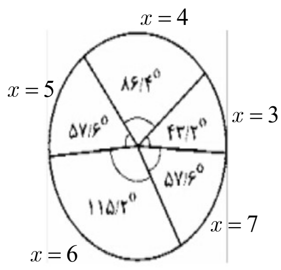
نمودار سوزنی



نمودار ستونی



نمودار میله‌ای



نمودار دایره ی

۲۵   ۳ ، ۳۶۰ x , $x = \frac{۳۶۰}{۲۵} \times ۳ = ۴۳/۲$

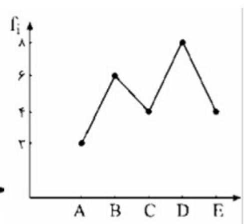
۲۵   ۴ ، ۳۶۰ x , $x = \frac{۳۶۰}{۲۵} \times ۴ = ۵۷/۶$

۲۵   ۶ ، ۳۶۰ x , $x = \frac{۳۶۰}{۲۵} \times ۶ = ۸۶/۴$

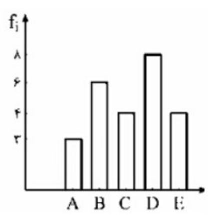
۲۵   ۸ ، ۳۶۰ x , $x = \frac{۳۶۰}{۲۵} \times ۸ = ۱۱۵/۲$

برای جدول زیر نمود رهای سوزنی، میله‌ای، ستونی، چندضلعی فراوانی و د یره‌ای را رسم کنید.

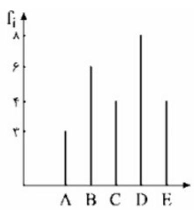
x_i	f_i
A	۳
B	۶
C	۴
D	۸
E	۴
$\sum$	۲۵



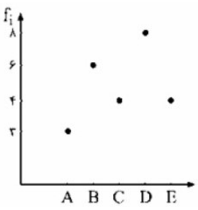
چند ضلعی فراوانی



نمودار ستونی



نمودار میله‌ای



نمودار سوزنی

$$x_i = \frac{\text{فراوانی مطلق}}{\text{کل فروانی}} \times ۳۶۰$$

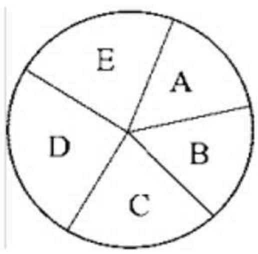
$$x_A = \frac{۳}{۲۵} \times ۳۶۰ = ۴۳/۲$$

$$x_B = \frac{۶}{۲۵} \times ۳۶۰ = ۸۶/۴$$

$$x_C = \frac{۴}{۲۵} \times ۳۶۰ = ۵۷/۶$$

$$x_D = \frac{۸}{۲۵} \times ۳۶۰ = ۱۱۵/۲$$

$$x_E = \frac{۴}{۲۵} \times ۳۶۰ = ۵۷/۶$$



نمودار دایره ی

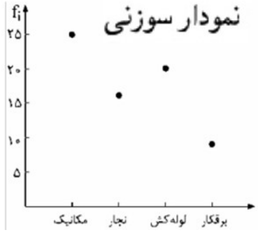
توزیع فراوانی کارگرن یک کارخانه بر حسب نوع فعالیت به صورت زیر است: نمودارهای مربوطه ر رسم نمایید.

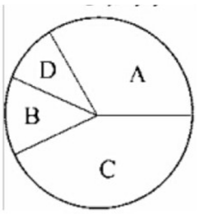
نوع فعالیت	فراوانی
مکانیک	۲۵
نجار	۱۶
لوله‌کش	۲۰
برق‌کار	۹
$\sum$	۷۰

حل:

نمودار میله‌ای

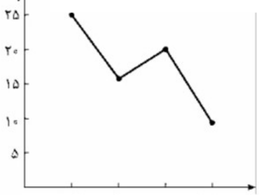
نمودار سوزنی



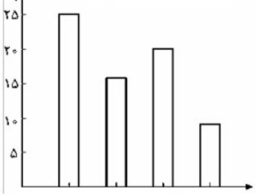


نمودار دایره ی

چند ضلعی فراوانی



نمودار ستونی



زاویه قطاع مربوط به مکانیک $x_A = \frac{۲۵}{۷۰} \times ۳۶۰ = ۱۲۸/۶$

زاویه قطاع مربوط به نجار $x_B = \frac{۱۶}{۷۰} \times ۳۶۰ = ۸۲/۳$

زاویه قطاع مربوط به لوله‌کش $x_C = \frac{۲۰}{۷۰} \times ۳۶۰ = ۱۰۲/۸$

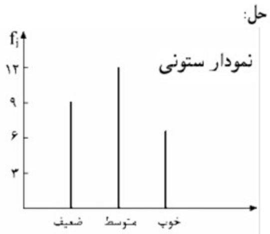
زاویه قطاع مربوط به برق‌کار $x_D = \frac{۹}{۷۰} \times ۳۶۰ = ۴۶/۳$

مثال ۳۳. در جدول زیر نظر ۲۸ مشتری کتابفروشی پیمان را در خصوص کیفیت کتاب‌های عرضه شده مشاهده می‌کنید. (مشتریان به صورت تصادفی انتخاب شده‌اند). نمودارهای مربوطه را رسم نمایید.

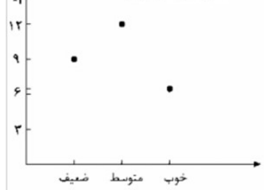
کیفیت	تعداد
ضعیف	۹
متوسط	۱۲
خوب	۷
$\sum$	۲۸

حل:

نمودار ستونی



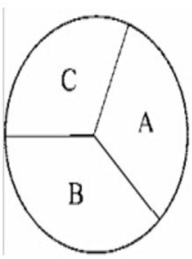
نمودار سوزنی



زاویه قطاع مربوط به ضعیف $x_A = \frac{۹}{۲۸} \times ۳۶۰ = ۱۱۵/۶$

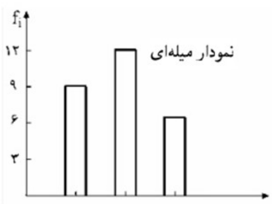
زاویه قطاع مربوط به متوسط $x_B = \frac{۱۲}{۲۸} \times ۳۶۰ = ۱۵۴/۴$

زاویه قطاع مربوط به خوب $x_C = \frac{۷}{۲۸} \times ۳۶۰ = ۹۰$

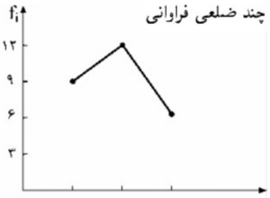


نمودار دایره ی

نمودار میله‌ای



چند ضلعی فراوانی



هیستوگرام (بافت نگار) :

نموداری است مرکب از چند مستطیل که از روی جدول توزیع فراوانی دسته‌بندی شده ساخته می‌شود.

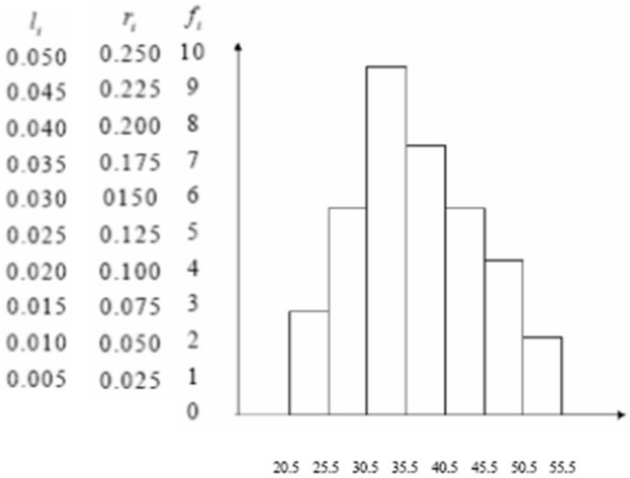
قاعده هر مستطیل روی محور xها جا دارد و طولش با طول رده برابر است و مرکز آن نماینده رده است.

ارتفاع هر مستطیل میتواند ○ فراوانی (f_i) ○ فراوانی نسبی (r_i) ○ رده‌ها $\frac{r_i}{I}$ باشد.

نکته:

- در صورتیکه ارتفاع مستطیل‌ها فراوانی (f_i) باشد، آنگاه $\sum$ ارتفاع مستطیل ها $= n$
- در صورتیکه ارتفاع فراوانی نسبی (r_i) باشد، آنگاه $\sum$ ارتفاع مستطیل ها $= \sum_{i=1}^{n} r_i = 1$
- در صورتیکه l_i ها باشد، آنگاه $\sum$ مساحت مستطیل ها $= \sum_{i=1}^{n} l_i \times I_i = \sum_{i=1}^{n} r_i = 1$
- در صورتیکه ارتفاع مستطیل‌ها $\frac{r_i}{I}$ باشد، مساحت تمام مستطیل‌های یک هیستوگرام برابر یک واحد مربع می‌باشد.
- اگر رده‌ها در جدول فراوانی دارای طول‌های مختلف باشند، قاعده‌های مستطیل برابر نخواهد بود. در این موارد باید ارتفاع مستطیل‌ها را $\frac{r_i}{I_i}$ اختیار کرد که مساحت تمام مستطیل برابر یک واحد مربع شود.

مثال: : برای مثال ۱ هیستوگرام را رسم کنید



i	رده ها	x_i	f_i	f_{ci}	r_i	r_{ci}	$l_i = \frac{r_i}{I_i}$
1	20.5-25.5	23	3	3	0.075	0.075	۰.۰۱۵
2	25.5-30.5	28	6	9	0.15	0.225	۰.۰۳۰
3	30.5-35.5	33	10	۱۹	0.25	0.475	۰.۰۵۰
4	35.5-40.5	38	8	27	0.20	0.675	۰.۰4۰
5	40.5-45.5	43	6	33	0.15	0.825	۰.۰۳۰
6	45.5-50.5	48	5	38	0.125	0.95	۰.۰۲۵
7	50.5-55.5	53	2	40	0.05	1	۰.۰۱
$\sum$	---	---	40	---	1	---	---

نمودار پلی گون (polygon) یا چند ضلعی:

این نمودار روی بافت نگار رسم می‌شود.

نحوه رسم بدین ترتیب است که وسط عرض باید مستطیل‌ها را مشخص کنیم

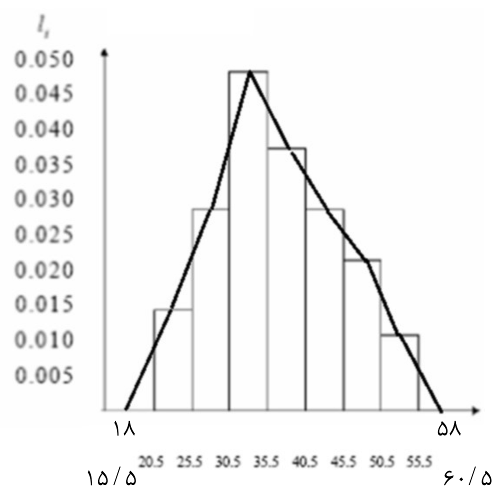
و به اندازه نصف طول رده روی محور Ox قبل و بعد از مستطیل ها معیین میکنیم

از به هم وصل کردن نقاط حاصل نمودار چند ضلعی یا پلی گون به دست می‌آید.

توجه داریم که :

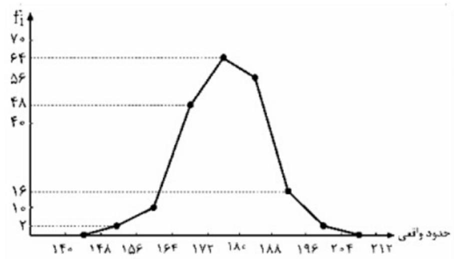
$$\sum_{i=1}^{n} l_i \times I_i = \sum_{i=1}^{n} r_i = 1$$ = مساحت زیر نمودار چند ضلعی

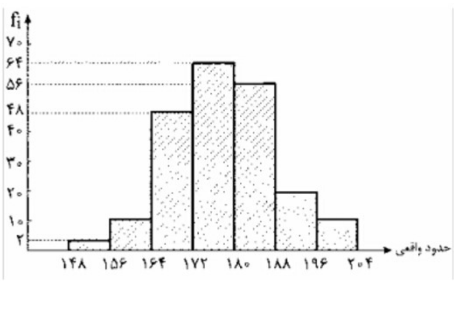
مثال: : برای مثال ۱ نمودار پلی گون را رسم کنید



نمود رهای هیستوگرم و چندضلعی فراو نی (پلی‌گن) مربوط به جدول توزیع فراوانی زیر را رسم نمایید.

فرونی	حدود طبقات
۲	۱۴۸ – ۱۵۶
۱۰	۱۵۶ – ۱۶۴
۴۸	۱۶۴ – ۱۷۲
۶۴	۱۷۲ – ۱۸۰
۵۶	۱۸۰ – ۱۸۸
۱۶	۱۸۸ – ۱۹۶
۴	۱۹۶ – ۲۰۴





f_{p_i}	f_i	x_i	حدود طبقات
۰/۰۱	۲	۱۵۲	۱۴۸ – ۱۵۶
۰/۰۵	۱۰	۱۶۰	۱۵۶ – ۱۶۴
۰/۲۴	۴۸	۱۶۸	۱۶۴ – ۱۷۲
۰/۳۲	۶۴	۱۷۶	۱۷۲ – ۱۸۰
۰/۲۸	۵۶	۱۸۴	۱۸۰ – ۱۸۸
۰/۰۸	۱۶	۱۹۲	۱۸۸ – ۱۹۶
۰/۰۲	۴	۲۰۰	۱۹۶ – ۲۰۴
۱	۲۰۰		

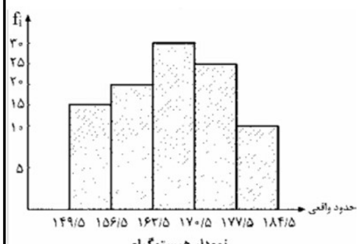
جدول توزیع فراوانی زیر ر در نظر بگیرید:

حدود طبقات	فراوانی مطلق
۱۴۹/۵ – ۱۵۶/۵	۱۵
۱۵۶/۵ – ۱۶۳/۵	۲۰
۱۶۳/۵ – ۱۷۰/۵	۳۰
۱۷۰/۵ – ۱۷۷/۵	۲۵
۱۷۷/۵ – ۱۸۴/۵	۱۰

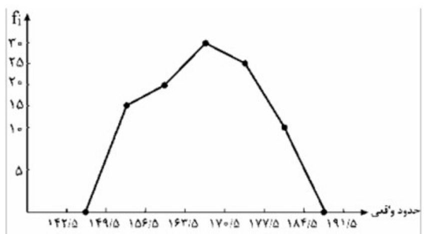
نمودارهای هیستوگرام و چندضلعی فراوانی (پلی‌گن) را رسم نمایید.

حل:

حدود طبقات	x_i	f_i
۱۴۹/۵ – ۱۵۶/۵	۱۵۳	۱۵
۱۵۶/۵ – ۱۶۳/۵	۱۶۰	۲۰
۱۶۳/۵ – ۱۷۰/۵	۱۶۷	۳۰
۱۷۰/۵ – ۱۷۷/۵	۱۷۴	۲۵
۱۷۷/۵ – ۱۸۴/۵	۱۸۱	۱۰
		۱۰۰



نمودار هیستوگرام



نمودار پلی‌گن

مثال ۳۹. حقوق روزنه ۲۵ نفر بر حسب هزار تومان در جدول زیر د ده شده است:

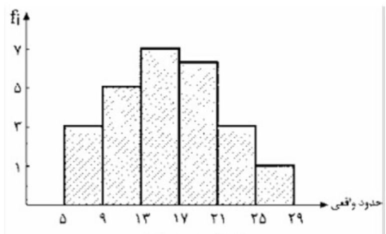
حدود طبقات	۵ – ۹	۹ – ۱۳	۱۳ – ۱۷	۱۷ – ۲۱	۲۱ – ۲۵	۲۵ – ۲۹
تعداد	۳	۵	۷	۶	۳	۱

الف. جدول توزیع فراوانی را تشکیل دهید.

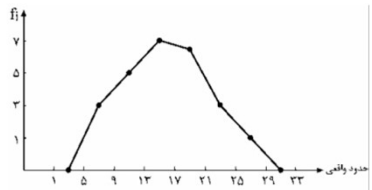
ب. نمود رهای هیستوگرام و پلی‌گن را رسم نمایید.

$$I = \frac{R}{K}$$

حدود طبقات	x_i	f_i	f_{p_i}	P_i	F_{C_i}	F_{CP_i}	P_C
۵ – ۹	۷	۳	۰/۱۲	۱۲	۳	۰/۱۲	۱۲
۹ – ۱۳	۱۱	۵	۰/۲۰	۲۰	۸	۰/۳۲	۳۲
۱۳ – ۱۷	۱۵	۷	۰/۲۸	۲۸	۱۵	۰/۶۰	۶۰
۱۷ – ۲۱	۱۹	۶	۰/۲۴	۲۴	۲۱	۰/۸۴	۸۴
۲۱ – ۲۵	۲۳	۳	۰/۱۲	۱۲	۲۴	۰/۹۶	۹۶
۲۵ – ۲۹	۲۷	۱	۰/۰۴	۴	۲۵	۱	۱۰۰
		۲۵	۱	۱۰۰			



نمودار هیستوگرام



نمودار پلی‌گن

مثال ۳۵. نمودار چند ضلعی درصد فروانی را برای جدول توزیع فروانی زیر رسم کنید.

	حدود طبقات	f_i
۸/۵–۱۱/۵	۹ – ۱۱	۴
۱۱/۵–۱۴/۵	۱۲ – ۱۴	۲
۱۴/۵–۱۷/۵	۱۵ – ۱۷	۶
۱۷/۵–۲۰/۵	۱۸ – ۲۰	۳
۲۰/۵–۲۳/۵	۲۱ – ۲۳	۵
	$\sum$	۲۰

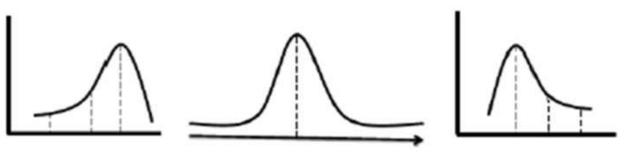
داده ها را به صورت پیوسته رده بندی می کنیم

حدود طبقات	f_i	x_i	f_{pi}
۹ – ۱۱	۴	۱۰	۰٫۲
۱۲ – ۱۴	۲	۱۳	۰٫۱
۱۵ – ۱۷	۶	۱۶	۰٫۳
۱۸ – ۲۰	۳	۱۹	۰٫۱۵
۲۱ – ۲۳	۵	۲۲	۰٫۲۵

نمودار توزیع (توزیع فراوانی):

اگر فاصله دسته‌ها در نمودار پلی گون کوچک و کوچکتر شود در حالی که تعداد داده ها بیشتر و بیشتر شود نمودار پلی گون به یک منحنی نزدیک می‌شود که آن منحنی را توزیع می‌نامیم.

این منحنی به طور تقریب با برازش آن یک منحنی مشتق پذیر به نقاط کثیر الاضلاع به دست می‌آید.

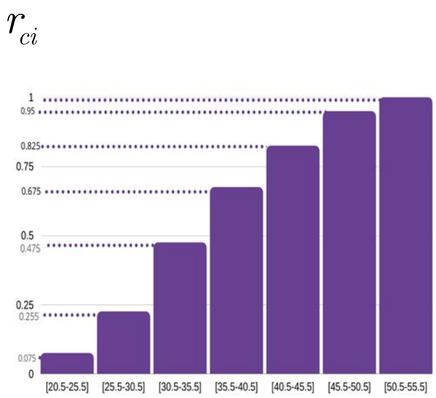


۱= مساحت زیر منحنی با محور افقی

نمودار فراوانی نسبی تجمعی یا فراوانی تجمعی: در این نمودار فراوانی نسبی تجمعی یا فراوانی تجمعی در مقابل رده‌ها رسم می‌شود و بصورت مستطیل‌ها نشان داده می‌شود.

مثال: نمودار فراوانی نسبی تجمعی را برای داده‌های برای مثال ۱ رسم کنید. (داده های کره)

i	رده ها	x_i	f_i	f_{ci}		r_{ci}		
1	20.5-25.5	23	3	3		0.075		
2	25.5-30.5	28	6	9		0.225		
3	30.5-355	33	10	19		0.475		
4	35.5-40.5	38	8	27		0.675		
5	40.5-45.5	43	6	33		0.825		
6	15.5-505	48	5	38		0.95		
7	50.5-55.5	53	2	40		1		
Σ	---	---	40	---		---		



$$r_{ci} = \frac{f_{ci}}{n}$$

نمودار اجایو:

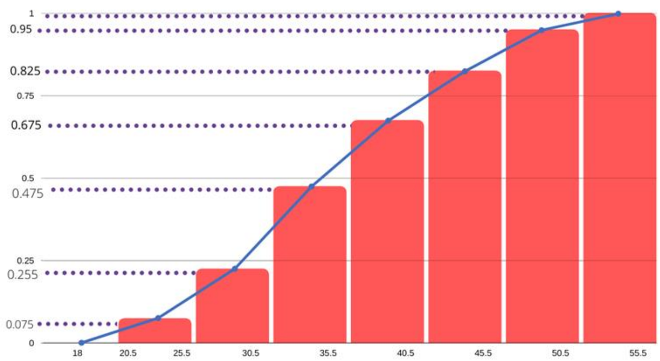
این نمودار می‌تواند به دو صورت از نمودار فراوانی نسبی تجمعی به دست آید.

الف) مرز بالایی رده‌ها را بهم وصل کنیم.

ب)وسط رده‌ها را بهم وصل کرده و به اندازه نصف رده اول عقب می‌رویم.

مثال: نمودار اجایو را برای داده‌های برای مثال ۱ رسم کنید.

i	رده ها	x_i	f_i	f_{ci}		r_{ci}
1	20.5-25.5	23	3	3		0.075
2	25.5-30.5	28	6	9		0.225
3	30.5-355	33	10	19		0.475
4	35.5-40.5	38	8	27		0.675
5	40.5-45.5	43	6	33		0.825
6	15.5-505	48	5	38		0.95
7	50.5-55.5	53	2	40		1
Σ	---	---	40	---		---



منحنی اجایو: اگر تعداد رده‌ها زیاد شده و عرض رده‌ها کوچک شود نمودار اجایو به منحنی اجایو همگرا می‌شود. منحنی S

مثال ۴۴. طول عمر ۱۰۰ لامپ ۶۰ وت، بر حسب ساعت در جدول زیر داده شده است:

طول عمر	فراوانی
۳۰۰ — ۳۲۰	۱۰
۳۲۰ — ۳۴۰	۲۰
۳۴۰ — ۳۶۰	۳۸
۳۶۰ — ۳۸۰	۲۵
۳۸۰ — ۴۰۰	۷
	$\Sigma =$ ۱۰۰

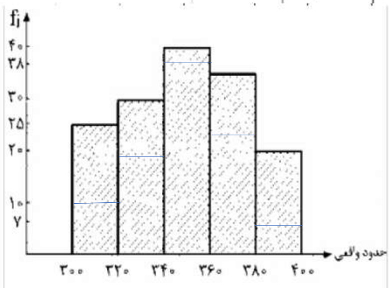
الف. جدول توزیع فراونی را تشکیل دهید.

ب. نمودهای هیستوگرام و چندضلعی فراوانی (پلی‌گن) را رسم کنید.

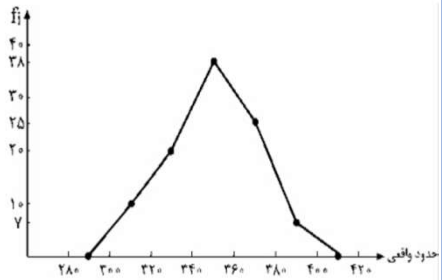
ج. نمودار فروانی تجمعی را رسم نمایید.

حل: الف. جدول توزیع فراوانی

حدود طبقات	x_i	f_i	r_i f_{P_i}	P_i	f_{ci} F_{C_i}	r_{ci} F_{CP_i}	P_C
۳۰۰ — ۳۲۰	۳۱۰	۱۰	٪۱۰	۱۰	۱۰	۰٫۱۰	۱۰
۳۲۰ — ۳۴۰	۳۳۰	۲۰	۰٫۲۰	۲۰	۳۰	۰٫۳۰	۳۰
۳۴۰ — ۳۶۰	۳۵۰	۳۸	۰٫۳۸	۳۸	۶۸	۰٫۶۸	۶۸
۳۶۰ — ۳۸۰	۳۷۰	۲۵	۰٫۲۵	۲۵	۹۳	۰٫۹۳	۹۳
۳۸۰ — ۴۰۰	۳۹۰	۷	۰٫۰۷	۷	۱۰۰	۱	۱۰۰
		۱۰۰	۱				



نمودار هیستوگرام



نمودار پلی‌گن

مثال ۴۵. داده‌های زیر را در نظر بگیرید:

۱۷	۳۵	۳۰	۱۰	۱۴	۳
۲۱	۲۸	۲۴	۸	۶	۷
۴۵	۵۷	۱۹	۱۱	۳۵	۱۸
۳۲	۵۹	۱۷	۳۳	۱۲	۱۴

الف. جدول توزیع فراونی را در ۶ طبقه تنظیم نمایید.

$L = 0,\ I = 10$

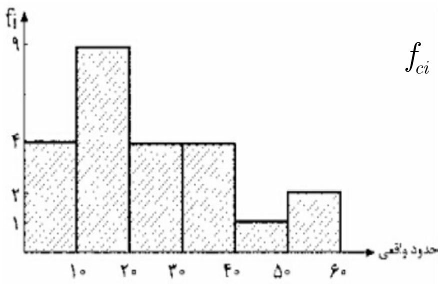
ب. نمودهای هیستوگرام و فراونی تجمعی را رسم کنید.

حل: الف.

فاصله طبقات = (کوچکترین داده − بزرگترین داده) / تعداد طبقات

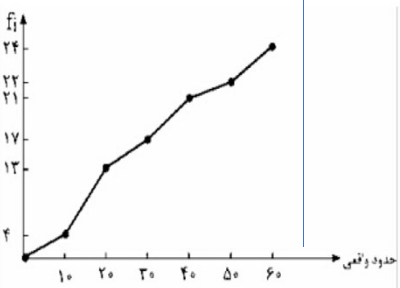
حدود طبقات	x_i	f_i	r_i f_{P_i}	P_i	f_{ci} F_{C_i}	F_{CP_i}	P_C
۰ — ۱۰	۵	۴	۰٫۱۷	۱۷	۴	۰٫۱۷	۱۷
۱۰ — ۲۰	۱۵	۹	۰٫۳۷	۳۷	۱۳	۰٫۵۴	۵۴
۲۰ — ۳۰	۲۵	۴	۰٫۱۷	۱۷	۱۷	۰٫۷۱	۷۱
۳۰ — ۴۰	۳۵	۴	۰٫۱۷	۱۷	۲۱	۰٫۸۸	۸۸
۴۰ — ۵۰	۴۵	۱	۰٫۰۴	۴	۲۲	۰٫۹۲	۹۲
۵۰ — ۶۰	۵۵	۲	۰٫۰۸	۸	۲۴	۱	۱۰۰
Σ		۲۴	۱	۱۰۰			

جدول توزیع فراونی



نمودار هیستوگرام

f_{ci}



نمودار فراوانی تجمعی

نمودار ساقه و برگ: برای توضیح این نمودار به مثال زیر توجه کنید:

مثال: در یک تست انگلیسی ۸۰ سوالی از دانشجویان نتایج زیر بدست

نمودار ساقه برگ آن را رسم کنید. کلید نمودار ۵۲=۲ ۵ است.

25	30	40	61	71	30	40	31	50	31
42	37	50	52	35	35	50	61	40	44
51	45	44	44	52	68	52	37	45	54
40	32	41	50	67	32	32	47	57	47
50	45	56	46	38	56	39	47	57	49
64	66	38	56	46	46	56	49		

حل: ابتدا داده ها را بصورت صعودی می‌نویسیم

25	30	30	31	31	32	32	32	35	35
37	37	38	38	39	40	40	40	40	41
42	44	44	44	45	45	46	46	46	47
47	47	49	49	50	50	50	50	50	51
52	52	52	54	56	56	56	56	57	57
61	61	64	67	71	78				

برای رسم نمودار ساقه و برگ برای این داده‌ها، توجه کنید

که تمام داده‌ها دو رقمی هستند

لذا برای تشکیل ساقه از رقم دهگان و برای تشکیل برگ‌ها

از رقم یکان استفاده می‌کنیم.

مثلاً عدد ۷۸ را بصورت ۸ ۷ می‌نویسیم که در آن ۷ روی

ساقه و ۸ روی برگ‌ها قرار می‌گیرد.

بدین ترتیب نمودار زیر را بدست می‌آید:

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

تجزیه و تحلیل مشاهدات :

تعریف: هر اندازه یا کمیتی که مرکزی از توزیع را نشان دهد، معیار تمایل به مرکز نامیده می‌شود.

تعریف: هر مقدار عددی که میزان پراکنش داده ها را حول یک نقطه مرکزی را نشان دهد، اندازه پراکندگی گوییم.

معیارهای گرایش به مرکز:

شاید مهم‌ترین نکته در مطالعه توزیع داده ها تعیین یک مقدار مرکزی باشد

یعنی یک مقدار نماینده که اندازه‌ها در اطراف آن توزیع شده‌اند.

هر معیار عددی را که معرف مرکز و مجموعه ای از داده‌ها باشد، معیار گرایش به مرکز می‌نامند.

متداول‌ترین معیارهای گرایش به مرکز عبارتند از: **میانگین – میانه – مد**

انواع میانگین‌ها:

- میانگین حسابی
- -میانگین هندسی
- میانگین هارمونیک
- میانگین موزون
- میانگین وینروری
- میانگین پیراسته

تعریف: برای یک مجموعه n تایی از اعداد $X_1,\dots,X_n$ میانگین حسابی با $\bar{X}$ نشان داده می‌شود و بصورت زیر تعریف شده است:

۱) داده‌ها به صورت اصلی و خام هستند

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

مثال: میانگین ۱۰ داده آماری برابر ۵ محاسبه شده است پس از محاسبه معلوم گردید که دو مقدار ۱۰ و ۱۲ نیز باید به داده ها اضافه شود. میانگین جدید را بدست آورید.

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \Rightarrow 5=\frac{1}{10}\sum_{i=1}^{10}X_i \Rightarrow \sum_{i=1}^{10}X_i=50 \Rightarrow \sum_{i=1}^{12}X_i=\sum_{i=1}^{10}X_i+X_{11}+X_{12}=50+10+12 \Rightarrow \sum_{i=1}^{12}X_i=72 \Rightarrow \bar{X}=\frac{1}{12}\sum_{i=1}^{12}X_i=\frac{72}{12}=6$$

۲) داده ها دارای تکرار هستند

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{k}f_iX_i$$

مثال: تعداد افراد تحت تکفل پرسنل یک دبیرستان به شرح زیر است میانگین آن را بدست آورید.

x_i	0	1	2	3	4	5	Σ
f_i	2	1	5	3	2	1	14
x_if_i	۰	۱	۱۰	۹	۸	۵	۳۳

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{k}f_iX_i=\frac{1}{14}(0\times2+1\times1+2\times5+3\times3+4\times2+5\times1)=\frac{33}{14}$$

۳) داده ها رده بندی شده هستند

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{k}f_i\hat{X}_i$$

$\hat{x}_i =$ نقطه وسط رده i ام

$$=\frac{\text{مرز پایینی رده } i\text{ ام} + \text{مرز بالایی رده } i\text{ ام}}{2}$$

مثال ۴. جدول مقابل، جدول توزیع فراوانی نمرات درس فیزیک ۵۰ دانشجو را نشان می‌دهد. میانگین را محاسبه کنید.

حدود طبقات	f_i
۳ – ۵	۹
۶ – ۸	۱۰
۹ – ۱۱	۹
۱۲ – ۱۴	۱۰
۱۵ – ۱۷	۶
۱۸ – ۲۰	۶
	۵۰

حل: با توجه به فرمول (۲)، به دو ستون، یکی مربوط به x_i و دیگری مربوط به f_ix_i نیاز داریم لذا، ابتدا این دو ستون را تشکیل می‌دهیم

داده‌های رده‌بندی شده تعمیم یافته	حدود طبقات	f_i	x_i	f_ix_i
۲٫۵ – ۵٫۵	۳ – ۵	۹	۴	۳۶
۵٫۵ – ۸٫۵	۶ – ۸	۱۰	۷	۷۰
	۹ – ۱۱	۹	۱۰	۹۰
	۱۲ – ۱۴	۱۰	۱۳	۱۳۰
	۱۵ – ۱۷	۶	۱۶	۹۶
	۱۸ – ۲۰	۶	۱۹	۱۱۴
				۵۳۶

$$\bar{x}=\frac{۱}{n}\sum_{i=۱}^{۶}f_ix_i=\frac{۵۳۶}{۵۰}=۱۰٫۷۲$$

مثال ۵. جدول زیر، مربوط به توزیع فراوانی طول قد ۱۰۰ دانشجو می‌باشد. میانگین طول قد را محاسبه کنید.

حدود طبقات	f_i	x_i	f_ix_i
۱۵۳ – ۱۵۶٫۹	۲	۱۵۴٫۹۵	۳۰۹٫۹
۱۵۷ – ۱۶۰٫۹	۶	۱۵۸٫۹۵	۹۵۳٫۷
۱۶۱ – ۱۶۴٫۹	۲۵	۱۶۲٫۹۵	۴۰۷۳٫۵
۱۶۵ – ۱۶۸٫۹	۳۶	۱۶۶٫۹۵	۶۰۱۰٫۲
۱۶۹ – ۱۷۲٫۹	۲۳	۱۷۰٫۹۵	۳۹۳۱٫۸۵
۱۷۳ – ۱۷۶٫۹	۷	۱۷۴٫۹۵	۱۲۲۴٫۶۵
۱۷۷ – ۱۸۰٫۹	۱	۱۷۸٫۹۵	۱۷۸٫۹۵
	۱۰۰		۱۶۶۸۳٫۰۰

$$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{k} f_i x_i = \frac{۱۶۶۸۳}{۱۰۰} = ۱۶۶٫۸۳$$

خواص میانگین

میانگین مجموع دو متغیر

اگر مشاهده متغیر X، به صورت مجموع دو متغیر دیگر باشند یعنی داشته باشیم

$$Z_i = X_i + Y_i \ , \quad i = ۱, ۲, \ldots, n$$

در این صورت

$$\bar{Z} = \bar{X} + \bar{Y}$$

زیرا،

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)$$
$$= \frac{1}{n}\left(\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} Y_i$$
$$= \bar{X} + \bar{Y}$$

۱. مجموع انحرافات داده‌ها از میانگین برابر صفر است. زیرا، اگر داده‌ها طبقه‌بندی شده باشند

$$\sum_{i=۱}^{n}(x_i - \bar{x}) = \sum_{i=۱}^{n} x_i - n\bar{x} = \sum_{i=۱}^{n} x_i - n\frac{۱}{n}\sum_{i=۱}^{n} x_i = ۰$$

و اگر داده‌ها طبقه‌بندی شده باشند.

$$\sum f_i(x_i - \bar{x}) = \sum f_i x_i - \bar{x}\sum f_i$$
$$= \sum f_i x_i - \bar{x}n$$
$$= \sum f_i x_i - n\frac{۱}{n}\sum f_i x_i = ۰$$

مثال ۶. توزیع فراوانی صفت متغیر X به صورت زیر داده شده است:

x_i	f_i
۱	۵
۲	۱۵
۳	۲۰
۴	۳۰
۵	۴۰

مطلوب است:**الف.** محاسبه میانگین.

ب. نشان دهید که مجموع انحرافات از میانگین برابر صفر است.

حل: فرض کنید $d_i = x_i - \bar{x}$، نشان دهنده انحراف داده‌ها از میانگین باشد، جدول زیر را تشکیل می‌دهیم داریم:

x_i	f_i	$f_i x_i$	d_i	$f_i d_i$
۱	۵	۵	$-۲/۷۵$	$-۱۳/۷۵$
۲	۱۵	۳۰	$-۱/۷۵$	$-۲۶/۲۵$
۳	۲۰	۶۰	$-۰/۷۵$	-۱۵
۴	۲۰	۸۰	۰/۲۵	۵
۵	۴۰	۲۰۰	۱/۲۵	۵۰
Σ	۱۰۰	۳۷۵		۰ همیشه

$$\bar{x} = \frac{۱}{n}\sum f_i x_i$$

$$= \frac{۳۷۵}{۱۰۰} = ۳/۷۵$$

۲. اگر به تمام داده‌ها مقدار ثابت a را اضافه یا از تمام داده‌ها مقدار ثابت a را کم کنیم، میانگین به اندازه a اضافه و یا کم می‌شود.

$$y_i = x_i \pm a \Longrightarrow \sum y_i = \sum (x_i \pm a) = \sum x_i \pm \sum a = \sum x_i \pm na$$

با تقسیم طرفین بر n داریم

$$\bar{y} = \bar{x} \pm a$$

اگر داده‌ها گروه‌بندی شده باشند.

$$\bar{y} = \frac{۱}{n}\sum f_i(x_i \pm a) = \frac{۱}{n}\sum f_i x_i \pm \frac{۱}{n}\sum f_i a = \frac{۱}{n}\sum f_i x_i \pm \frac{a}{n}\sum f_i \quad \overset{\sum_{i=1}^{n} f_i = n}{\Longrightarrow}$$

$$\bar{y} = \frac{۱}{n}\sum f_i x_i \pm \frac{na}{n} = \bar{x} \pm a$$

بنابراین فرمول زیر را داریم

$$y_i = x_i \pm a \Longrightarrow \bar{y} = \bar{x} \pm a$$

مثال ۷. توزیع سن دانشجویان یک‌کلاس در جدول زیر داده شده، مطلوب است محاسبه میانگین.

x_i	f_i
۱۹	۱۰
۲۰	۱۰
۲۱	۲۰
۲۲	۴۱
۲۳	۱۰
۲۴	۹

حل: با انتخاب $a = ۲۲$ و متغیر $y_i = x_i - a$ داریم:

$$y_۱ = x_۱ - a = ۱۹ - ۲۲ = -۳$$
$$y_۲ = x_۲ - a = ۲۰ - ۲۲ = -۲$$
$$y_۳ = x_۳ - a = ۲۱ - ۲۲ = -۱$$
$$y_۴ = x_۴ - a = ۲۲ - ۲۲ = ۰$$
$$y_۵ = x_۵ - a = ۲۳ - ۲۲ = ۱$$
$$y_۶ = x_۶ - a = ۲۴ - ۲۲ = ۲$$

x_i	f_i	y_i	$f_i y_i$
۱۹	۱۰	-۳	-۳۰
۲۰	۱۰	-۲	-۲۰
۲۱	۲۰	-۱	-۲۰
۲۲	۴۱	۰	۰
۲۳	۱۰	۱	۱۰
۲۴	۹	۲	۱۸
	۱۰۰		-۴۲

$$\bar{y} = \frac{۱}{n}\sum f_i y_i = \frac{-۴۲}{۱۰۰} = -۰{,}۴۲$$

$$\bar{y} = \bar{x} - a \Longrightarrow \bar{x} = \bar{y} + a = -۰{,}۴۲ + ۲۲ = ۲۱{,}۵۸$$

توجه: a عددی دلخواه است ولی در حل مسائل ترجیح می‌دهیم a را یکی از x_iها انتخاب کنیم و توصیه می‌شود a را x_i وسطی انتخاب کنید.

۳. هرگاه تمام داده‌ها را در عدد ثابتی مانند b ضرب کنیم، میانگین در b ضرب می‌شود، زیرا:

$$y_i = bx_i \Longrightarrow \sum y_i = b\sum x_i$$

با تقسیم طرفین بر n داریم

$$\frac{۱}{n}\sum y_i = b\frac{۱}{n}\sum x_i \Longrightarrow \bar{y} = b\bar{x}$$

$$\left.\begin{array}{l} y_i = bx_i \Longrightarrow \bar{y} = b\bar{x} \\ \text{به طور کلی } y_i = bx_i + a \Longrightarrow \bar{y} = b\bar{x} + a \end{array}\right\}$$

فرمول مهم

یادداشت (روش میانگین فرضی):

روش کوتاه یا روش کدگذاری

و رده‌بندی‌ها پیوسته

اگر داده‌ها به صورت رده بندی شده باشند برای محاسبه میانگین، میتوان از رابطه زیر استفاده کرد.

$$\bar{X} = A + I\frac{1}{n}\sum_{i=1}^{k} f_i d_i$$

که در آن

$$d_i = \frac{\hat{X}_i - A}{I}$$

$$\hat{X}_i = Id_i + A$$

حد وسط دسته مد دار $= A$

تفاوت حد وسط دو دسته متوالی $I = ۱$

اثبات:

$$\left.\begin{array}{l} d_i = \dfrac{\hat{X}_i - A}{I} \Rightarrow \hat{X} = Id_i + A \\ \bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} f_i \hat{X}_i \end{array}\right\} \Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{k} f_i \left(Id_i + A\right)$$

$$= I\frac{1}{n}\sum_{i=1}^{k} f_i d_i + \frac{1}{n}A\sum_{i=1}^{k} f_i$$

$$\overset{\sum_{i=1}^{k} f_i = n}{\Rightarrow} \quad \bar{X} = A + I\frac{1}{n}\sum_{i=1}^{k} f_i d_i$$

مثال: : برای مثال ۱ میانگین داده های رده بندی شده را به صورت مستقیم و با استفاده از روش میانگین فرضی بدست آورید

(داده های کره)

i	رده ها	x_i	f_i	$x_i f_i$	d_i	$d_i f_i$
1	20.5-25.5	23	3	69	-2	-6
2	25.5-30.5	28	6	168	-1	-6
3	30.5-355	33 =A	10	330	0	0
4	35.5-40.5	38	8	304	1	8
5	40.5-45.5	43	6	258	2	12
6	15.5-505	48	5	240	3	15
7	50.5-55.5	53	2	106	4	8
$\sum$	---	---	40	1475	---	31

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} f_i \hat{X}_i$$

$$= \frac{1475}{40} = 36.875$$

$$d_i = \frac{\hat{X}_i - A}{I}$$

$$\bar{X} = A + I\frac{1}{n}\sum_{i=1}^{k} f_i d_i$$

$$= 33 + 5 \times \frac{1}{40} \times 31$$

$$= 36.865$$

مثال جدول توزیع فراوانی زیر، وزن ۲۰ کودک را نشان می‌دهد، میانگین وزن را با روش کدگذاری محاسبه کنید.

حدود طبقات	f_i	x_i
۹ − ۱۱	۴	۱۰
۱۲ − ۱۴	۲	۱۳
۱۵ − ۱۷	۶	۱۶
۱۸ − ۲۰	۳	۱۹
۲۱ − ۲۳	۵	۲۲
	۲۰	

حدود طبقات	f_i	x_i	u_i	f_iu_i
۸٫۵-۱۱٫۵ ۹ − ۱۱	۴	۱۰	−۲	−۸
11.5 – 14.5 ۱۲ − ۱۴	۲	۱۳	−۱	−۲
۱۵ − ۱۷	۶	$A=$ ۱۶	۰	۰
۱۸ − ۲۰	۳	۱۹	۱	۳
۲۱ − ۲۳	۵	۲۲	۲	۱۰
	۲۰			۳

$I=3$ $\qquad \bar{u} = \frac{۱}{n}\sum f_iu_i = \frac{۳}{۲۰}$

$$\overline{X} = A + I\frac{1}{n}\sum_{i=1}^{k} f_id_i = A + I\bar{u} = 16 + 3\times\frac{3}{20} = 16.45$$

مثال ۱۰. جدول توزیع فراوانی مربوط به وزن ۵۰ قطعه ساخته شده توسط یک کارخانه به صورت زیر بیان شده است. میانگین وزن را به دو روش مستقیم و غیرمستقیم محاسبه کنید.

حدود طبقات	f_i
۵۰ − ۵۹	۸
۶۰ − ۶۹	۱۰
۷۰ − ۷۹	۱۷
۸۰ − ۸۹	۱۳
۹۰ − ۹۹	۲
	۵۰

$$\bar{x} = \frac{۱}{n}\sum f_ix_i = \frac{۳۶۳۵}{۵۰} = ۷۲٫۷$$

حدود طبقات	f_i	x_i	f_ix_i	u_i	f_iu_i
۴۹٫۵-۵۹٫۵	۸	۵۴٫۵	۴۳۶	−۲	−۱۶
۶۰ − ۶۹	۱۰	۶۴٫۵	۶۴۵	−۱	−۱۰
۷۰ − ۷۹	۱۷	$A=$ ۷۴٫۵	۱۲۶۶٫۵	۰	۰
۸۰ − ۸۹	۱۳	۸۴٫۵	۱۰۹۸٫۵	۱	۱۳
۹۰ − ۹۹	۲	۹۴٫۵	۱۸۹	۲	۴
	۵۰		۳۶۳۵		−۹

$\bar{u} = \frac{۱}{n}\sum f_iu_i = \frac{-۹}{۵۰} = -۰٫۱۸$ $\qquad I=10$

$$\overline{X} = A + I\frac{1}{n}\sum_{i=1}^{k} f_id_i = A + I\bar{u} = 74.5 + 10\times(-0.18) = 72.7$$

میانگین هندسی: اگر $X_1,\ldots,X_n$ اعداد نا منفی باشند آنگاه میانگین هندسی آن‌ها که با M_G $(\overline{X}_G)$ نشان داده می‌شود برابر است با:

$$\overline{X}_G=\sqrt[n]{X_1X_1\ldots X_n},\qquad \ln \overline{X}_G=\frac{1}{n}\sum_{i=1}^{n}\ln X_i$$

یعنی، لگاریتم میانگین هندسی برابر است با میانگین حسابی لگاریتم داده‌ها.

مثال: مبلغ ۱۰۰۰ دلار با بهره ۴ درصد در اول ژانویه سال ۱۹۶۰ سرمایه‌گذاری شده است. اگر بهره در اول ژانویه هر سال به این مبلغ اضافه گردد با بکار بردن میانگین هندسی مقدار متوسط پول سرمایه‌گذاری شده از اول ژانویه سال ۱۹۶۰ تا ۳۱ دسامبر ۱۹۶۳ را حساب کنید.

سال	مقدار سرمایه
۱۹۶۰	۱۰۰۰/۰۰
۱۹۶۱	۱۰۴۰/۰۰
۱۹۶۲	۱۰۸۰/۶۱
۱۹۶۳	۱۱۲۴/۸۶

حل:

$$\overline{X}_G=\sqrt[n]{X_1X_1\ldots X_n}=\sqrt[4]{1000\times1040\times1081\times1124}=1004.5$$

مثال: میزان سود یک شرکت در پنج سال گذشته بر حسب درصد فروش به ترتیب 12 9 9 4 2 بوده است متوسط درصد فروش سالیانه را محاسبه کنید.

حل: چون مقادیر برحسب درصد بیان شده‌اند، لذا متوسط درصد فروش سالیانه بر حسب میانگین هندسی محاسبه می‌شود:

$$\overline{X}_G=\sqrt[n]{X_1X_1\ldots X_n}=\sqrt[5]{2\times4\times9\times9\times12}=6$$

تعریف: اگر یک جدول فراوانی با حد وسط دسته‌های غیر منفی بوده و با $X_1,\ldots,X_n$ نشان داده می‌شود و فراوانی دسته‌ها به ترتیب $f_1,\ldots,f_n$ باشد میانگین هندسی عبارت است از:

$$\overline{X}_G=\sqrt[n]{x_1^{f_1}\cdots x_k^{f_k}};\ n=\sum_{i=1}^{k}f_i$$

مثال ۱۲. اگر جدول توزیع فراوانی به صورت زیر بیان شده باشد، مطلوب است محاسبه

حدود طبقات	f_i
۱۰ − ۱۲	۳
۱۳ − ۱۵	۵
۱۶ − ۱۸	۲

حل:

	حدود طبقات	f_i	x_i
۹/۵−۱۲/۵	۱۰ − ۱۲	۳	۱۱
۱۲/۵−۱۵/۵	۱۳ − ۱۵	۵	۱۴
۱۵/۵−۱۸/۵	۱۶ − ۱۸	۲	۱۷
		۱۰	

$$G=\sqrt[۱۰]{۱۱^{۳}\times۱۴^{۵}\times۱۷^{۲}}$$

میانگین هارمونیک : برای یک مجموعه n تایی از اعداد $X_1,\ldots,X_n$ ، میانگین هارمونیک که با $\overline{X}_h$ نشان داده می‌شود برابراست با

$$\overline{X}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

مثال : هواپیمایی یک مسیر را در ثلث اول و سوم آن با سرعت ۲۵۰ مایل در ساعت و ثلث دوم را با سرعت ۳۰۰ مایل در ساعت طی می‌نماید، متوسط سرعت هواپیما چقدر است؟

حل:

$$\overline{X}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}} = \frac{3}{\frac{1}{250}+\frac{1}{300}+\frac{1}{250}} = 264.71$$

مثال ۲۲. جدول توزیع فراوانی مصرف برق ۱۰۰ خانوار، به صورت زیر داده شده، میانگین همساز را حساب کنید.

حدود طبقات	f_i	x_i
۶۰ − ۶۲	۵	۶۱
۶۳ − ۶۵	۱۸	۶۴
۶۶ − ۶۸	۴۲	۶۷
۶۹ − ۷۱	۲۷	۷۰
۷۲ − ۷۴	۸	۷۳

حل:

$$H = \frac{n}{\sum_{i=۱}^{k} f_i/x} = \frac{۱۰۰}{(\frac{۵}{۶۱}+\frac{۱۸}{۶۴}+\frac{۴۲}{۶۷}+\frac{۲۷}{۷۰}+\frac{۸}{۷۳})} \simeq ۶۸{,}۰۳$$

میانگین حسابی وزنی

	سنگینی	
نمره فیزیک	ضریب ۳	نمره ۱۲
نمره ریاضی	ضریب ۲	نمره ۱۵
نمره ادبیات	ضریب ۱	نمره ۱۴

$$\bar{x} = \frac{۳\times۱۲+۲\times۱۵+۱\times۱۴}{۳+۲+۱}$$

$$= \frac{۸۰}{۶} = ۱۳{,}۳$$

تعریف: به طور کلی اگر داده‌های $x_۱$، $x_۲$، ...، x_k به ترتیب دارای ضرایب وزنی $\omega_۱$، $\omega_۲$، ...، ω_k باشند، آن‌گاه

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

میانگین درجه دوم

میانگین درجه دوم داده‌های آماری $x_۱$، $x_۲$، ...، x_n عبارت است از:

$$Q=\left(\frac{۱}{n}\sum_{i=۱}^{n} x_i^{۲}\right)^{\frac{۱}{۲}}$$

و اگر داده‌ها گروه‌بندی شده باشند

$$Q=\left(\frac{۱}{n}\sum_{i=۱}^{k} f_i x_i^{۲}\right)^{\frac{۱}{۲}}$$

که در آن x_i نماینده و f_i فراوانی طبقه iام می‌باشند.

مثال ۲۳. میانگین درجه دوم داده‌های زیر را حساب کنید.

۱ ۲ ۳ ۴

حل:

$$Q=\left(\frac{۱}{۴}(۱^{۲}+۲^{۲}+۳^{۲}+۴^{۲})\right)^{\frac{۱}{۲}}$$
$$=\left(\frac{۳۰}{۴}\right)^{\frac{۱}{۲}}=\sqrt{۷/۵}$$

توجه: رابطه بین میانگین‌های بیان شده به صورت زیر می‌باشد:

$$H\leq G\leq \bar{x}\leq Q$$

چون در محاسبه میانگین‌های پیراسته و وینزوری،
باید از چارک‌های اول و سوم استفاده نمود،
لذا بحث این میانگین‌ها را بعد از معرفی چارک‌ها
بیان می‌کنیم.

میانه: اندازه عددی است که به عنوان یک معیار مرکزی به کار می‌رود به طوری که نصف داده‌ها کوچکتر یا مساوی آن و نصف داده‌ها بزرگتر یا مساوی آن باشند.

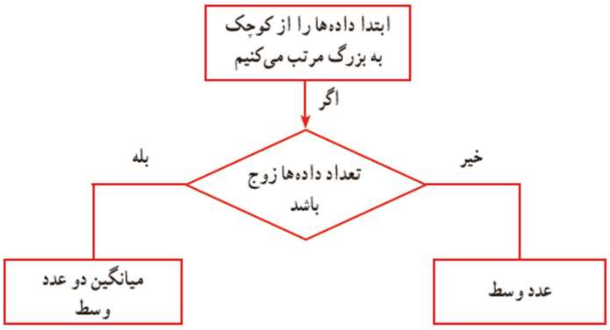
محاسبه میانه: اگر داده‌ها به صورت خام باشند، ابتدا داده‌ها را به صورت صعودی منظم می کنیم. دو حالت می تواند مطرح باشد:

حالت اول n تعداد داده‌ها فرد است در این صورت $\frac{n+1}{2}$ شماره عددی را مشخص می کند که اندازه میانه است

بعبارت دیگر میانه برابر است با $x_{\frac{n+1}{2}}$.

حالت دوم n زوج است در این صورت میانه برابر است با $\dfrac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$

نحوهٔ به دست آوردن میانه داده‌ها



مثال: میانه اعداد ۶ و ۷ و ۹ و ۱۲ و ۱۶ و ۲۰ را پیدا کنید.

مثال : میانه توزیع را به دست آورید.

$$M_d = \frac{x_3 + x_4}{2} = \frac{9+12}{2} = 10.5$$

مثال : میانه توزیع را به دست آورید.

x_i	۰	۱	۲	۳	۴	۵	Σ
f_i	۳	۵	۹	۴	۲	۱	۲۴
f_{ci}	۳	۸	۱۷	۲۱	۲۳	۲۴	--

$$\Rightarrow \frac{n}{2} = 12 \Rightarrow \frac{n}{2} + 1 = 13 \Rightarrow M_d = 2$$

مثال ۲۶. جدول توزیع فراوانی نمرات دانشجو، به صورت زیر داده شده، میانه را محاسبه کنید.

x_i	f_i
۱۲	۴
۱۴	۸
۱۵	۶
۱۷	۴

حل: ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم

f_{ci}

x_i	f_i	$F(x_i)$
۱۲	۴	۴
۱۴	۸	۱۲
۱۵	۶	۱۸
۱۷	۴	۲۲
۱۹	۳	۲۵
	۲۵	

$$k = \frac{n+۱}{۲} = \frac{۲۶}{۲} = ۱۳$$

$$\Rightarrow \quad M_d = ۱۵$$

مثال ۲۷. در جدول توزیع فراوانی زیرکه تعداد داده‌ها زوج می‌باشد، میانه را محاسبه کنید.

f_{ci}

x_i	f_i	$F(x_i)$
۱۴	۳	۳
۱۵	۶	۹
۱۷	۵	۱۴
۱۸	۲	۱۶
۱۹	۴	۲۰
	۲۰	

$$k = \frac{n}{۲} = ۱۰ \quad \Rightarrow \quad x_{10} = x_{11} = 17$$

$$\Rightarrow \quad Md = 17$$

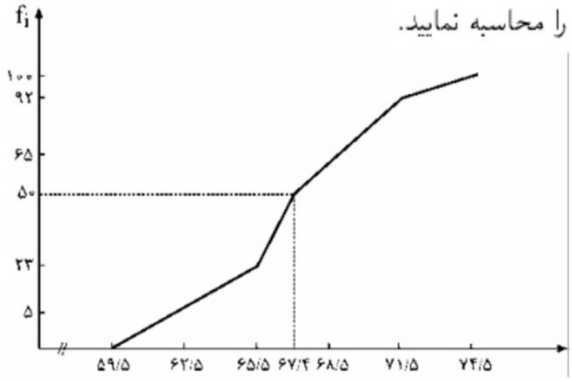
محاسبه میانه با استفاده از نمودار فراوانی تجمعی نمودار اجایو

برای محاسبه میانه می‌توان از روش ترسیمی استفاده نمود. بدین ترتیب که منحنی فراوانی تجمعی را رسم کرده و چون میانه، طول نقطه‌ای است که فراوانی تجمعی آن $\frac{n}{۲}$ است، کافی است روی محور فراوانی تجمعی، نقطه $\frac{n}{۲}$ را مشخص کرده و خطی موازی با محور xها رسم کنیم. سپس از این نقطه، خطی عمود بر محور xها رسم کرده پای عمود تا مبدأ، میانه می‌باشد.

F_{ci} r_{ci}

مثال ۳۰. برای مثال (۲۸) با استفاده از روش ترسیمی، میانه را محاسبه نمایید.

$\frac{n}{2}$



$$Md = 67.4$$

چارک‌ها: اندازه‌های عددی هستند که داده‌ها را به چهار قسمت مساوی تقسیم می‌کنند، بدین ترتیب

چارک اول Q_1 اندازه عددی است که $1/4$ از داده‌ها کوچکتر یا مساوی آن و $3/4$ از داده‌ها بزرگتر یا مساوی آن است.

چارک دوم Q_2 اندازه عددی است که $1/2$ از داده‌ها کوچکتر یا مساوی آن و $1/2$ از داده‌ها بزرگتر یا مساوی آن است.

چارک سوم Q_3 اندازه عددی است که $3/4$ از داده‌ها کوچکتر یا مساوی آن و $1/4$ از داده‌ها بزرگتر یا مساوی آن است

برای محاسبه میانه و چارک‌ها وقتی داده‌ها رده بندی شده باشند فرمول کلی ارائه می‌شود.

دهک‌ها: اندازه‌های عددی هستند که داده‌ها را به ده قسمت مساوی تقسیم می‌کنند و آن ها را با $D_9 \cdots D_2 D_1$ نشان می‌دهند. بدین ترتیب

D_1 یا دهک اول اندازه‌یست عددی که $1/10$ داده‌ها کوچکتر یا مساوی آن و $9/10$ از داده‌ها بزرگتر یا مساوی آن است

و... و دهک پنجم همان میانه یا چارک دوم است و...

و دهک نهم اندازه عددی است که $9/10$ از داده‌ها کوچکتر یا مساوی آن و $1/10$ از داده‌ها بزرگتر یا مساوی آن است.

برای محاسبه دهک‌ها برای تعداد زیاد و داده‌های رده بندی شده فرمول کلی ارائه می‌شود.

صدک ها: صدک ها اندازه‌های عددی هستند که داده‌ها را به ۱۰۰ قسمت مساوی تقسیم می‌کنند و آن‌ها را با نمادهای $H_{99} \cdots H_2 H_1$ نشان می‌دهند.

صدک اول اندازه عددی است که $1/100$ داده‌ها کوچکتر یا مساوی آن و $99/100$ داده‌ها بزرگتر یا مساوی آن است و...

و صدک دهم همان دهک اول

و صدک بیست و پنجم همان چارک اول

و صدک پنجاهم همان میانه یا چارک دوم یا دهک پنجم

و صدک هفتاد و پنجم همان چارک سوم می باشد. بدین ترتیب تمام چندک ها قابل تبدیل به صدک‌ها می‌باشند .

محاسبه صدک‌ها:

الف) حالتی که داده‌ها رده بندی شده باشند:

برای محاسبه صدک‌ها فرمول کلی

$$H_h = L + \frac{\frac{nh}{100} - f_{c(m-1)}}{f_m}(U - L) \qquad h = 1, \ldots, 99$$

را داریم که در آن

شماره رده شامل صدک h ام $= m$

مرز پایینی رده و بالایی و فراوانی رده m $= L, U, f_m$

صدک h ام $= H_h$

فراوانی تجمعی رده ماقبل رده mام $= f_{c(m-1)}$

تعداد کل داده‌ها $= n$

سوال: چگونه شماره رده شامل صدک h ام را پیدا کنیم؟

برای محاسبه صدک‌ها ابتدا فراوانی تجمعی رده‌ها و مقدار $\frac{nh}{100}$ را بدست می‌آوریم سپس $\frac{nh}{100}$ را با ستون فراوانی تجمعی از بالا به پائین مقایسه می‌کنیم. اولین جایی که فراوانی تجمعی از عدد $\frac{nh}{100}$ بزرگتر یا مساوی بود رده‌ای را مشخص می‌کند که صدک hام در آن قرار دارد.

مثال : داده‌های رده بندی شده زیر مفروضند. مطلوبست:

۱- صدک ۷۴ ام   ۲- میانه

۳- چارک اول   ۴- دهک سوم

i	رده‌ها	f_i	f_{ci}
۱	۰-۵	۳	۳
۲	۵-۱۰	۵	۸
۳	۱۰-۱۵	۹	۱۷
۴	۱۵-۲۰	۱۳	۳۰
۵	۲۰-۲۵	۵	۳۵
۶	۲۵-۳۰	۳	۳۸
۷	۳۰-۳۵	۲	٤۰

$$H_h = L + \frac{\frac{nh}{100} - f_{c(m-1)}}{f_m}(U - L)$$

محاسبه صدک ۷۴ام:

$H_{74} \Rightarrow h = 74$

$$\frac{nh}{100} = \frac{40 \times 74}{100} = 29.6 < 30 \Rightarrow m = 4, L = 15, U = 20, f_m = 13, f_{c(m-1)} = 17$$

$$\Rightarrow H_{74} = 15 + \frac{29.6 - 17}{13}(20 - 15) = 19.85$$

محاسبه میانه:

$m_d = H_{50} \Rightarrow h = 50$

$$\frac{nh}{100} = \frac{40 \times 50}{100} = 20 < 30 \Rightarrow m = 4, L = 15, U = 20, f_m = 13, f_{c(m-1)} = 17 \Rightarrow M_d = H_{50} = 15 + \frac{20 - 17}{13}(20 - 15) = 16.15$$

محاسبه چارک اول:

$Q_1 = H_{۲۵} \Rightarrow h = ۲۵$

$$\frac{nh}{100} = \frac{40 \times 25}{100} = 10 < 17 \Rightarrow m = 3, L = 10, U = 15, f_m = 9, f_{c(m-1)} = 8 \Rightarrow Q_1 = H_{25} = 10 + \frac{10 - 8}{9}(15 - 10) = 11.11$$

محاسبه دهک سوم:

$D_3 = H_{۳۰} \Rightarrow h = ۳۰$

$$\frac{nh}{100} = \frac{40 \times 30}{100} = 12 < 17 \Rightarrow m = 3, L = 10, U = 15, f_m = 9, f_{c(m-1)} = 8 \Rightarrow D_3 = H_{30} = 10 + \frac{12 - 8}{9}(15 - 10) = 12.22$$

مثال: جدول فراوانی زیر داده شده است میانه را بدست آورید

دسته	18 – 20	21 – 23	24 – 26	27 – 29	30 – 32	33 – 35	36 – 38	39 – 41	42 – 44	45 – 47	48 – 50
فراوانی	1	2	3	6	7	8	8	6	4	3	2

حل: تصحیح پیوستگی می کنیم و فراوانی تجمعی را بدست می آوریم

$$H_h = L + \frac{\frac{nh}{100} - f_{c(m-1)}}{f_m}(U - L)$$

h=۱,...,۹۹

دسته ها	17.5 20.5	20.5 23.5	23.5 26.5	26.5 29.5	29.5 32.5	32.5 35.5	35.5 38.5	38.5 41.5	41.5 44.5	44.5 47.5	47.5 50.5
فراوانی تجمعی	۱	۳	۶	۱۲	۱۹	۲۷	۳۵	۴۱	۴۵	۴۸	۵۰

$$m_d = H_{50},\quad \frac{nh}{100} = 25 < 27 \Rightarrow m = 6 \Rightarrow L = 32.5,\ f_{c(m-1)} = 19,\ f_m = 8,\ I = 3 \Rightarrow m_d = 32.5 + \frac{25-19}{8} \times 3 = 34.75$$

مثال ۴۰. جدول زیر، جدول توزیع فراوانی نمرات در دروس ریاضی ۵۰ دانشجو را نشان می‌دهد. مطلوب است محاسبه:

الف. چارک اول **ب.** چارک سوم **پ.** دهک هفتم **ت.** صدک ۶۸ام

حدود طبقات	f_i	F_c
۵ – ۷ ۴/۵–۷/۵	۸	۸
۸ – ۱۰ ۷/۵–۱۰/۵	۷	۱۵
۱۱ – ۱۳	۱۰	۲۵
۱۴ – ۱۶	۱۲	۳۷
۱۷ – ۱۹	۱۳	۵۰
	۵۰	

$$\frac{nh}{۱۰۰} = np = ۵۰ \times ۰/۲۵ = ۱۲/۵ \qquad Q_۱ = ۷/۵ + \frac{۱۲/۵ - ۸}{۷} \times ۳ = ۹/۴۳$$

$$p = \frac{h}{۱۰۰} \qquad np = ۵۰ \times ۰/۷۵ = ۳۷/۵ \qquad Q_۳ = ۱۶/۵ + \frac{۳۷/۵ - ۳۷}{۱۳} \times ۳ = ۱۶/۶۲$$

$$np = ۵۰ \times ۰/۷ = ۳۵ \qquad D_۷ = ۱۳/۵ + \frac{۳۵ - ۲۵}{۱۲} \times ۳ = ۱۶$$

$$np = ۵۰ \times ۰/۶۸ = ۳۴ \qquad H_{68} = ۱۳/۵ + \frac{۳۴ - ۲۵}{۱۲} \times ۳ = ۱۵/۷۵$$

مثال ۳۵. فرض کنید بین دو شهر که فاصله آن‌ها ۱۰۰ کیلومتر است، **۷** گاراژ وجود داشته باشد، در جدول زیر تعداد اتومبیل‌های هرگاراژ داده شده و فواصل گاراژها نیز مشخص شده است.

فواصل گاراژها	۰ — ۲۰	۲۰ — ۴۰	۴۰ — ۵۰	۵۰ — ۸۰	۸۰ — ۱۰۰
تعداد اتومبیل	۵۰	۱۰۰	۲۰	۴۰	۱۲۰

می‌خواهیم یک پمپ بنزین در این جاده احداث کنیم، محلی را تعیین کنید که جمع کل مسافت‌های پیموده شده برای بنزین‌گیری توسط ماشین‌ها، حداقل باشد.

حل: فرض کنید a محل احداث پمپ بنزین باشد، x_i محل گاراژ iام، مسافت پیموده شده توسط هر ماشین از گاراژ iام تا پمپ بنزین برابر است با $|x_i - a|$. بنابراین باید $\Sigma f_i|x_i - a|$ به حداقل برسد، بعد باید a برابر با میانه باشد، حال میانه جدول بالا را محاسبه می‌کنیم

حدود طبقات	f_i	$F(x_i)$
۰ — ۲۰	۵۰	۵۰
۲۰ — ۴۰	۱۰۰	۱۵۰
۴۰ — ۵۰	۲۰	۱۷۰
۵۰ — ۸۰	۴۰	۲۱۰
۸۰ — ۱۰۰	۱۲۰	۳۳۰

$$h = ۵۰ \Rightarrow \frac{nh}{۱۰۰} = \frac{۳۳۰ \times ۵۰}{۱۰۰}$$

$$= ۱۶۵ < ۱۷۰ \Rightarrow m = ۳$$

$$H_h = L + \frac{\frac{nh}{100} - f_{c(m-1)}}{f_m}(U - L)$$

$$Md = ۴۰ + \frac{۱۶۵ - ۱۵۰}{۲۰}(۵۰ - ۴۰)$$

$$= ۴۷/۵$$

ب) حالتی که داده ها خام هستند:

فرض کنید n داده را بصورت $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ مرتب کنیم

حال برای بدست آوردن صدک hام،

اگر $\frac{(n+1)h}{100}$ عدد صحیح، r باشد، صدک hام $H_h = x_{(r)}$ است

در غیر اینصورت جزء صحیح $\frac{(n+1)h}{100}$ را مساوی r و اختلاف آن را با r برابر w می‌گیریم H_h را از فرمول زیر حساب می‌کنیم.

$$H_h = (1-w)x_{(r)} + wx_{(r+1)}$$

مثال: ۹ کارگر صنعتی تحت آزمون قدرت پنجه قرار گرفتند و اندازه‌های زیر بدست آمده‌اند:

137.7, 105.8, 132.1, 125, 152.4, 116.6, 93.9, 106.5, 128.3

چارک‌ها را برای این داده‌ها محاسبه کنید.

حل: ابتدا داده ها را بصورت غیر نزولی

93.9, 105.8, 106.5, 116.6, 125, 128.3 132.1, 137.7, 152.4

مرتب می کنیم. برای محاسبه چارک اول

$$p = 0.25 \Rightarrow (n+1)p = 10 \times 0.25 = 2.5 \Rightarrow r = 2, w = 0.5$$

$$Q_1 = (1-0.5)x_{(2)} + 0.5x_{(3)} = 0.5 \times 105.8 + 0.5 \times 106.5 = 106.15$$

برای محاسبه چارک دوم و سوم

$p = 0.75$ و $p = 0.5$

قرار داده و داریم:

$$Q_2 : (n+1)p = 5 \Rightarrow r = 5, w = 0 \Rightarrow \quad Q_2 = md = x_{(5)} = 125,$$

$$Q_3 : (n+1)p = 7.5 \Rightarrow r = 7, w = 0.5 \Rightarrow$$

$$Q_3 = (1-0.5)x_{(7)} + 0.5x_{(8)} = 0.5 \times 132.1 + 0.5 \times 136.7 = 134.4$$

مثال ۳۸. داده‌های زیر مربوط به نمرات دروس یک دانش‌آموز می‌باشد **الف.** چارک اول **ب.** چارک دوم **پ.** چارک سوم **ت.** دهک سوم **ث.** صدک ۴۵ام را حساب کنید

۱۷ , ۱۶ , ۱۹ , ۱۳ , ۱۴ , ۱۸ , ۱۶ , ۱۵ , ۱۷

حل: ابتدا داده‌ها را به طور غیر نزولی مرتب می‌کنیم. ۱۹ , ۱۸ , ۱۷ , ۱۷ , ۱۶ , ۱۶ , ۱۵ , ۱۴ , ۱۳

$$H_h = (1-w)x_{(r)} + wx_{(r+1)}$$

الف. چارک اول

$(n+۱)p = (۹+۱) \times ۰/۲۵ = ۲/۵ = ۲ + ۰/۵ \quad Q_۱ = (۱ - ۰/۵)x_۲ + ۰/۵x_۳ = ۰/۵ \times ۱۴ + ۰/۵ \times ۱۵ = ۱۴/۵$

ب. چارک دوم

$(n+۱)p = (۹+۱) \times ۰/۵۰ = ۵ \quad Q_۲ = x_۵ = ۱۶$

پ. چارک سوم $Q_۳ = Q_{۰/۷۵}$

$(n+۱)p = (۹+۱) \times ۰/۷۵ = ۷/۵ = ۷ + ۰/۵ \quad Q_۳ = (۱ - ۰/۵)۱۷ + ۰/۵ \times ۱۸ = ۱۷/۵$

ت. $D_۳ = Q_{۰/۳۰}$

$(n+۱) \times p = (۹+۱) \times ۰/۳۰ = ۳ \quad D_۳ = x_۳ = ۱۵$

ث. $P_{۴۵} = Q_{۰/۴۵}$

$(n+۱)p = (۹+۱) \times ۰/۴۵ = ۴/۵ = ۴ + ۰/۵ \quad P_{۴۵} = (۱ \quad ۰/۵)x_۴ + ۰/۵x_۵ = ۰/۵ \times ۱۶ + ۰/۵ \times ۱۶ = ۱۶$

مثال ۳۹. داده‌های زیر، ساعات اضافه کاری کارمندان یک شرکت را نشان می‌دهد مطلوب است محاسبه: **الف.** چارک اول **ب.** چارک دوم **پ.** چارک سوم **ت.** دهک ۶ام **ث.** صدک ۶۳ام

۱۶ ۲۴ ۲۴ ۲۱ ۱۸ ۱۹ ۱۸ ۱۹ ۱۸ ۲۲ ۲۲ ۲۴ ۱۰ ۱۲ ۱۶ ۱۵ ۱۶ ۱۳ ۱۳ ۱۵

۲۴ ۲۴ ۲۴ ۲۲ ۲۲ ۲۱ ۱۹ ۱۸ ۱۸ ۱۸ ۱۶ ۱۶ ۱۶ ۱۵ ۱۵ ۱۳ ۱۳ ۱۲ ۱۰ مرتب کردن بصورت صعودی

$Q_۱ = Q_{۰/۲۵} \quad (n+۱)p = (۲۰+۱) \times ۰/۲۵ = ۵/۲۵ = ۵ + ۰/۲۵$

$Q_۱ = (۱ - ۰/۲۵)x_۵ + ۰/۲۵x_۶ = ۰/۷۵ \times ۱۵ + ۰/۲۵ \times ۱۵ = ۱۵$

$Q_۲ = Q_{۰/۵۰} \quad (n+۱)p = (۲۰+۱) \times ۰/۵۰ = ۱۰/۵ = ۱۰ + ۰/۵$

$Q_۲ = (۱ - ۰/۵)x_{۱۰} + ۰/۵x_{۱۱} = ۰/۵ \times ۱۸ + ۰/۵ \times ۱۸ = ۱۸$

$Q_3 = Q_{۰/۷۵} \quad (n+۱)p = (۲۰+۱) \times ۰/۷۵ = ۱۵/۷۵$

$Q_۳ = (۱ - ۰/۷۵)x_{۱۵} + ۰/۷۵x_{۱۶} = ۰/۲۵ \times ۲۱ + ۰/۷۵ \times ۲۲ = ۲۱/۷۵$

$D_۶ = Q_{۰/۶۰} \quad (n+۱)p = (۲۰+۱) \times ۰/۶ = ۱۲/۶ = ۱۲ + ۰/۶$

$D_۶ = (۱ - ۰/۶)x_{۱۲} + ۰/۶x_{۱۳} = ۰/۴ \times ۱۸ + ۰/۶ \times ۱۹ = ۱۸/۶$

$P_{۶۳} = Q_{۰/۶۳} \quad (n+۱)p = (۲۰+۱) \times ۰/۶۳ = ۱۳/۲۳ = ۱۳ + ۰/۲۳$

$P_{۶۳} = (۱ - ۰/۲۳)x_{۱۳} + ۰/۲۳x_{۱۴} = ۰/۷۷ \times ۱۹ + ۰/۲۳ \times ۱۹ = ۱۹$

نقاط دور افتاده:

یک مقدار دور افتاده در مقایسه با بقیه مقادیر داده ها یک مقدار کوچک یا بزرگ افراطی است.

روش تحقیق برای بررسی داده های دور افتاده:

گامهای زیر امکان بررسی اینکه آیا مقدار مفروضی در مجموعه داده ها را می‌توان به عنوان داده دور افتاده رده بندی کرد یا خیر به ما میدهد.

گام یک: $Q_1 = H_{25}$ $\quad Q_3 = H_{75}$ $\quad IR = Q_3 - Q_1$ محاسبه می کنیم

گام دو: $x < Q_1 - 1.5IR$ یا $x > Q_3 + 1.5IR$ باشد، آنگاه x را داده دور افتاده می‌نامیم

مثال: داده‌های زیر ۲۰ کشور را با بیشترین تعداد مدال‌هایی که در بازی المپیک سال ۱۹۹۶ آتلانتا کسب کرده‌اند نشان می‌دهد، از جمله ایالات متحده آمریکا که ۱۰۱ مدال دارد. آیا تعداد مدالهای آمریکا نسبت به تعداد مدالهای ۱۹ کشور دیگر یک مقدار دور افتاده است؟

63 65 50 37 35 41 25 23 27 21

15 17 17 20 19 22 15 15 15 101

حل: داده ها را بصورت صعودی مرتب می کنیم

15 15 15 15 17 17 19 20 21 22

23 25 27 35 37 41 50 63 65 101

$$Q_1: \ \frac{(n+1)h}{100} = 21 \times 0.25 = 5.25 \ \Rightarrow r = 5, \ w = 0.25$$

$$\Rightarrow Q_1 = 0.75x_5 + 0.25x_6 = 0.75 \times 17 + 0.25 \times 17 = 17$$

$$Q_3: \ \frac{(n+1)h}{100} = 21 \times 0.75 = 15.75 \ \Rightarrow r = 15, \ w = 0.75$$

$$\Rightarrow Q_3 = 0.75x_{15} + 0.25x_{16} = 0.25 \times 37 + 0.75 \times 41 = 40$$

$$IR = Q_3 - Q_1 = 40 - 17 = 23$$

$$Q_1 - 1.5IR = 17 - 1.5 \times 23 = -17.5$$

$$Q_3 + 1.5IR = 40 + 1.5 \times 23 = 74.5$$

چون $101 > 72$ است تعداد مدال‌های آمریکا نسبت به سایر کشورها یک داده دور افتاده است.

مثال: در مثال قالب‌های کره داده‌های پرت را مشخص کنید.

i	رده ها	x_i	f_i	f_{ci}
1	20.5-25.5	23	3	3
2	25.5-30.5	28	6	9
3	30.5-355	33	10	۱۹
4	35.5-40.5	38	8	27
5	40.5-45.5	43	6	33
6	15.5-505	48	5	38
7	50.5-55.5	53	2	40
Σ	---	---	40	---

$$H_h = L + \frac{\frac{nh}{100} - f_c(m-1)}{f_m}(U - L)$$

$$Q_1 = H_{25} = ? \qquad \frac{nh}{100} = \frac{40 \times 25}{100} = 10 \rightarrow m = 3, \ f_c(\mathrm{m}-1) = 9, \ L = 30.5, \ U = 35.5, \ f_m = 10$$

$$H_{25} = 30.5 + \frac{10 - 9}{10} \times (35.5 - 30.5) = 31 = \mathrm{Q}_1$$

$$Q_3 = H_{75} = ?$$

$$\frac{nh}{100} = \frac{40 \times 75}{100} = 30 \rightarrow m = 5, \ f_c(\mathrm{m}-1) = 27, \ L = 40.5, \ U = 45.5, \ f_m = 6$$

$$H_{25} = 40.5 + \frac{30 - 27}{6} \times (45.5 - 40.5) = 43 = \mathrm{Q}_3 \qquad IR = Q_3 - Q_1 = 43 - 31 = 12$$

$$x < Q_1 - 1.5\,IR \ , \ x > Q_3 + 1.5\,IR \Rightarrow \begin{cases} x < 31 - 1.5 \times 12 \rightarrow x < 13 \\ x > 43 + 1.5 \times 12 \rightarrow x > 61 \end{cases}$$

هیچ داده ای پرت نیست.

قبل از خاتمه این بحث، به معرفی دو معیار گرایش به مرکزی دیگری می‌پردازیم که مرکز داده‌ها را بهتر توصیف می‌کنند، این دو عبارتند از:

الف) **میانگین پیراسته** ب) **میانگین وینزوری**

طرز به دست آوردن میانگین پیراسته:

برای این منظور تمام مشاهدات کوچکتر از چارک اول و تمام مشاهدات بزرگتر از چارک سوم را بر می‌داریم سپس میانگین مشاهدات باقیمانده را حساب می‌کنیم.

طرز به دست آوردن میانگین وینزوری:

برای این منظور به جای هر یک از مشاهدات کوچکتر از چارک اول، مقدار چارک اول و به جای هر یک از مشاهدات بزرگتر از چارک سوم ، مقدار چارک سوم را قرار میدهیم، بقیه مشاهدات را تغییر نداده سپس میانگین تمام مشاهدات را محاسبه می‌کنیم.

سوال: چرا میانگین میانگین پیراسته و میانگین وینزوری اهمیت دارند؟

- در نظر داشته باشید که حتی معدودی مشاهده که به طور غیر عادی بزرگ یا کوچک باشند، می‌توانند بر میانگین اثر بگذارند.
- حال آن که میانه، بعد از آن که داده‌ها مرتب شدند، به جز یک یا دو مقدار وسطی، با بقیه کاری ندارد.
- میانگین‌های پیراسته و وینزوری را می‌توان تلفیقی از میانگین نمونه و میانه نمونه‌ای دانست.
- وجود معدودی مشاهده غیر معمول یا اشتباه انگیز که خیلی کوچک یا خیلی بزرگ باشند، تاثیری در میانگین پیراسته و میانگین وینزوری نمی‌گذارد.
- معیار اولی، مشاهداتی را که کمتر از چارک اول یا بیشتر از چارک سوم هستند کنار می‌گذارد،
- معیار دومی، قبل از متوسط‌گیری، این مشاهدات را با چارک های متناظر جایگزین می‌کند.
- می بینیم که در توزیع‌های متقارن، این دو معیار تقریباً به خوبی میانگین هستند ولی در صورت وجود مقادیر غیرمنتظره هر دو از میانگین بهترند.
- وقتی منحنی توزیع یک دنباله کشیده داشته باشد، این معیارها به جای میانگین و میانه به عنوان گرایش به مرکز به کار می‌روند.

مثال: در مثال کارگران صنعتی قدرت پنجه آنها بترتیب غیر نزولی مرتب شده اند.

$$93.9,\ 105.8,\ \ 106.5,\ 116.6,\ 125,\ 128.3\ 132.1,\ \ 137.7,\ 152.4$$

$$\underset{\uparrow}{} \qquad\qquad \underset{\uparrow}{}$$

$$Q_1 = 106.15 \qquad\qquad Q_۳ = ۱۳۴/۴$$

میانگین پیراسته و وینزوری داده‌های بالا کدامند؟

حل:

$$\text{میانگین پیراسته} = \frac{106.5+116.6+125+128.3+132.1}{5} = 121.7$$

$$\text{میانگین وینزوری} = \frac{106.15+106.15+106.5+116.6+125+128.3+132.1+134.4+134.4}{9} = 121.05$$

تمرینات صفحه 145 الی 146 حل شود

نمودار جعبه‌ای:

نمودارهایی که تاکنون شناخته‌ایم هر کدام به طریقی داده‌ها را نمایش می‌دادند و برای مقایسه داده‌ها بسیار مفیدند ولی هیچ کدام به سوالاتی از قبیل اینکه

- آیا داده‌ها به هم نزدیک هستند؟
- آیا داده‌ها بیشتر در اطراف میانگین متمرکزند
- یا بیشتر اطراف کمترین داده یا بیشترین داده متمرکزند؟

پاسخ نمی دهند.

نمودار جعبه‌ای نموداری تصویری است که داده ها را بر اساس پنج مقدار نمایش می‌دهد. این مقادیر عبارتند از:

- کوچکترین داده
- چارک اول
- میانه
- چارک سوم
- بزرگترین داده.

مثال: فرض کنید تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان بصورت زیر باشند.

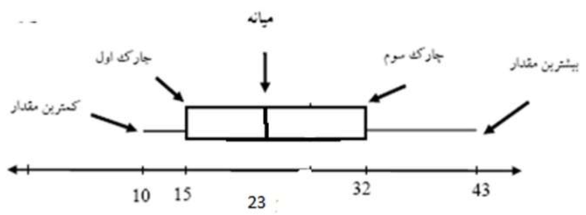
12 10 15 23 14 27 16 34 41 43 32 18 25 31 19

نمودار جعبه‌ای داده های فوق را رسم کنید.

حل: داده‌های فوق را بصورت غیر نزولی مرتب می‌کنیم.

10 12 14 15 16 18 19 23 25 27 31 32 34 41 43

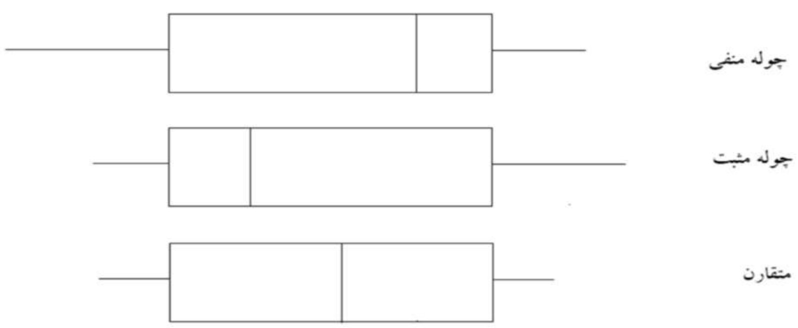
بنابراین: ۱- کوچکترین داده = ۱۰ ۲- چارک اول= ۱۵ ۳- میانه = ۲۳ ۴- چارک سوم = ۳۲ ۵- بزرگترین داده = ۴۳



اطلاعاتی که می‌توان از یک نمودار جعبه‌ای بدست آورد

۱-اگر میانه به مرکز جعبه نزدیک باشد، توزیع مقادیر داده‌ها، تقریباً متقارن است.
۲-اگر میانه در سمت چپ مرکز جعبه باشد، توزیع مقادیر داده‌ها، به طور مثبت چوله است.
۳-اگر میانه در سمت راست مرکز جعبه باشد، توزیع مقادیر داده‌ها، به طور منفی چوله است.
۴-اگر دنباله‌ها تقریباً دارای طول یکسان باشند، توزیع مقادیر داده‌ها تقریباً متقارن است.
۵-اگر دنباله راست طویل‌تر از دنباله چپ باشد، توزیع مقادیر داده‌ها به طور مثبت چوله است.
۶-اگر دنباله چپ طویل‌تر از دنباله راست باشد، توزیع مقادیر داده‌ها به طور منفی چوله است.

شکل ۸-۴ ، نمودارهای جعبه‌ای که این مشخصه‌ها را نمایش می‌دهند، نشان می‌دهد.



مثال: اطلاعات زیر مفروض است، نمودارهای جعبه‌ای هر یک را رسم کرده و در مورد چولگی آنها اظهار نظر کنید.

الف) $\max=70,\ Q_3=52,\ Q_2=33,\ Q_1=30,\ \min=25$ ج) $\max=80,\ Q_3=65,\ Q_2=45,\ Q_1=25,\ \min=10$

ب) $\max=65,\ Q_3=60,\ Q_2=55,\ Q_1=40,\ \min=5$

حل: (آ) ملاحظه کنید که طول دنباله چپ $30-25=5$ و طول دنباله راست $70-52=18$ همچنین فاصله میانه از Q_1 تنها برابر $33-30=3$ واحد و فاصله میانه از Q_3 برابر $52-33=19$ واحد است. از این اطلاعات می‌توان نتیجه گرفت که توزیع چوله به راست یا با چولگی مثبت است. نمودار جعبه‌ای بصورت زیر است.

چوله مثبت

ب) ملاحظه کنید که طول دنباله چپ $40-5=35$ و طول دنباله راست $65-60=5$ است. همچنین فاصله میانه از Q_1 برابر $60-55=5$ واحد و از Q_3 برابر $55-40=15$ واحد است. از این اطلاعات، می‌توان نتیجه گرفت که توزیع چوله به چپ است. نمودار جعبه‌ای برای این اطلاعات در شکل زیر نشان داده شده است.

چوله منفی

ج) ملاحظه کنید که طول دنباله چپ $25-10=15$ و طول دنباله راست $80-65=15$. همچنین فاصله میانه از Q_1 برابر $45-25=20$ واحد و از Q_3 تنها برابر $65-45=20$ واحد است. از این اطلاعات می‌توان نتیجه گرفت که این توزیع متقارن است یا با چولگی صفر است. نمودار جعبه‌ای بصورت زیر است.

متقارن

مد یا نما : مد یا نما اندازه‌ای است عددی که در بین داده‌ها بیشترین تکرار را دارد.

- اگر همه عناصر در مجموعه داده‌ها ، فراوانی یکسانی داشته باشند، در این صورت گفته می‌شود که مجموعه داده‌ها دارای نما نیست .
- اگر مجموعه داده‌ها دارای یک مقدار باشد که فراوانی آن بیشتر از بقیه مقادیر باشد در این صورت گفته می شود که مجموعه داده‌ها تک نمایی است
- اگر دو داده از مجموعه داده‌ها دارای بالاترین فراوانی یکسانی باشند آنگاه گفته می‌شود که مجموعه داده‌ها دو نمایی است.

مثال ۳۱. توزیع فراوانی گروه خونی ۳۵ نفر به صورت زیر داده شده است. مد یا نما را مشخص کنید.

گروه خونی	A	B	O	AB
فراوانی	۵	۷	۱۳	۱۰

حل: چون بیشترین فراوانی مربوط به گروه خونی O می‌باشد، لذا $M_o=\mathrm{O}$.

توجه: مد تنها شاخص مرکزی برای صفات کیفی است.

اگر صفت x یک صفت کمی ناپیوسته با توزیع فراوانی به صورت زیر باشد:

x_i	x_1	x_2	$\cdots$	x_k	$\cdots$	x_n
فراوانی	f_1	f_2	$\cdots$	f_k	$\cdots$	f_n

و اگر f_k ماکسیمم f_iها باشد، $M_o=x_k$.

مثال ۳۲. توزیع فراوانی تعداد افراد خانوار به صورت زیر داده شده، مد را تعیین کنید.

x_i	۱	۲	۳	۴	۵	۶	۷
f_i	۵	۱۰	۲۰	۱۰	۴۰	۲۰	۱۰

حل: بیشترین فراوانی مربوط به خانوارهای ۵ نفره است، پس $M_o = ۵$.

محاسبه مد:

اگر داده‌ها بصورت رده‌ها بیان شود،

برای محاسبه مد از رابطه زیر استفاده می کنیم:

$$Mo = L + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}(U-L)$$

$U, L, f_m =$ حد بالا و پایین و فراوانی دسته مد دار

$f_{m-1}, f_{m+1} =$ فراوانی دسته ما قبل و ما بعد دسته مد دار

مثال: مد داده‌های رده بندی شده زیر را بیابید

حل:

i	رده‌ها	f_i	f_{ci}
۱	۰-۵	۳	۳
۲	۵-۱۰	۵	۸
۳	۱۰-۱۵	۹	۱۷
۴	۱۵-۲۰	۱۳	۳۰
۵	۲۰-۲۵	۵	۳۵
۶	۲۵-۳۰	۳	۳۸
۷	۳۰-۳۵	۲	٤۰

۴
۸

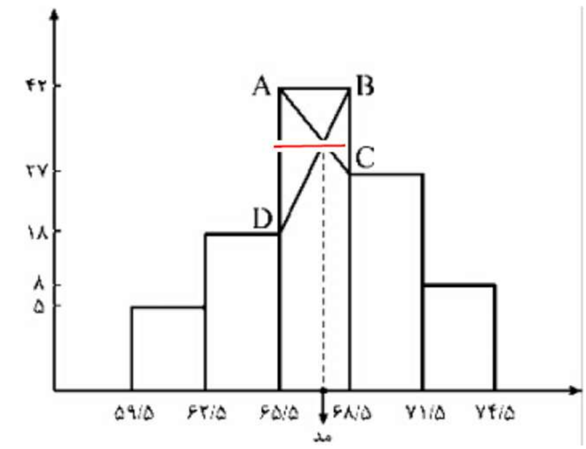
$$M_o = L + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}(U-L)$$

$$= 15 + \frac{13-9}{2\times 13 - 9 - 5}(20-15)$$

$$= 15 + \frac{20}{12} = 16.7$$

محاسبه مد به روش ترسیمی

هیستوگرام فراوانی را رسم می‌کنیم. بلندترین مستطیل را در نظر می‌گیریم و از A به C و از B به D وصل می‌کنیم، از محل تقاطع این دو خط، بر محور xها عمود می‌کنیم، پای عمود تا مبدأ را مد می‌نامیم، نمودار زیر مربوط به مثال بالا می‌باشد.



$$Mo = L + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}(U-L)$$

مزایای نسبی میانگین، میانه و مد:

- میانگین متداول‌ترین معیار تمایل به مرکز است.
- محاسبه آن آسان و برای عملیات جبری قالب ریزی شده است.
- اگر داده پرت نداشته باشیم میانگین حسابی نسبت به میانه و نما با ثبات‌تر بوده و نوسانات آن کمتر است یعنی اگر میانگین حسابی دو نمونه را با یکدیگر مقایسه کنیم نزدیکتر از میانه آن دو نمونه خواهد بود.
- میانگین تحت تاثیر تمام داده‌ها بوده و داده‌های پرت روی آن تاثیر دارند.

مثال: داده‌های $x_i:\ 1\quad 2\quad 3\quad 3\quad 4\quad 5$ را در نظر بگیرید، داریم $\overline{x} = m_d = M_0 = 3$

حال داده‌های $x_i: 1,2,3,3,4,6005$ را در نظر بگیرید داریم $m_d = M_0 = 3,\ \overline{x} = 1003$

- در نتیجه میانگین تحت تاثیر داده‌ها پرت قرار می‌گیرد
- اما میانه کمتر تحت تاثیر داده‌های پرت قرار می‌گیرد
- در صورتیکه مد و میانه کمتر از داده پرت تاثیر می‌پذیرند.
- نما به علت مبهم بودن، حتی از میانه نیز اهمیت کمتری دارد
- ولی در حالت کلی از لحاظ اهمیت در رده سوم قرار دارد.
- لازم به ذکر است که برای داده‌های کیفی شاخص مرکزی مناسب، نما است.

پارامترهای پراکندگی

در فصل قبل، در مورد پارامترهای مرکزی صحبت کردیم. ین پارامترها به تنهایی نمی‌تو نند سایر جنبه‌های داده‌ها را مشخص کنند.

گروه اول	گروه دوم
۶۵	۴۲
۶۶	۵۴
۶۷	۵۸
۶۸	۶۲
۷۱	۶۷
۷۳	۷۷
۷۴	۷۷
۷۷	۸۵
۷۷	۹۳
۷۷	۱۰۰

گروه اول $\bar{x} = ۷۱/۵\quad , \quad M_d = ۷۲\quad , \quad M_o = ۷۷$

گروه دوم $\bar{x} = ۷۱/۵\quad , \quad M_d = ۷۲\quad , \quad M_o = ۷۷$

از لحاظ شاخص‌های مرکزی، تفاوتی بین دو گروه وجود ند رد.

مید ن تغییرات صفت

در گروه اول، دامنه برابر با $۷۷ - ۶۵ = ۱۲$

در گروه دوم دامنه برابر با $۱۰۰ - ۴۲ = ۵۸$

پارامترهای مرکزی ین اختلاف دو توزیع را نشان نمی‌دهند.

پارامترهای پراکندگی:

یک معیار تغییر پذیری برای مجموعه از مقادیر داده‌ها عددی است که هدف آن بیان ایده پراکندگی برای این مجموعه داده‌ها است.

معمول‌ترین معیارهای تغییر پذیری برای داده‌های نمونه عبارتند از:

دامنه : دامنه برابر تفاضل بین بزرگترین و کوچکترین مقدار در مجموعه‌ای از داده‌ها است $R = \max - \min$

این تعریف برای یک نمونه و همچنین جامعه‌ای نا متناهی درست است.

نکته:

دامنه از مفهوم انحراف‌ها استفاده نمی‌کند.

دامنه از نقاط دور افتاده (مقدار بزرگ یا کوچک نسبت به بقیه مجموعه داده‌ها) تأثیر می‌پذیرد

همه اطلاعات در مجموعه داده‌ها را مورد استفاده قرار نمی‌دهد.

بنابراین معیار تغییر پذیری بسیار سود مند نیست.

دامنه میان چارکی: دامنه میان چارکی پراکندگی 50% وسط یک مجموعه داده‌های مرتب شده را اندازه می‌گیرد آن برابر است با :

$$IR = Q_3 - Q_1$$

دامنه صدک‌ها

$$\text{دمنه صدک} = P_{۹۰} - P_{۱۰}$$

میانگین انحراف مطلق:

تعریف می‌کنیم (μ میانگین X است)

$$D_i = X_i - \mu$$

بدیهی است که میانگین D صفر است

لذا برای به دست آوردن یک معیار برای اندازه‌گیری پراکندگی تعریف می‌کنیم

$$M_D = \frac{1}{n}\sum_{i=1}^{n}|X_i - \mu|$$

مثال: توزیع فراوانی زیر مفروض است.

x_i	1	2	3	4	5
f_i	10	40	10	20	20

میانگین قدر مطلق انحراف هارا محاسبه کنید.

حل: جدول محاسباتی زیر را تشکیل می‌دهیم

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} f_i X_i$$

$$= \frac{300}{100} = 3,$$

$$M_D = \frac{1}{n}\sum_{i=1}^{n}|X_i - \mu|$$

$$= \frac{1}{100}\sum_{i=1}^{5} f_i |X_i - \bar{X}| = 1.2$$

x_i	f_i	$x_i f_i$	$\lvert X_i - \bar{X}\rvert$	$f_i\lvert X_i - \bar{X}\rvert$
۱	۱۰	۱۰	۲	۲۰
۲	۴۰	۸۰	۱	۴۰
۳	۱۰	۳۰	۰	۰
٤	۲۰	۸۰	۱	۲۰
٥	۲۰	۱۰۰	۲	۴۰
Σ	۱۰۰	۳۰۰	---	۱۲۰

واریانس و انحراف معیار:

واریانس نمونه‌ای یک متوسط تقریبی از توان دوم انحراف‌ها از میانگین نمونه است و آن را با S^2 نشان می‌دهیم و از رابطه زیر محاسبه می‌شود:

- داده‌های خام:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \qquad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

- داده‌ها با تکرار:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}f_i(X_i-\bar{X})^2 \qquad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}f_iX_i$$

- داده‌های رده بندی شده

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}f_i(\hat{X}_i-\bar{X})^2 \quad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}f_i\hat{X}_i$$

ثابت می‌شود

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

$$=\sum_{i=1}^{n}X_i^2-\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n},$$

$$\sum_{i=1}^{n}f_i(X_i-\bar{X})^2=\sum_{i=1}^{n}f_iX_i^2-n\bar{X}^2$$

$$=\sum_{i=1}^{n}f_iX_i^2-\frac{\left(\sum_{i=1}^{n}f_iX_i\right)^2}{n},$$

انحراف معیار: ریشه دوم و مثبت واریانس است

$$S=\sqrt{S^2}$$

مثال ۵. واریانس داده‌های زیر را محاسبه کنید.

۱۲ , ۶ , ۷ , ۳ , ۱۵ , ۱۰ , ۱۸ , ۵

$$\bar{x}=\frac{۱}{n}\sum x_i=\frac{۱}{۸}(۱۲+۶+۷+۳+۱۵+۱۰+۱۸+۵)=۹/۵$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{8-1}\Big((۱۲-۹/۵)^۲+(۶-۹/۵)^۲+(۷-۹/۵)^۲+(۳-۹/۵)^۲$$

$$+(۱۵-۹/۵)^۲+(۱۰-۹/۵)^۲+(۱۸-۹/۵)^۲+(۵-۹/۵)^۲\Big)$$

$$=۲۳/۷۵$$

$$s=\sqrt{۲۳/۷۵}=۴/۸۷$$

مثال ۶. جدول توزیع فراوانی نمرات درس آمار ۱۰۰ دانشجو به صورت زیر مفروض است مطلوب است محاسبه واریانس.

	حدود طبقات	f_i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^۲$	$f_i(x_i - \bar{x})^۲$	$f_i x_i$
۵۹/۵−۶۲/۵	۶۰ − ۶۲	۵	۶۱	−۶٫۴۵	۴۱٫۶۰۲۵	۲۰۸٫۰۱۲۵	۳۰۵
۶۲/۵−۶۵/۵	۶۳ − ۶۵	۱۸	۶۴	−۳٫۴۵	۱۱٫۹۰۲۵	۲۱۴٫۲۴۵۰	۱۱۵۲
۶۵/۵−۶۸/۵	۶۶ − ۶۸	۴۲	۶۷	−۰٫۴۵	۰٫۲۰۲۵	۸٫۵۰۵۰	۲۸۱۴
۶۸/۵−۷۱/۵	۶۹ − ۷۱	۲۷	۷۰	۲٫۵۵	۶٫۵۰۲۵	۱۷۵٫۵۶۷۵	۱۸۹۰
۷۱/۵−۷۴/۵	۷۲ − ۷۴	۸	۷۳	۵٫۵۵	۳۰٫۸۰۲۵	۲۶۴٫۴۲۰۰	۵۸۴
		۱۰۰				۸۵۲٫۷۵۰۰	۶۷۴۵

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} f_i \widehat{X}_i = \frac{۶۷۴۵}{۱۰۰} = ۶۷٫۴۵$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} f_i(\widehat{X}_i - \overline{X})^2 = \frac{852.75}{100-1} = 8.53$$

$$s = \sqrt{۸٫۵۲۷۵} = ۲٫۹۲$$

خواص واریانس

الف. اگر به تمام داده‌ها، مقدار ثابت b را اضافه کنیم، واریانس داده‌های جدید برابر با واریانس داده‌های قبلی است.

$$Var\left(X+a\right) = Var\left(X\right)$$

ب. اگر تمام داده‌ها ر در عدد ثابتی مانند a ضرب کنیم، و ریانس داده‌های جدید در $a^۲$ ضرب می‌شود. فرض کنید $Y = aX$.

$$Var\left(aX\right) = a^2 Var\left(X\right)$$

پ. به سادگی می‌توان نشان داد که

$$Var\left(aX+b\right) = a^2 Var\left(X\right)$$

$$S^2_{aX+b} = a^2 S^2_X$$

محاسبه واریانس بااستفاده‌از میانگین فرضی

اگر داده‌ها بصورت رده بندی ارائه شود، می‌توان با استفاده از رابطه زیر واریانس نمونه‌ای را به دست آورد:

$$d_i = \frac{\hat{X}_i - A}{I}$$

$$\hat{X}_i = Id_i + A$$

$$S^2 = I^2 S_d^2 = I^2 \frac{1}{n-1} \sum_{i=1}^{n} f_i (d_i - \bar{d})^2$$

$$= I^2 \frac{\sum_{i=1}^{n} fd_i^2 - n\bar{d}^2}{n-1} = I^2 \frac{\sum_{i=1}^{n} fd_i^2 - \frac{\left(\sum_{i=1}^{n} fd_i\right)^2}{n}}{n-1}$$

مثال: در مثال ***1*** واریانس وزن قالب های کره را بدست آورید؟

i	رده ها	x_i	f_i	$x_i f_i$	d_i	$d_i f_i$	$x_i^2 f_i$	$d_i^2 f_i$
١	٢٠/٥-٢٥/٥	٢٣	٣	٦٩	-٢	-٦	١٥٨٧	١٢
٢	٢٥/٥-٣٠/٥	٢٨	٦	١٦٨	-١	-٦	٤٧٠٤	٦
٣	٣٠/٥-٣٥/٥	٣٣	١٠	٣٣٠	٠	٠	١٠٨٩٠	٠
٤	٣٥/٥-٤٠/٥	٣٨	٨	٣٠٤	١	٨	١١٥٥٢	٨
٥	٤٠/٥-٤٥/٥	٤٣	٦	٢٥٨	٢	١٢	١١٠٩٤	٢٤
٦	٤٥/٥-٥٠/٥	٤٨	٥	٢٤٠	٣	١٥	١١٥٢٠	٤٥
٧	٥٠/٥-٥٥/٥	٥٣	٢	١٠٦	٤	٨	٥٦١٨	٣٢
$\sum$	---	---	٤٠	١٤٧٥	---	٣١	٥٦٩٦٥	١٢٧

راه حل اول:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} f_i \hat{X}_i = \frac{1475}{40} = 36.875,$$

راه حل دوم:

$$\bar{X} = A + I \frac{1}{n} \sum_{i=1}^{k} f_i \hat{d}_i = 33 + 5 \frac{31}{40} = 36.875$$

راه حل اول:

$$S^2 = \frac{\sum_{i=1}^{n} f_i \hat{X}_i^2 - \frac{\left(\sum_{i=1}^{n} f_i \hat{X}_i\right)^2}{n}}{n-1} = \frac{56965 - \frac{1475^2}{40}}{39} = 66.00$$

راه حل دوم:

$$S^2 = I^2 \frac{1}{n-1} \left(\sum_{i=1}^{n} fd_i^2 - \frac{\left(\sum_{i=1}^{n} fd_i\right)^2}{n} \right)$$

$$= \frac{25}{39} \left(127 - \frac{31^2}{40} \right) = 66.00 \Rightarrow S = 8.12$$

مثال ۶. جدول توزیع فراوانی نمرات درس آمار ۱۰۰ دانشجو به صورت زیر مفروض است مطلوب است محاسبه واریانس.

حدود طبقات	f_i	x_i	u_i	f_i	f_iu_i	$f_iu_i^2$
۶۰ − ۶۲	۵	۶۱	−۲	۵	−۱۰	۲۰
۶۳ − ۶۵	۱۸	۶۴	−۱	۱۸	−۱۸	۱۸
۶۶ − ۶۸	۴۲	۶۷	۰	۴۲	۰	۰
۶۹ − ۷۱	۲۷	۷۰	۱	۲۷	۲۷	۲۷
۷۲ − ۷۴	۸	۷۳	۲	۸	۱۶	۳۲
	۱۰۰			۱۰۰	۱۵	۹۷

حدود جدید

59.5 – 62.5
62.5 – 65.5
65.5 – 68.5
68.5 – 71.5
71.5 – 74.5

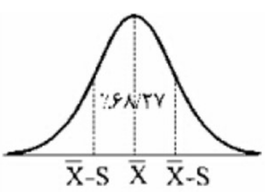
$$S^2 = I^2\,\frac{1}{n-1}\left(\sum_{i=1}^{n} fd_i^2 - \frac{\left(\sum_{i=1}^{n} fd_i\right)^2}{n}\right) \Rightarrow S^2 = 3^2\,\frac{1}{100-1}\left(97-\frac{15^2}{15}\right) = 3^2 \times 0.9475$$

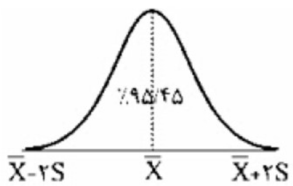
در توزیع‌های نرمال روابط زیر برقرار است.

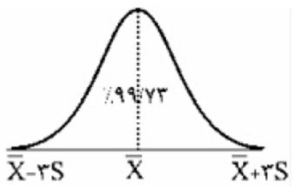
۱. در حدود ۶۸٫۲۷ درصد داده‌ها در فاصله $(\overline{X}-S\ ,\ \overline{X}+S)$ قرار دارند. $1-\frac{1}{3}$

۲. در حدود ۹۵٫۴۵ درصد داده‌ها در فاصله $(\overline{X}-2S\ ,\ \overline{X}+2S)$ قرار دارند. $1-\frac{1}{20}$

۳. در حدود ۹۹٫۷۳ درصد داده‌ها در فاصله $(\overline{X}-3S\ ,\ \overline{X}+3S)$ قرار دارند. $1-\frac{1}{100}$







مثال ۱۱. جدول توزیع ضریب هوش ۴۸۰ محصل به صورت زیر داده شده، مطلوب است درصد افرادی که در فوصل زیر قرار دارند:

الف. $\bar{x} \pm s$ **ب.** $\bar{x} \pm ۲s$ **پ.** $\bar{x} \pm ۳s$

۶۸ − ۷۲

x_i	۷۰	۷۴	۷۸	۸۲	۸۶	۹۰	۹۴	۹۸	۱۰۲	۱۰۶	۱۱۰	۱۱۴	۱۱۸	۱۲۲	۱۲۶
f_i	۴	۹	۱۶	۲۸	۴۵	۶۶	۸۵	۷۲	۵۴	۳۸	۲۷	۱۸	۱۱	۵	۲

$$\frac{360}{480} = 0.75$$

$$\frac{449}{480} = 0.93$$

$$\frac{480}{480} = 1$$

x_i	$u_۲$	f_i	f_iu_i	$f_iu_i^۲$
۷۰	−۶	۴	−۲۴	۱۴۴
۷۴	−۵	۹	−۴۵	۲۲۵
۷۸	−۴	۱۶	−۶۴	۲۵۶
۸۲	−۳	۲۸	−۸۴	۲۵۲
۸۶	−۲	۴۵	−۹۰	۱۸۰
۹۰	−۱	۶۶	−۶۶	۶۶
۹۴	۰	۸۵	۰	۰
۹۸	۱	۷۲	۷۲	۷۲
۱۰۲	۲	۵۴	۱۰۸	۲۱۶
۱۰۶	۳	۳۸	۱۱۴	۳۴۲
۱۱۰	۴	۲۷	۱۰۸	۴۳۲
۱۱۴	۵	۱۸	۹۰	۴۵۰
۱۱۸	۶	۱۱	۶۶	۳۹۶
۱۲۲	۷	۵	۳۵	۲۴۵
۱۲۶	۸	۲	۱۶	۱۲۸
		۴۸۰	۲۳۶	۳۴۰۴

$$\bar{u} = \frac{۱}{n}\sum f_iu_i = \frac{۲۳۶}{۴۸۰}$$

$$x = I\bar{u} + A = 94 + 4\bar{u}$$

$$\Rightarrow \bar{x} = A + I\bar{u} = 94 + 4\bar{u}$$

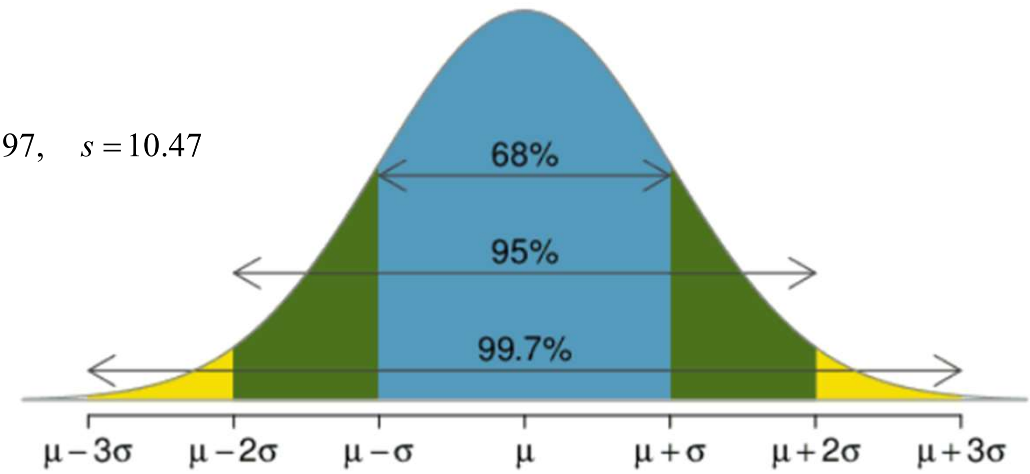
$$= ۹۴ + ۴ \times \frac{۲۳۶}{۴۸۰} = ۹۵/۹۷$$

$$S^2 = I^2 \frac{1}{n-1}\left(\sum_{i=1}^{n} f_i d_i^2 - \frac{\left(\sum_{i=1}^{n} f_i d_i\right)^2}{n}\right)$$

$$= 4^2 \frac{1}{480-1}\left(3404 - \frac{236^2}{480}\right)$$

$$= 16 \times 6.84 \qquad \Rightarrow s = 10.47$$

$\bar{x} = 95.97, \quad s = 10.47$



۶۴/۶	۷۵	۸۵/۵	۹۵/۹۷	۱۰۶/۴	۱۱۶/۹	۱۲۷/۴
$\bar{x} - 3s$	$\bar{x} - 2s$	$\bar{x} - s$	$\bar{x}$	$\bar{x} + s$	$\bar{x} + 2s$	$\bar{x} + 3s$

پارامترهای نسبی پراکندگی

۱. ضریب تغییرات

۲. ضریب چولگی

۳. ضریب کشیدگی

ضریب تغییرات:

نسبت انحراف معیار به میانگین یعنی

$$CV = \frac{s}{\bar{x}}$$

این ضریب اغلب بصورت درصد بیان می‌شود

را ضریب تغییرات می‌نامند.

به واحد اندازه‌گیری بستگی ندارد.

در عمل برای مقایسه پراکندگی بین دو متغیر که دارای واحدهای متفاوتند و یا واحدهای آنها یکسان ولی میانگین آنها مختلفند، بکار می‌رود.

مثال: کارخانه‌ای دو نوع تسمه تولید می‌کند. تسمه نوع اول دارای میانگین طول عمر ۲۰۰ ساعت و انحراف معیار ۲۰ساعت و تسمه نوع دوم دارای میانگین طول عمر ۳۶۰ ساعت با انحراف معیار ۲۵ ساعت است. کدام نوع تسمه بهتر است؟

حل:

$$\bar{x}_1 = 200, \qquad s_1 = 20, \qquad CV_1 = \frac{s_1}{\bar{x}_1} = \frac{20}{200} = 10\%$$

$$\bar{x}_2 = 360, \qquad s_2 = 25, \qquad CV_2 = \frac{s_2}{\bar{x}_2} = \frac{25}{360} = 7\%$$

بنابراین تسمه نوع دوم بهتر است چون دارای میانگین طول عمر بیشتر و ضریب تغییرات کمتری است.

مثال ۱۲. فرض کنید میانگین و انحراف معیار طول قد دانشجویان دو کلاس به صورت زیر مفروض باشند.

$$\bar{x} = ۱۷۵ \ , \quad S_X = ۵ \ , \quad \bar{y} = ۱۶۰ \ , \quad S_Y = ۵$$

$$C \cdot V_X = \frac{S_X}{\bar{X}} \times ۱۰۰ = \frac{۵}{۱۷۵} \times ۱۰۰ = \%۲/۸۵$$

$$C \cdot V_Y = \frac{S_Y}{\bar{Y}} \times ۱۰۰ = \frac{۵}{۱۶۰} \times ۱۰۰ = \%۳/۱۲۵$$

$\Rightarrow$ یعنی پراکندگی صفت Y بیش ز پرکندگی صفت X است.

مثال ۱۳. یک تولیدکننده لامپ تصویر تلویزیون دو نوع لامپ تصویر تولید می‌کند. نوع A و B، عمر متوسط A برابر ۱۴۹۵ ساعت و انحراف معیار آن برابر ۲۸۰ ساعت است. عمر متوسط نوع B برابر ۱۸۷۵ ساعت و انحراف معیار آن ۳۱۰ ساعت است. کدام یک از این دو نوع لامپ تصویر دارای پراکندگی نسبی بیشتری است.

$$C \cdot V_A = \frac{S_A}{\bar{X}_A} \times ۱۰۰ = \frac{۲۸۰}{۱۴۹۵} \times ۱۰۰ = \%۱۸/۷$$

$$C \cdot V_B = \frac{S_B}{\bar{X}_B} \times ۱۰۰ = \frac{۳۱۰}{۱۸۷۵} = \%۱۶/۵۳$$

بنابراین مشاهده می‌شود که لامپ تصویر A دارای پرکندگی بیشتری است.

مثال ۱۶. توزیع فروانی قد و وزن ۱۰۰ نفر در جدول زیر داده شده، این دو صفت ر از لحاظ پراکندگی با هم مقایسه کنید.

x_i	f_i	u_i	f_iu_i	$f_iu_i^۲$
۱۶۵	۱۰	−۲	−۲۰	۴۰
۱۷۰	۲۰	−۱	−۲۰	۲۰
۱۷۵	۴۰	۰	۰	۰
۱۸۰	۲۰	۱	۲۰	۲۰
۱۸۵	۱۰	۲	۲۰	۴۰
	۱۰۰		۰	۱۲۰

$$\bar{x} = C\bar{u} + A = ۵ \times ۰ + ۱۷۵ = ۱۷۵$$

$$V(U) = \frac{۱}{۱۰۰}\left[۱۲۰ - \frac{۱}{۱۰۰}(۰)^۲\right] = ۱/۲$$

$$V(X) = C^۲V(U) = ۵^۲ \times ۱/۲ = ۳۰$$

$$\Longrightarrow S_X = \sqrt{۳۰} = ۵/۴۷$$

$$C \cdot V_X = \frac{S_X}{\overline{X}} \times ۱۰۰ = \frac{۵/۴۷}{۱۷۵} \times ۱۰۰ = ٪۳/۱۵$$

$\Downarrow$

پراکندگی Y از پراکندگی X بیشتر ست.

y_i	f_i	u_i	f_iu_i	$f_iu_i^۲$
۴۰	۲۰	−۲	−۴۰	۸۰
۵۰	۱۰	−۱	−۱۰	۱۰
۶۰	۲۰	۰	۰	۰
۷۰	۳۰	۱	۳۰	۳۰
۸۰	۲۰	۲	۴۰	۸۰
	۱۰۰		۲۰	۲۰۰

$$\overline{y} = C\overline{u} + A = ۱۰ \times \frac{۲۰}{۱۰۰} + ۶۰ = ۶۲$$

$$V(U) = \frac{۱}{۱۰۰}\left[۲۰۰ - \frac{۱}{۱۰۰}(۲۰)^۲\right] = ۱/۹۶$$

$$V(Y) = C^۲V(U) = ۱۰۰ \times ۱/۹۶ = ۱۹۶$$

$$\Longrightarrow S_Y = ۱۴ \quad C \cdot V_Y = \frac{۱۴}{۶۲} \times ۱۰۰ = ٪۲۲/۵$$

گشتاور و گشتاور مرکزی داده‌ها:

گشتاور مرتبه r ام $(r \in N)$ داده‌های $x_1, x_2, \dots, x_k$ با فراوانی‌های $f_1, f_2, \dots, f_k$ را با m'_r نمایش می‌دهند و به این صورت تعریف می‌شود:

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} f_i X_i^r$$

گشتاور مرکزی رتبه r ام داده های بالا را با m_r نشان داده و تعریف می کنند:

$$m_r = \frac{1}{n}\sum_{i=1}^{k} f_i(\bar{X}_i - \bar{X})^r$$

واضح است که

- m'_1 برابر $\bar{x}$ و m_1 برابر صفر و m_2 برابر σ^2 می باشد.
- اگر داده‌ها نسبت به میانگین، متقارن باشند، گشتاورهای مرکزی مرتبه فرد آن ها برابر صفر هستند.

چولگی: میزان عدم تقارن منحنی فراوانی را چولگی می‌نامند. هر کدام از ملاک‌های زیر را می‌توان به عنوان معیار چولگی به کار برد.

- ضریب چولگی پیرسن بر حسب میانه به صورت زیر است.

$$(s.k)_1 = \frac{3\left(\bar{X} - md\right)}{s}$$

- ضریب چولگی پیرسن بر حسب مد به شکل زیر است

$$(s.k)_2 = \frac{\bar{X} - Mo}{s}$$

- ضریب چولگی پیرسن بر حسب چارک ها به صورت زیر است.

$$(s.k)_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

- ضریب چولگی پیرسن بر حسب صدک ها به صورت زیر است.

$$(s.k.)_H = \frac{H_{90} - 2H_{50} + H_{10}}{H_{90} - H_{10}}$$

- ضریب گشتاوری چولگی به شکل می‌باشد

$$g = \frac{m_3}{s^3}$$

مثال ۲۵. ضریب چولگی پیرسن را برای جدول توزیع فراوانی زیر به دست آورید.

حدود طبقات	x_i	f_i
۵۰ − ۵۹٫۹۹	۵۵	۸
۶۰ − ۶۹٫۹۹	۶۵	۱۰
۷۰ − ۷۹٫۹۹ [69.995, 79.995)	۷۵	۱۶
۸۰ − ۸۹٫۹۹	۸۵	۱۴
۹۰ − ۹۹٫۹۹	۹۵	۱۰
۱۰۰ − ۱۰۹٫۹۹	۱۰۵	۵
۱۱۰ − ۱۱۹٫۹۹	۱۱۵	۲

حل: $\Rightarrow$

x_i	f_i	u_i	f_iu_i	$f_iu_i^2$	F_c
۵۵	۸	−۲	−۱۶	۳۲	۸
۶۵	۱۰	−۱	−۱۰	۱۰	۱۸
۷۵	۱۶	۰	۰	۰	۳۴
۸۵	۱۴	۱	۱۴	۱۴	۴۸
۹۵	۱۰	۲	۲۰	۴۰	۵۸
۱۰۵	۵	۳	۱۵	۴۵	۶۳
۱۱۵	۲	۴	۸	۳۲	۶۵
	۶۵		۳۱	۱۷۳	

$$\bar{x} = A + c\bar{u} = A + c\frac{۱}{n}\sum f_iu_i = ۷۵ + ۱۰ \times \frac{۳۱}{۶۵} = ۷۹٫۷۷$$

$$S^2 = I^2\frac{1}{n-1}\left(\sum_{i=1}^{n} f_id_i^2 - \frac{\left(\sum_{i=1}^{n} f_id_i\right)^2}{n}\right) = 10^2\frac{1}{65-1}\left(173 - \frac{31^2}{64}\right) = 15.6^2 \Rightarrow s = 15.6$$

$$M_d = ۶۹٫۹۹۵ + \frac{۳۲٫۵ - ۱۸}{۱۶} \times ۱۰ = ۷۹٫۰۶ \qquad M_o = ۶۹٫۹۹۵ + \frac{۶}{۶ + ۲} \times ۱۰ = ۷۷٫۵$$

ضریب چولگی اول پیرسن

$$S_k = \frac{\text{میانگین} - \text{مد}}{\text{انحراف معیار}} = \frac{۷۹٫۷۷ - ۷۷٫۵۰}{۱۵٫۶۰}$$

$$= ۰٫۱۴۴۹$$

ضریب چولگی دوم پیرسن

$$S_k = \frac{۳\,(\text{میانگین} - \text{میانه})}{\text{انحراف معیار}}$$

$$= \frac{۳(۷۹٫۷۷ - ۷۹٫۰۶)}{۱۵٫۶۰} = ۰٫۱۳۶۵$$

چون ضرایب چولگی، مثبت هستند، توزیع دارای چولگی مثبت می‌باشد. ضمناً توجه کنید

میانگین > میانه > مد

مثال ۱۷. برای طول قد ۱۰۰ دانشجو که جدول توزیع فراوانی آن ضریب چولگی بر حسب چارک‌ها را محاسبه کنید.

حدود طبقات	f_i	F_c	
۶۰ − ۶۲	۵	۵	۵۹/۵ − ۶۲/۵
۶۳ ۶۵	۱۸	۲۳	۶۲/۵ − ۶۵.۵
۶۶ − ۶۸	۴۲	۶۵	۶۵/۵ − ۶۸/۵
۶۹ − ۷۱	۲۷	۹۲	۶۸/۵ − ۷۱/۵
۷۲ − ۷۴	۸	۱۰۰	۷۱/۵ − ۱۴/۵

چارک اول: $np = ۱۰۰ \times \frac{۱}{۴} = ۲۵$

$$Q_1 = ۶۵٫۵ + \frac{۲۵ - ۲۳}{۴۲} \times ۳ = ۶۵٫۶۴$$

بری محاسبه $Q_2 = M_d$

$$Q_2 = M_d = ۵٫۵ + \frac{۵۰ - ۲۳}{۴۲} \times ۳ = ۶۷٫۴۳$$

چارک سوم: $np = ۱۰۰ \times \frac{۳}{۴} = ۷۵$

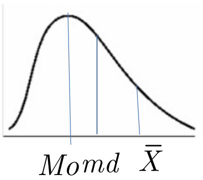
$$Q_3 = ۶۸٫۵ + \frac{۷۵ - ۶۵}{۲۷} \times ۳ = ۶۹٫۶۱$$

$$H_h = L + \frac{\frac{nh}{100} - f_{c(m-1)}}{f_m}(U - L)$$

$$S_Q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{Q_3 + Q_1 - ۲Q_2}{Q_3 - Q_1}$$

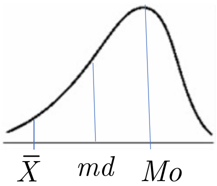
$$= \frac{۶۹٫۶۱ + ۶۵٫۶۴ - ۲ \times ۶۷٫۴۳}{۶۹٫۶۱ - ۶۷٫۴۳} = \frac{۰٫۳۹}{۲٫۱۸} = ۰٫۱۷۹$$

فرم‌های توزیع های فراوانی: • چولگی مثبت • چولگی منفی • متقارن



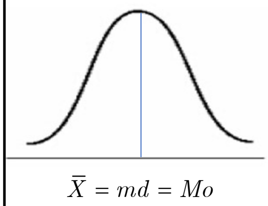
توزیع با چولگی مثبت:

در یک توزیع با چولگی مثبت بیشتر مقادیر داده‌ها در سمت چپ میانگین قرار دارند و دم این توزیع به سمت راست است . بعلاوه میانگین در طرف راست میانه است و نما در سمت چپ میانه است



توزیع با چولگی منفی:

در یک توزیع با چولگی منفی بیشتر مقادیر داده‌ها در طرف راست میانگین قرار دارد و دم توزیع به سمت چپ است به علاوه میانگین در طرف چپ میانه و نما در طرف راست میانه است.



توزیع متقارن:

در یک توزیع متقارن، مقادیر داده‌ها بطور یکسان در دو طرف میانگین توزیع شده‌اند. همچنین وقتی توزیع تک نمایی باشد، میانگین ، میانه و نما با هم برابرند و در مرکز توزیع قرار دارند. در صورتی که داده‌ها نسبت به میانگین متقارن باشند، ضرایب بالا همگی صفرند

یادداشت: هرگاه میزان چولگی خفیف باشد ، بین میانگین میانه و مد رابطه تقریبی زیر(رابطه تجربی پیرسون) برقرار می‌باشد.

$3(\bar{x}-m)=\bar{x}-Mo$ (میانه − میانگین)۳ ≃ مد − میانگین

مثال ۳۶. جدول توزیع فراوانی میزان بارندگی در چند سال گذشته بر حسب میلی‌متر به صورت زیر داده شده است. آیا توزیع متقارن است؟

میزان بارندگی	f_i
۰٫۵ − ۵۰٫۵	۱۵
۵۰٫۵ − ۱۰۰٫۵	۱۷
۱۰۰٫۵ − ۱۵۰٫۵	۱۱
۱۵۰٫۵ − ۲۰۰٫۵	۱۳
۲۰۰٫۵ − ۲۵۰٫۵	۱۴
۲۵۰٫۵ − ۳۰۰٫۵	۱۰
	۸۰

حل: برای بررسی تقارن یا عدم تقارن توزیع، باید میانگین، میانه و مد را محاسبه نمود.

$$\bar{u}=\frac{۱}{n}\sum f_i u_i=\frac{-۵۶}{۸۰}=-۰٫۷ \quad \bar{x}=c\bar{u}+A=۵۰(-۰٫۷)+۱۷۵٫۵=۱۴۰٫۵$$

میزان بارندگی	f_i	x_i	u_i	$f_i u_i$	F_c
۰٫۵ − ۵۰٫۵	۱۵	۲۵٫۵	−۳	−۴۵	۱۵
۵۰٫۵ − ۱۰۰٫۵	۱۷	۷۵٫۵	−۲	−۳۴	۳۲
۱۰۰٫۵ − ۱۵۰٫۵	۱۱	۱۲۵٫۵	−۱	−۱۱	۴۳
۱۵۰٫۵ − ۲۰۰٫۵	۱۳	۱۷۵٫۵	۰	۰	۵۶
۲۰۰٫۵ − ۲۵۰٫۵	۱۴	۲۲۵٫۵	۱	۱۴	۷۰
۲۵۰٫۵ − ۳۰۰٫۵	۱۰	۲۷۵٫۵	۲	۲۰	۸۰
	۸۰			−۵۶	

$$M_o=۵۰٫۵+\frac{۲}{۲+۶}\times ۵۰=۶۳$$

$$M_d=۱۰۰٫۵+\frac{۴۰-۳۲}{۱۱}\times ۵۰=۱۳۶٫۸۶$$

(میانه − میانگین)۳ ≃ مد − میانگین

$$3(\bar{x}-m)=3(140.5-136.9)=10.8$$

$$\bar{x}-Mo=140.5-63=77.5$$

چون مقدار میانه، بین مد و میانگین قرار دارد، لذا توزیع دارای چولگی می‌باشد.

مثال ۳۷. جدول توزیع فراوانی زیر، مربوط به طول عمر لامپ‌های ساخته شده توسط یک کارخانه می‌باشد. آیا توزیع متقارن است؟

حدود طبقات	f_i	x_i	u_i	f_iu_i	F_c
۴۰ − ۴۲٫۹	۲	۴۱٫۴۵	−۳	−۶	۲
۴۳ − ۴۵٫۹	۴	۴۴٫۴۵	−۲	−۸	۶
۴۶ − ۴۸٫۹	۲۶	۴۷٫۴۵	−۱	−۲۶	۳۲
۴۹ − ۵۱٫۹	۴۷	۵۰٫۴۵	۰	۰	۷۹
۵۲ − ۵۴٫۹	۱۵	۵۳٫۵۴	۱	۱۵	۹۴
۵۵ − ۵۷٫۹	۶	۵۶٫۴۵	۲	۱۲	۱۰۰
	۱۰۰			−۱۳	

$$\bar{u} = \frac{۱}{n}\sum f_iu_i = \frac{-۱۳}{۱۰۰} = -۰٫۱۳$$

$$\bar{x} = c\bar{u} + A = ۳(-۰٫۱۳) + ۵۰٫۴۵ = ۵۰٫۰۶$$

$$M_o = ۴۸٫۹۵ + \frac{۲۱}{۲۱+۳۲} \times ۳ \simeq ۵۰٫۱۴$$

$$M_d = ۴۸٫۹۵ + \frac{۵۰ - ۳۲}{۴۷} \times ۳ \simeq ۵۰٫۱$$

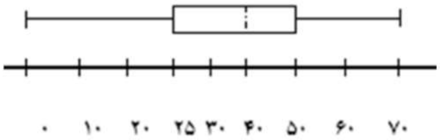
ملاحظه می‌کنید که این سه عدد، تقریبا نزدیک به هم هستند و این خود دلیل بر متقارن بودن توزیع می‌باشد. از طرفی رابطه تجربی نیز برقرار است.

$$3(\bar{x} - m) = \bar{x} - Mo$$

$$۵۰٫۰۶ - ۵۰٫۱۴ \simeq ۳(۵۰٫۰۶ - ۵۰٫۱)$$

تمرینات صفحه 131 الی 136 حل شود

مثال : با توجه به نمودار جعبه‌ای زیر ضریب چولگی چارکی داده‌ها و نوع توزیع آن‌ها کدام است؟



۱) ۰.۲ – و چوله به چپ　　　۳) ۰.۴– و چوله به چپ

۲) ۰.۴ و چوله به راست　　　۴) ۰.۲ و چوله به راست

پاسخ : گزینه « 1 » صحیح است زیرا:

$$(s.k)_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{50 - 2\times 40 + 25}{50 - 25} = -0.2$$

کشیدگی میزان کشیدگی یا پخی منحنی فراوانی را نسبت به منحنی نرمال استاندارد، کشیدگی آن می نامند و با معیار $k = \frac{m_4}{m_2^2} - 3$ اندازه‌گیری می‌کنند.

k را ضریب گشتاوری کشیدگی می‌نامند.

k برای منحنی نرمال استاندارد برابر صفر است.

بر حسب آن که k مثبت یا منفی باشد، منحنی فراوانی کشیده‌تر از نرمال یا پخ تر از آن می باشد

اگر k نزدیک صفر باشد، کشیدگی منحنی فراوانی، طبیعی است.

مثال: در جامعه‌ای با حجم $N = 100$ کمیت‌های

$$\sum_{i=1}^{k} n_i(\hat{X}_i - \overline{X})^2 = 2500, \quad \sum_{i=1}^{k} n_i(\hat{X}_i - \overline{X})^4 = 190000$$

به دست آمده است، کدام مورد تفسیر کشیدگی توزیع است؟

۱ توزیع از توزیع نرمال کشیده‌تر است.

۲ توزیع از توزیع نرمال کوتاه‌تر است.

۳ کشیدگی توزیع به اندازه نرمال است

۴ کشیدگی توزیع غیر قابل اغماض است.

پاسخ :

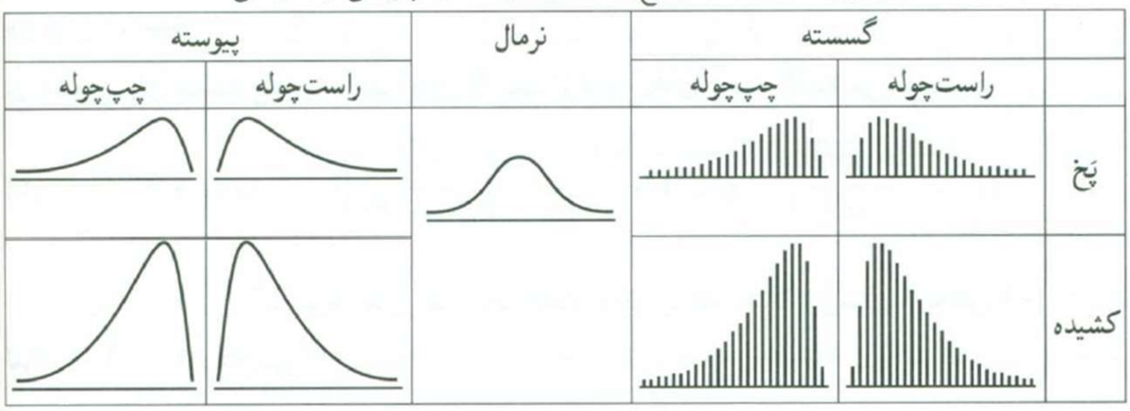
$$m_4 = \frac{1}{n}\sum_{i=1}^{k} n_i(\hat{X}_i - \overline{X})^4 = \frac{190000}{100} = 1900$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(\hat{X}_i - \overline{X})^2 = \frac{2500}{100} = 25$$

$$k = \frac{m_4}{\sigma_X^4} - 3 = \frac{1900}{25^2} - 3 = 0.04 \quad \Rightarrow$$ گزینه ۳ صحیح است

رده‌بندی توزیع‌های احتمال در قیاس با توزیع نرمال

جدول ۱ رده‌بندی توزیع‌های آماری برحسب چولگی و کشیدگی



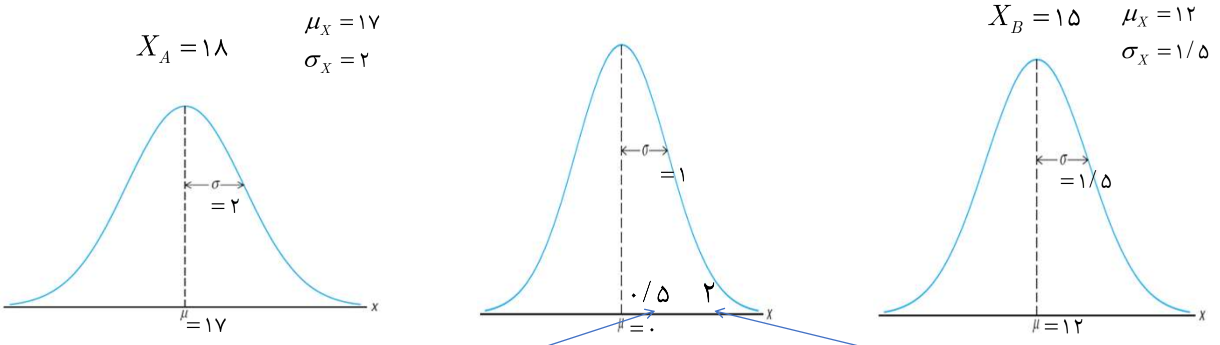
نمره Z یا نمره استاندارد شده:

استاد الف

استاد ب

$X_A = ۱۸$ $\mu_X = ۱۷$ $\sigma_X = ۲$

$X_B = ۱۵$ $\mu_X = ۱۲$ $\sigma_X = ۱/۵$



$$z_A = \frac{۱۸-۱۷}{۲} = ۰/۵$$

$$Z = \frac{X-\mu_X}{\sigma_X}$$

$\mu_Z = ۰$

$\sigma_Z = ۱$

$$z_B = \frac{۱۵-۱۲}{۱/۵} = ۲$$

نمره Z یا نمره استاندارد شده:

نمره Z برای یک مقدار نمونه در یک مجموعه داده ها با کم کردن میانگین و تقسیم داده‌ها بر انحراف معیار بدست می‌آید.

اگر $\bar{x}$ و S بترتیب میانگین و انحراف معیار نمونه‌ای باشند آنگاه $Z=\dfrac{x-\bar{x}}{S}$

اگر μ و σ بترتیب میانگین و انحراف معیار جامعه باشند آنگاه $Z=\dfrac{x-\mu}{\sigma}$

نکته:

- نمره Z تعداد انحراف معیارهایی است که یک مقدار داده ها بالای (نمره Z مثبت) یا پائین (نمره Z منفی) میانگین برای این مجموعه داده‌ها قرار می گیرد.
- نمره Z از مقدار داده دور افتاده در مجموعه داده ها تاثیر پذیر است زیرا این مقدار دور افتاده (مقدار بسیار کوچک یا بسیار بزرگ) بطور مستقیم میانگین یا انحراف معیار را تحت تاثیر قرار می دهد.
- متغیر تصادفی استاندارد شده دارای میانگین صفر و واریانس یک و بی بعد است
- اصولاً Z نمره ایست که نشان می‌دهد یک داده، چند برابر انحراف معیار از میانگین فاصله دارد.

نشان دهید

- متغیر تصادفی استاندارد شده دارای میانگین صفر و واریانس یک و بی بعد است

$$\bar{z}=\frac{1}{n}\sum f_i\left(\frac{x_i-\bar{x}}{s}\right)=\frac{1}{s}\times\frac{1}{n}\sum f_i(x-\bar{x})=\frac{1}{s}\left[\frac{1}{n}\sum f_i x_i-\frac{\bar{x}\sum f_i}{n}\right]=\frac{1}{s}[\bar{x}-\bar{x}]=0$$

$$\left(\sum f_i=n\right)$$

$$V(Z)=\frac{1}{n}\sum f_i\left(\frac{x_i-\bar{x}}{s}\right)^2=\frac{1}{s^2}\times\frac{1}{n}\sum f_i(x_i-\bar{x})^2=\frac{1}{s^2}\times s^2=1$$

مثال: نمره Z برای مقدار ۱۴ در مجموعه داده های نمونه

$$3, 8, 6, 14, 4, 12, 7, 10$$

چیست؟

حل: میانگین و واریانس نمونه‌ای برابر است با

$$\bar{x} = \frac{3+8+\cdots+10}{8} = 8,$$

$$s^2 = \frac{(3-8)^2+(8-8)^2+\cdots+(10-8)^2}{8-1} = 3.8173^2$$

$$z = \frac{14-8}{3/81} = 1.57$$

بنابراین مقدار ۱۴، ۱/۵۷ انحراف معیار از میانگین ۸ بیشتر است.

مثال: نمره Z برای مقدار ۹۵ درمجموعه داده‌های نمونه

96 114 100 97 101 102 92 95 90

چیست؟

حل:

$$\bar{x} = \frac{96+\cdots+90}{9} = 99.33,$$

$$s^2 = \frac{(96-99.33)^2+\cdots+(90-99.33)^2}{9-1} = 6.59^2$$

$$z = \frac{95-99.33}{6.59} = -0.66$$

مثال ۲۳. دانشجویی در متحان درس اقتصاد ۸۴ گرفته است که میانگین نمرات ۷۶ و انحرف معیار نمرات ۱۰ بوده است. وی در امتحان درس آمار که دارای میانگین ۸۲ و نحراف معیار ۱۶ بوده است نمره ۹۰ گرفته است درکدام درس نمره‌ی وی به نسبت بهتر بوده است.

حل:

نمره استاندارد شده اقتصاد

$$Z = \frac{X-\overline{X}}{S} \Longrightarrow Z_A = \frac{۸۴-۷۶}{۱۰} = ۰/۸$$

نمره استاند رد شده آمار

$$Z = \frac{X-X}{S} \Longrightarrow Z_B = \frac{۹۰-۸۲}{۱۶} = ۰/۵$$

بنابراین دانشجوی مورد نظر نمره‌ای معادل ۰/۸ انحرف معیار بالاتر از میانگین نمرات اقتصاد گرفته است ولی نمرهٔ آمار وی فقط ۰/۵ انحراف معیار بالاتر از میانگین نمرات آمار است پس در مقام مقایسه وی به نسبت نمره بهتری در اقتصاد آورده است تا آمار. □

تمرینات صفحه 185 الی 191 حل شود

نرم‌افزار $SPSS$

https://www.spss-iran.com/download/#_SPSS_26

مثال ۱.۳ جدول زیر چند سطر از اطلاعات کارکنان یک بخش را نشان می‌دهد

ردیف	نام	سن	جنس	میزان تحصیلات	شغل
۱	اکبری، حسن	۴۲	مرد	دیپلم	تکنسین
۲	بهادری، پروین	۳۷	زن	دکتری	پزشک
۳	جواهری، لاله	۴۰	زن	کارشناسی‌ارشد	سرپرستار
۴	دایی، فاطمه	۳۵	زن	کارشناسی	پرستار

هر سطر داده‌های یک فرد را نشان می‌دهد. هر ستون مقدارهای یک متغیر را بیان می‌کند. به غیر از نام کارکنان، چهار متغیر سن، جنس، میزان تحصیلات، و شغل ثبت شده‌اند. جنس و شغل متغیرهای رسته‌ای‌اند. سن و میزان تحصیلات (تعداد سال‌هایی که درس خوانده‌اند) متغیرهای کمّی‌اند. اغلب نرم‌افزارهای آماری از این قالب برای واردکردن داده‌ها استفاده می‌کنند. ■

تشکیل پرونده‌ی داده‌ها در نرم‌افزار $SPSS$.

پس از نصب نسخه‌ای از نرم‌افزار $SPSS$ در رایانه، عملیات زیر را برای تشکیل پرونده و وارد کردن داده‌ها به ترتیب انجام می‌دهیم.

۱. نرم‌افزار $SPSS$ را فرامی‌خوانیم.

۲. در پنجره‌ای که ظاهر می‌شود، گزینه‌ی $Type\ in\ Data$ (وارد کردن داده‌ها) را علامت زده روی Ok کلیک می‌کنیم.

۳. در صفحه‌ای که ظاهر می‌شود، درگوشه‌ی جنوب‌غربی، دو گزینه داریم.
$Variable\ view$ (نمای متغیر) برای تعریف متغیرها و $Data\ view$ (نمای داده‌ها) برای وارد کردن داده‌ها روی $Variable\ view$ کلیک می‌کنیم.

۴. در صفحه‌ای که ظاهر می‌شود و بسیار شبیه به جدول مثال ۱.۳ است از چپ به راست موارد زیر را می‌بینیم:
نام متغیر ($Name$) که به اختیار پژوهشگر نامی انتخاب می‌شود.
نوع متغیر ($Type$) که با کلیک کردن در آن خانه پنجره‌ای با انواع متغیرها ظاهر می‌شود و بسته به ماهیت متغیر یکی از آن‌ها را علامت می‌زنیم.

(نوع $Numeric$ (عددی) برای متغیرهای کمّی و نوع String (رشته‌ای) برای متغیرهای رسته‌ای است. در هر پنجره پس از گزینش گزینه‌ی مورد نظر بسته مورد روی Ok یا $Continue$ کلیک می‌کنیم.)

پهنا ($Width$) برای تعداد رقم‌های متغیر است که با کلیک کردن در آن خانه می‌توان تعداد رقم‌ها را کم یا زیاد کرد.

رقم‌های اعشاری ($Decimals$) که با کلیک کردن کم و زیاد می‌شوند.

برچسب ($Label$) برای دادن نشانه یا نامی خاص به متغیر است.

مقدارها ($Values$) (این گزینه برای متغیرهای رسته‌ای کاربرد دارد. با کلیک کردن روی آن پنجره‌ای ظاهر می‌شود که در آن در سطر $Value$ کد حالت اول متغیر رسته‌ای را مثلاً با ۱ تعریف و در سطر $Label$ برچسبی برای آن در نظر می‌گیریم. سپس روی Add کلیک می‌کنیم تا کد و برچسب تعریف شده به مستطیل بزرگ مقابل منتقل شود. مثلاً خواهیم داشت "CC" = ۱. این کار به تعداد حالت‌های متغیر تکرار می‌شود. درباره‌ی متغیرهای کمّی از این گزینه استفاده نمی‌شود.)

داده‌های گمشده ($Missing$). با کلیک کردن روی آن پنجره‌ای را خواهیم دید که در آن گزینه‌های بدون مقدارهای گمشده ($No\ missing\ values$) و دو گزینه‌ی دیگر را داریم. این گزینه‌ها برای پرونده‌های بزرگ که بخواهند در آن‌ها برخی داده‌ها را علامت‌گذاری کنند و از تحلیل کنار بگذارند مصرف دارند. در مثال‌ها و تمرین‌ها معمولاً با چنین مسأله‌ای روبه‌رو نیستیم. از این‌رو همان گزینه‌ی اول بدون داده‌ی گمشده را انتخاب می‌کنیم.
ستون‌ها ($Columns$) تعداد ستون‌های اختصاص‌یافته به متغیر را نشان می‌دهد و قابل تنظیم است.

تراز ($Align$) نوع چینش داده‌ها در ستون را به‌صورت راست‌چین، مرکزی، و چپ‌چین نشان می‌دهد که به اختیار پژوهشگر است.

مقیاس ($Measure$) نوع مقیاس اسمی، ترتیبی، رتبه‌ای، و نسبتی را نشان می‌دهد. (برای هر متغیر مقیاس مناسب آن انتخاب می‌شود و این کار با کلیک کردن در آن خانه و آوردن فهرست مقیاس‌ها انجام می‌یابد.) پس از تعریف خصوصیات هر متغیر در یک سطر از صفحه‌ی نمای متغیرها با کلیک کردن روی $Data\ view$ به صفحه‌ی نمای داده‌ها می‌رویم. اگر درست عمل کرده باشیم، در این صفحه به تعداد متغیرهای تعریف شده و با همان نام‌ها ستون‌هایی از متغیرها را خواهیم دید. اکنون با دو وضعیت مواجه می‌شویم

۱ـ اگر متغیرمان از نوع عددی است، مقدارهای عددی آن را در سطرهای مختلف وارد می‌کنیم.

۲ـ اگر متغیرمان از نوع رسته‌ای است، در هر سطر کد مربوط به آن فرد را وارد می‌کنیم. مثلاً اگر متغیر جنس (زن یا مرد) را داشته باشیم به تعداد زنان کد ۱ و به تعداد مردان کد ۲ را در سطرها وارد می‌کنیم.

بعد از تکمیل صفحه‌ی نمای داده‌ها، لازم است پرونده‌ی داده‌ها را تحت نامی معین ذخیره کنیم تا قابل بازیابی برای تحلیل‌های بعدی باشد.

۵. فرض کنید پرونده‌ای به نام $Blood.data.sav$ در $My\ Documents$ ذخیره شده است. پس از فراخواندن $SPSS$، در پنجره‌ی ظاهر شده، گزینه‌ی $Open\ an\ existing\ data\ source$ (منبعی موجودی از داده‌ها را باز کنید) را برگزیده روی Ok کلیک می‌کنیم. در پنجره‌ای که باز می‌شود روی نام $Blood.data.sav$ کلیک می‌کنیم، پرونده‌ی داده‌ها ظاهر می‌شود که آماده‌ی تحلیل است.

۶. روش تحلیل داده‌ها را در هر مورد خاص در جای خود توضیح خواهیم داد. نکته‌های مشترک آن است که در نوار بالایی صفحه، گزینه‌ی $Transform$ برای تبدیل متغیرها و تعریف متغیرهای جدید به‌کار می‌رود، $Analyze$ برای تحلیل‌های آماری است، و $Graphs$ برای رسم نمودارهاست.

محاسبه‌ی آماره‌های توصیفی با $SPSS$.

برای محاسبه‌ی آماره‌های خلاصه‌ی هر توزیع مسیر زیر را طی کنید

$$Analyze \longrightarrow Descriptive\ statistics \longrightarrow 123Frequencies$$

در پنجره‌ی $Frequencies$، متغیر (یا متغیرهای) مورد نظر را به خانه‌ی $Variables$ بفرستید. گزینه‌ی $statistics$ آماره‌های مختلف را حساب می‌کند. در پنجره‌ای که باز می‌شود در قسمت $Percentile\ values$ (مقدارهای صدکی) خانه‌ی $Quartiles$ (چارک‌ها) را تیک بزنید و تعداد صدک‌های لازم را با تیک زدن و درج تعداد لازم در گزینه‌ی $Cut\ points\ for\ \square\ equal\ group$ مشخص کنید. مثلاً ۴ برای چارک، ۱۰ برای دهک و ۱۰۰ برای صدک و غیره.

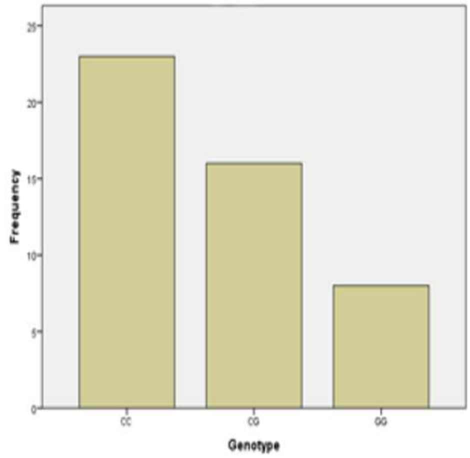
در قسمت $Centraltendency$ (تمایل به مرکز، میانگین، میانه، و مد را تیک بزنید.) در قسمت $Dispersion$ (پراکندگی) آماره‌های لازم را تیک بزنید. با کلیک روی $Continue$ به پنجره‌ی اصلی برگردید. در گزینه‌ی $Charts$ (نمودارها) می‌توانید نمودار مطلوب خود را بخواهید که برحسب بسامدها یا درصدها رسم شود. با $Continue$ به پنجره‌ی اصلی برگردید. در پایین پنجره $Display\ frequency\ tables$ را تیک بزنید تا جدول توزیع بسامدی نیز محاسبه شود. در پایان روی Ok کلیک کنید.

مثال ۲.۳ (امامقلی‌پور و همکاران، ۱۳۸۸: ۱۲۳ـ۱۲۹) در بررسی «همراهی چند شکلی پروموتر ژن رزیستین با دیابت نوع ۲» تعداد ۴۷ بیمار مبتلا به دیابت نوع ۲ و ۶۶ فرد سالم را از لحاظ دگره[1]های ژن C و G بررسی کرده‌اند.

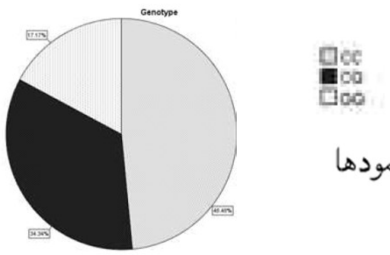
توزیع ژن‌نمود[2] رزیستین در بین نمونه‌ی افراد دیابتی به قرار جدول ۱.۳ است. **نمودار میله‌ای**[3] این داده‌ها در شکل ۱.۳ رسم شده است. این داده‌ها را با **نمودار کلوچه‌ای**[4] نیز می‌توان نمایش داد که در شکل ۲.۳ رسم شده است.

جدول ۱.۳ توزیع بسامدی و درصدی ژن‌نمودها

Genotype					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CC	23	48.9	48.9	48.9
	CG	16	34.0	34.0	83.0
	GG	8	17.0	17.0	100.0
	Total	47	100.0	100.0	



شکل ۱.۳ نمودار میله‌ای توزیع بسامدی ژن‌نمودها



شکل ۲.۳ نمودار کلوچه‌ای توزیع درصدی ژن‌نمودها

را با برچسب خاص آن برای افراد دیابتی تعریف می‌کنیم. کدهای زیر را برای آن در خانه‌ی Values وارد می‌کنیم. $"CC" = 1$، $"CG" = 2$، و $"GG" = 3$. در صفحه‌ی نمای داده‌ها، تعداد ۲۳ فقره عدد یک، ۱۶ فقره عدد ۲، و ۸ فقره عدد ۳ را در ستون ۱ وارد می‌کنیم. پرونده‌ی داده‌ها را با نام معینی ذخیره می‌کنیم.

پس از فراخوان پرونده‌ی داده‌ها، برای رسم شکل ۱.۳ از مسیر

$$Analyze \longrightarrow Descriptive\ statistics \longrightarrow 123Frequencies \longrightarrow$$

متغیر رسته‌ای مورد نظر را به خانه‌ی متغیرها (Variables) می‌بریم. در پنجره‌ای که ظاهر می‌شود گزینه‌های مختلف را داریم. فعلاً با گزینه‌ی **statistics** کاری نداریم. در گزینه‌ی **Charts** نوع نمودار میله‌ای یا کلوچه‌ای را برمی‌گزینیم. در **Chart values** بسامدها یا درصدها را علامت می‌زنیم. با **Continue** به پنجره‌ی اصلی برمی‌گردیم و **Display frequency tables** را تیک می‌زنیم در پایان روی **Ok** کلیک می‌کنیم.

مثال ۵.۳ در آزمایشی مدت زنده‌مانی ۷۲ خوکچه‌ی هندی را پس از تزریق یک باکتری عفونت‌زا اندازه گرفتند. تعداد روزهایی که این خوکچه‌ها زنده ماندند به قرار زیر است.

۰۴

۴۳ ۴۵ ۵۳ ۵۶ ۵۶ ۵۷ ۵۸ ۶۶ ۶۷ ۷۳ ۷۴ ۷۹ ۸۰ ۸۰ ۸۱ ۸۱ ۸۱
۸۲ ۸۳ ۸۳ ۸۴ ۸۸ ۸۹ ۹۱ ۹۱ ۹۲ ۹۲ ۹۷ ۹۹ ۹۹ ۱۰۰ ۱۰۰ ۱۰۱
۱۰۲ ۱۰۲ ۱۰۲ ۱۰۳ ۱۰۴ ۱۰۷ ۱۰۸ ۱۰۹ ۱۱۳ ۱۱۴ ۱۱۸ ۱۲۱ ۱۲۳ ۱۲۶ ۱۲۸
۱۳۷ ۱۳۸ ۱۳۹ ۱۴۴ ۱۴۵ ۱۴۷ ۱۵۶ ۱۶۲ ۱۷۴ ۱۷۸ ۱۷۹ ۱۸۴ ۱۹۱ ۱۹۸
۲۱۱ ۲۱۴ ۲۴۳ ۲۴۹ ۳۲۹ ۳۸۰ ۴۰۳ ۵۱۱ ۵۲۲ ۵۹۸

نمودار ساقه و برگ این داده‌ها را رسم کنید و درباره‌ی شکل، مرکز، پراکنش، تک‌مدی بودن و تقارن آن بحث کنید.

حل.
برای رسم نمودار ساقه و برگ باید مسیر زیر را طی کرد $Analyze \longrightarrow Descriptive\ statistics \longrightarrow Explore$

متغیر مدت زنده مانی را به خانه‌ی Dependent list ببرید. فعلاً با گزینه‌های دیگر جز Plots کاری نداریم. در پنجره‌ی Plots از Descriptives در خانه‌ی Stem-and-leaf تیک بزنید و روی Continue کلیک کنید. در پنجره‌ی Explore در بخش پایینی Plots را انتخاب کنید و روی Ok کلیک کنید.

شکل ۳.۳ نمودار ساقه و برگ داده‌های زنده‌مانی خوکچه‌های هندی

```
SurvTime Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2,00        0 .  44
    28,00        0 .  5555566777888888888889999999
    24,00        1 .  000000000001112222333444
     8,00        1 .  56777899
     4,00        2 .  1144
     6,00 Extremes    (>=329)

 Stem width:      100
 Each leaf:       1 case(s)
```

در این شکل، نرم‌افزار پهنای ساقه را ۱۰۰ فرض کرده یعنی رقم یکان مشاهدات را نادیده گرفته است. بدین ترتیب مثلاً ۴۳ و ۴۴ به ۰۴ تبدیل می‌شوند. مشاهدات بزرگ‌تر از ۳۲۹ را یک‌جا در نظر گرفته است. بسامد هر ساقه در سمت چپ داده شده است. در واقع نمودار ساقه شکسته را داریم.

شکل توزیع نامتقارن است. مرکز آن در حدود ۱۰۰ روز است. پراکنش زیاد دارد زیرا دامنه‌ی تغییرات آن $R = ۵۹۸ - ۴۳ = ۵۵۵$ روز است. تک‌مدی است زیرا یک قله‌ی عمده بیشتر ندارد و راست‌چوله است. ■

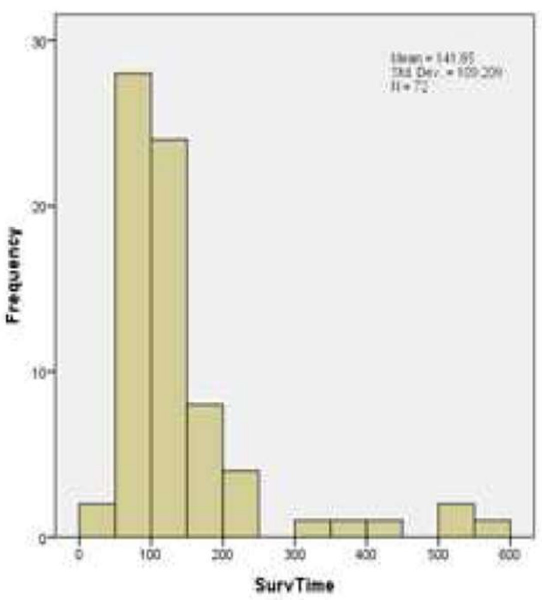
مثال ۶.۳ بافت‌نگار مدت زنده‌مانی خوکچه‌های هندی مثال ۵.۳ را رسم کنید. از روی شکل حاصل درباره‌ی تقارن، مرکز، و پراکنش متغیر اظهارنظر کنید.

حل. رسم بافت‌نگار با دست، به‌ویژه اگر اندازه‌ی نمونه کوچک نباشد، کاری کند و خسته‌کننده است. با استفاده نرم‌افزار به راحتی می‌توانیم بافت‌نگار دلخواه را به‌دست آوریم. برای این کار همان مسیر رسم نمودار مثال ۵.۳ را طی کنید و در پنجره‌ی Plots خانه‌ی Histogram را تیک بزنید. بقیه‌ی مراحل همانند مثال ۵.۳ طی می‌شود. با این ترتیب شکل ۵.۳ را برای بافت‌نگار زنده‌مانی خوکچه‌ها به‌دست می‌آوریم.

از شکل ۴.۳ پیداست که توزیع شدیداً راست‌چوله است. در حالی‌که اکثر خوکچه‌ها تا ۲۰۰ روز زنده می‌مانند، تعداد اندکی از آن‌ها تا ۵۵۰ روز هم باقی می‌مانند. عمر این خوکچه‌ها **دورافتاده** به نظر می‌رسد. مرکز توزیع حدود ۱۴۰ روز است. پراکنش زیاد است و چنان‌که در مثال ۵.۳ دیدیم حدود ۵۵۵ روز است. توزیع تک‌مدی است. این شکل در واقع همان شکل ۳.۳ به‌صورتی دیگر است. ■

رده	تعداد خوکچه‌ها	درصد
۰-۵۰	۲	۲٫۸
۵۱-۱۰۰	۲۸	۳۸٫۹
۱۰۱-۱۵۰	۲۴	۳۳٫۳
۱۵۱-۲۰۰	۸	۱۱٫۱
۲۰۱-۲۵۰	۴	۵٫۶
۲۵۱-۳۰۰	۰	۰
۳۰۱-۳۵۰	۱	۱٫۴
۳۵۱-۴۰۰	۱	۱٫۴
۴۰۱-۴۵۰	۱	۱٫۴
۴۵۱-۵۰۰	۰	۰
۵۰۱-۵۵۰	۲	۲٫۸
۵۵۱-۶۰۰	۱	۱٫۳
کل	۷۲	۱۰۰٫۰

جدول ۱.۳ جدول توزیع بسامدی روزهای زنده‌مانی خوکچه‌های هندی

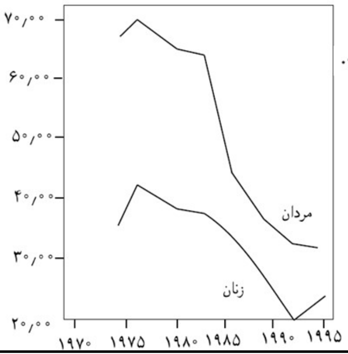


شکل ۴.۳ بافت‌نگار روزهای زنده‌مانی خوکچه‌های هندی پس از تزریق یک باکتری عفونت‌زا

نمودارهای زمانی برای رسم این نمودارها مسیر زیرا را طی کنید.

$$Graphs \longrightarrow Chart\ Builder \longrightarrow Gallery \longrightarrow Line$$

مثال ۷.۳ (هیل و همکاران، (۱۹۹۸): ۲۰۹-۲۱۳) داده‌های شش پیمایش از درصد سیگاری‌های مرد و زن استرالیا را طی سال‌های مختلف گزارش کرده‌اند. نمودار زمانی تغییرات درصد سیگاری‌ها را در شکل ۵.۳ آورده‌ایم.[۱]
نمودارهای مکانی را نیز می‌توان رسم کرد. نمونه‌ای از آن‌ها که مخاطره‌ی نسبی سرطان لب در ایران را نشان می‌دهد در مقاله‌ی (کاوسی، مشکانی و محمدزاده (۲۰۰۹): ۳۴۷_۳۵۹) درج شده است.



از این نمودار پیداست که درصد سیگاری‌ها طی دو دهه‌ی ۱۹۷۰ تا ۱۹۹۵ کاهش چشمگیری داشته است.

این کاهش در مورد مردان بیشتر از زنان است.

برای مرتب کردن داده‌ها از دستور زیر در $SPSS$ استفاده کنید:

$Data \longrightarrow Sort\ Cases \longrightarrow$ متغیر موردنظر را به خانه‌ی Sort by ببرید $\longrightarrow Ascending \longrightarrow Ok$

نمودار جعبه‌ای

برای رسم این نمودار مسیر زیر طی می‌شود.

$$Graphs \longrightarrow Chart\ Builder \longrightarrow Gallery \longrightarrow Box\ Plot$$

شکل نمودار جعبه‌ای تک را به صفحه‌ی مختصات می‌بریم. متغیر موردنظر را با مکان‌نمای رایانه کشیده روی محور عمودی می‌اندازیم. سپس روی Options و گزینه‌ی اندازه‌ی نمودار در خانه‌ی (Chart size) کلیک می‌کنیم. در پایان روی Ok و بعد از آن هم روی Ok کلیک می‌کنیم.

محاسبه‌ی آماره‌های توصیفی با $SPSS$

برای محاسبه‌ی آماره‌های خلاصه‌ی هر توزیع مسیر زیر را طی کنید

$$Analyze \longrightarrow Descriptive\ statistics \longrightarrow 123Frequencies$$

در پنجره‌ی $Frequencies$، متغیر (یا متغیرهای) مورد نظر را به خانه‌ی $Variables$ بفرستید. گزینه‌ی $statistics$ آماره‌های مختلف را حساب می‌کند. در پنجره‌ای که باز می‌شود در قسمت $Percentile\ values$ (مقدارهای صدکی) خانه‌ی $Quartiles$ (چارک‌ها) را تیک بزنید و تعداد صدک‌های لازم را با تیک زدن و درج تعداد لازم در گزینه‌ی $Cut\ points\ for\ \square\ equal\ group$ مشخص کنید. مثلاً ۴ برای چارک، ۱۰ برای دهک و ۱۰۰ برای صدک و غیره.

در قسمت $Centraltendency$ (تمایل به مرکز، میانگین، میانه، و مد را تیک بزنید.) در قسمت $Dispersion$ (پراکندگی) آماره‌های لازم را تیک بزنید. با کلیک روی $Continue$ به پنجره‌ی اصلی برگردید. در گزینه‌ی $Charts$ (نمودارها) می‌توانید نمودار مطلوب خود را بخواهید که برحسب بسامدها یا درصدها رسم شود. با $Continue$ به پنجره‌ی اصلی برگردید. در پایین پنجره $Display\ frequency\ tables$ را تیک بزنید تا جدول توزیع بسامدی نیز محاسبه شود. در پایان روی Ok کلیک کنید.