فصل اول : آمار توصیفی

۱.۲ مقدمه

در فصل اول در تعریف علم آمار ، جمع آوری ، تنظیم و خلاصه سازی ، تحلیل و تعبیر داده ها را وظایف مهم آن نامیدیم خلاصه کردن و توضیح خصوصیات مهم مجموعه داده ها را که معمولاً نمونه تصادفی از جامعه آماری مورد مطالعه است ، آمار توصیفی می نامند . درصورت داشتن در اختیار داشتن یک نمونه مناسب ، می توان انتظار داشت که مشخصات جامعه مورد بررسی تقریباً مشابه با مشخصات نمونه موجود باشند .

شناخت و توصیف داده ها در دو مرحله اصلی انجام می گیرد . مرحله خلاصه کردن که با ارائه جداول و رسم نمودار انجام می پذیرد و مرحله محاسبه شاخصهای عددی که به دو بخش محاسبه شاخص های تمرکز و شاخص های پراکندگی تقسیم می شود .

۲.۲ نمایش و توصیف داده ها توسط نمودار و جداول

۱.۲.۲ جدول های آماری

یک روش برای نمایش داده ها استفاده از جدول است . جدول آماری ، داده ها را با نظمی خاص در چند سطر و ستون ارائه می کند . یکی از پرکاربردترین جدول های آماری ، جدول فراوانی می باشد . هرگاه داده ها را به نحوی تقسیم کرده و آن ها را برحسب فراوانی ها در جدولی تنظیم شوند ، این جدول را جدول فراوانی یا جدول توزیع فراوانی می نامند .

در ستون اول با توجه به مقیاس اندازه گیری متغیر ، اندازه های متغیر یا فواصلی از اندازه های متغیر درج می شود . در ستون دوم نماینده هر رده مشخص می شود . اگر ستون اول به صورت فاصله ای رده بندی شده باشد ، نقطه وسط هر فاصله نماینده آن رده است در غیر اینصورت به هر رده یک عدد طبیعی نسبت داده می شود . در ستون سوم معمولاً فراوانی مشاهده شده (f_i) درج و در ستون چهارم فراوانی نسبی که از تقسیم فراوانی به تعداد کل مشاهدات ($r_i = \frac{f_i}{n}$) به دست می آید ، ثبت می شود . در ستون پنجم و ششم به ترتیب مقدار فراوانی تجمعی مطلق (g_i) و فراوانی نسبی تجمعی ($S_i = \frac{g_i}{n}$) هر رده نوشته می شود .

در تنظیم یک جدول فراوانی باید دقت کرد که طبقات در نظر گرفته شده نباید با یکدیگر تداخل داشته باشند و طبقات باید کاملاً فراگیر باشند به طوری که تمام مشاهدات را در بر بگیرند . تعداد طبقات و حدود طبقات به گونه ای اختیار شوند که تا حد ممکن طبقات خالی ایجاد نشود . معمولاً برای تعیین تعداد طبقات از دستور استورجس

$$k = 1 + 3.3 \log n$$

استفاده می شود که در آن n تعداد داده است و اگر k اعشاری حاصل شود آن را به نزدیکترین واحد معمولاً رو به بالا

گرد می کنند. برای تعیین فاصله در هر رده، برد رده ها را تعیین می کنند که آن را با R نشان می دهند

$$R = \text{بزرگترین مشاهده} - \text{کوچکترین مشاهده} + h$$

که در آن h با توجه به داده ها و به تعداد اعشار داده ها برابر ۱، ۰٫۱، ۰٫۰۱ و ... اختیار می شود.

و پس از محاسبه برد (R) و تعداد طبقات (K)، طول هر رده را از فرمول $C = \frac{R}{K}$ به دست می آورند و معمولا رو به بالا گرد می کنند.

برای تعیین L_1 مرز پایین رده اول از فرمول

$$\text{مرز پایین رده اول}\ (L_1) = \text{کوچکترین مشاهده} - h/2$$

استفاده می کنند.

مثال ۱: فرض کنید مشاهدات زیر حاصل یک آزمایش حاصل شده است جدول فراوانی را برای این داده ها تشکیل دهید

۲٫۳	۷٫۱۱	۶٫۱۳	۶٫۱۵	۹٫۱۵	۴٫۲۸	۱٫۲۹	۲٫۳۷
۸٫۴۱	۳٫۴۴	۹٫۴۳	۵٫۴۹	۸٫۴۹	۲٫۵۱	۶٫۵۱	۵٫۵۲
۷٫۵۳	۹٫۵۳	۰٫۵۴	۹٫۵۶	۸٫۶۳	۴٫۶۶	۶٫۶۸	۱٫۷۰
۰٫۷۱	۵٫۷۴	۷٫۷۶	۸٫۷۷	۹٫۸	۰٫۸۴	۷٫۸۵	۵٫۸۶
۹٫۸	۳٫۸۹	۴٫۹۱	۶٫۹۱	۸٫۹۸	۴٫۱	۲٫۱	۳٫۱۲

نکته: داده ها را در هفت رده تقسیم کنید. ($K = 1 + 3.3\log 40 = 6.3 \cong 7$)

حل: ابتدا برد را تعیین می کنیم

$$R = 9.8 - 0.54 + 0.01 = 9.27$$

مرز پایین اولین رده: $L_1 = 0.54 - \frac{0.01}{2} = 0.535$

طول رده: $C = \frac{9.27}{7} = 1.33$

بنابراین جدول فراوانی به صورت زیر تشکیل می شود:

0.535

رده	نماینده (x_i)	فراوانی (f_i)	فراوانی نسبی (r_i)	فراوانی تجمعی (F_i)	فراوانی نسبی تجمعی (R_i)
۰٫۵۳۵ – ۱٫۸۶۵	۱٫۲	۵	۰٫۱۲۵	۵	۰٫۱۲۵
۱٫۸۶۵ – ۳٫۱۹۵	۲٫۵۳	۴	۰٫۱	۹	۰٫۲۲۵
۳٫۱۹۵ – ۴٫۵۲۵	۳٫۸۶	۶	۰٫۱۵	۱۵	۰٫۳۷۵
۴٫۵۲۵ – ۵٫۸۵۵	۵٫۱۹	۵	۰٫۱۲۵	۲۰	۰٫۵
۵٫۸۵۵ – ۷٫۱۸۵	۶٫۵۲	۶	۰٫۱۵	۲۶	۰٫۶۵
۷٫۱۸۵ – ۸٫۵۱۵	۷٫۸۵	۵	۰٫۱۲۵	۳۱	۰٫۷۷۵
۸٫۵۱۵ – ۹٫۸۴۵	۹٫۱۸	۹	۰٫۲۲۵	۴۰	۱

مثال ۲: پزشکی ۲۰ بیمار قلبی دارد که گروه خونی آن ها عبارتند از

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

جدول فراوانی این مشاهدات را تشکیل دهید.

گروه	نماینده	فراوانی	فراوانی نسبی	فراوانی تجمعی	فراوانی نسبی تجمعی
A	۱	۶	۰/۳۰	۶	۰/۳
B	2	۳	۰/۱۵	۹	۰/۴۵
AB	۳	۴	۰/۲۰	۱۳	۰/۶۵
O	۴	۷	۰/۳۵	۲۰	۱

۲-۲ نمودارهای آماری

نمایش مشاهدات مطابق قراردادهای خاص به صورت هندسی، نمودار آماری نامیده می‌شود. از انواع نمودارهای معروف می‌توان به نمودار هستوگرام[1]، میله‌ای[2]، دایره‌ای[3] .... اشاره کرد.

Pie chart   Barchart   Histogram

- هستوگرام

هستوگرام یا بافت‌نگار از پرکاربردترین نمودارهای آماری است و برای نمایش هندسی توزیع‌های فراوانی مورد استفاده قرار می‌گیرد. در توزیع‌های فراوانی که دارای طبقات مساوی هستند هستوگرام ترکیبی از مستطیل‌هایی است که عرض آن برابر فاصله طبقات و ارتفاع آن فراوانی یا فراوانی نسبی طبقه است.

برای رسم هستوگرام می‌توان از نرم افزار R استفاده کرد. دستورات مورد نیاز برای مثال ۱ به شرح زیر است.

```
> x <- c(۲/۳, ۷/۱۱, ...., ۳/۱۲۱)
> hist(x)
```

دقت کنید با دستور his(x)، نرم افزار طول دسته را به طور خودکار تعیین می‌کند، اما می‌توان به صورت دستی حدود را مشخص کرد. برای این منظور از nclass برای تعیین تعداد طبقات و یا از breaks حدود رده را مشخص کرد. nclass برای حالتی که بیشتر از ۱۰ طبقه منظور باشد، استفاده می‌شود در غیر اینصورت از breaks استفاده می‌شود.

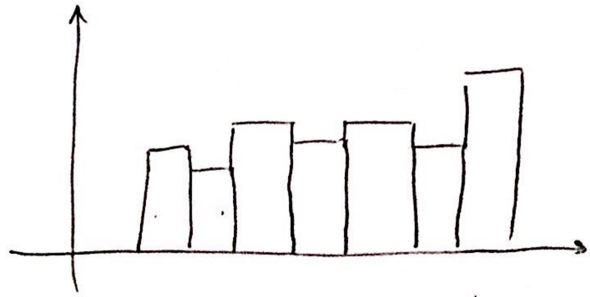
```
>hist(x, nclass =10)
>hist(x, nclass = "sturges")
>hist (x, breaks=c(0.535, ۱/۸۶۵, ۳/۱۹۵, ۴/۵۲۵, ..., ۹/۸۴۵))
```

درصورتی که هستوگرام براساس فراوانی نسبی مدنظر باشد باید prob=T در دستور hist اضافه شود.

```
> hist (x, nclass = 7 , prob=T )
```

نمودار مثال ۱

(با نرم افزار رسم می شود)



نمودار چندبر فراوانی (polygon)

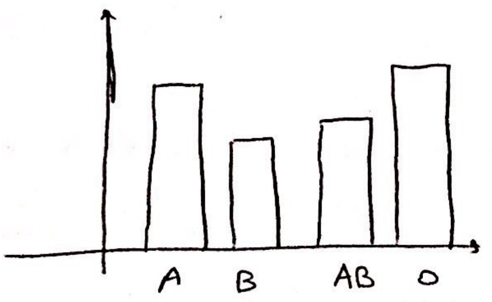
در صورتی که نقاط وسط مستطیلهای هیستوگرام و نقطه‌ای وسط در دو انتهای هیستوگرام با فراوانی صفر را به هم وصل کنیم نمودار چندبر فراوانی حاصل می شود. اگر عرض محور مختصات در نمودار هیستوگرام برابر فراوانی تجمعی یا فراوانی نسبی در نظر بگیریم نمودار فراوانی تجمعی به دست می آید و در صورتی که نقاط وسط مستطیلها در این حالت به هم پیوند داده شوند نمودار چندبر فراوانی تجمعی به دست می آید.

نمودار میله‌ای یا ستونی :

نمودار میله ای برای مشاهدات با ماهیت گسسته رسم می شود و ارتفاع میله تعداد یا فراوانی هر مشاهده را نشان می دهد به عبارت دیگر نمودار میله ای فقط برای مسائلی که به بررسی مشخصات کیفی می پردازند رسم می شود.
دستورات مورد نیاز برای رسم نمودار مثال ۲ با نرم افزار R به صورت

```
> Group <- rep(C( "A" , "B" , "AB" , "O" ) , c( 6 , 3 , 4 , 7 ))
> table (Group)
> bar plot (height = table(Group) , names = C("A","B","AB", "O" ))
```

خروجی نرم افزار

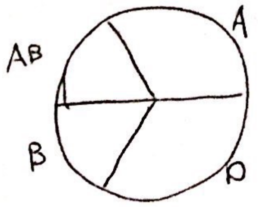


نمودار دایره‌ای :

نمودار دایره‌ای یکی از پرکاربردترین نمودارها برای مشاهدات کیفی است. برای رسم این نمودار دایره را به چند سمت تقسیم می کنند به طوری که سطح هر قطاع بیانگر فراوانی یا درصد فراوانی هر گروه است.

دستورات مورد نیاز برای رسم نمودار دایره‌ای مثال ۲ با نرم افزار R :

```
> pie ( table (Group))
```



۳.۲.۴ـ محاسبه شاخص ها

پس از نمایش مشاهدات توسط نمودارها و جداول می توان با محاسبه چند عدد ~~موقعیت~~ نتایج کلی تری در مورد مشاهدات و در نتیجه مشخص کردن موقعیت و ویژگی صفات مورد بررسی با محاسبه این اعداد امکان پذیر می شود. این اعداد شاخص ها یا معیارها نامیده می شوند و به دو نوع مرکزی و پراکندگی تقسیم می شوند. به عبارت دیگر ~~و~~ خلاصه کردن داده ها توسط این معیارها انجام می شود.

۴.۳.۲ـ معیارهای مرکزی :

مهمترین شاخص های تمرکز میانگین[1] ، میانه[2] و مد[3] (نما) می باشند، که هر یک کاربرد خواص به خصوصی دارند. مزیت میانگین این است که از تمام داده ها برای بدست آوردن آن استفاده می شود اما میانگین تحت تأثیر داده های پرت است به طوری که میانه تحت تأثیر داده های پرت قرار نمی گیرد اما چون از تمام داده ها برای بدست آوردن آن استفاده نمی شود ، معیار قابل قبولی برای بسیاری از بررسی های آماری نیست. یکی دیگر از ایرادهای میانه این است که ممکن است در داده ها نباشد. معمولاً برای داده های اسمی از نما و برای داده های رتبه ای از میانه به عنوان شاخص مرکزی استفاده می شود. برای داده های کمی به ترتیب اولویت با میانگین ، میانه و نما است. معیارهای مرکزی معمولاً در حوالی مرکز منحنی فراوانی قرار می گیرند.

- میانگین :

فرض کنید تعداد داده ها n به صورت $x_1, x_2, \dots, x_k$ با فراوانی های $f_1, f_2, \dots, f_k$ خلاصه شده باشند به طوری که $n = \sum_{i=1}^{k} f_i$ ، آن گاه اگر ماهیت داده ها کمی باشد و x_i نماینده رده i ام باشد (در جدول فراوانی) میانگین حسابی به صورت

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

تعریف می شود. میانگین انواع مختلفی دارد از جمله می توان به میانگین هندسی ، میانگین توافقی و میانگین ریشه ای رتبه دوم اشاره نمود که به صورت

پاورقی

1. mean
2. Median
3. Mode

(۱۰)

- میانگین هندسی

$$G = \sqrt[n]{x_1^{f_1} \cdots x_k^{f_k}}$$

- میانگین توافقی

$$H = \left(\frac{1}{n} \sum_{i=1}^{k} \frac{f_i}{x_i}\right)^{-1}$$

- میانگین ریشه‌ای رتبه ۲

$$M_2 = \left(\frac{1}{n} \sum_{i=1}^{k} f_i x_i^2\right)^{\frac{1}{2}}$$

تعریف می‌شوند. بین این چهار نوع میانگین نامساوی زیر برقرار است.

$$H \leq G \leq \bar{X} \leq M_2$$

برای بررسی نامساوی فوق کافیست نشان دهید $M_r = \left(\frac{1}{n} \sum_{i=1}^{k} f_i x_i^r\right)^{\frac{1}{r}}$ تابعی صعودی از r است و سپس $r = -1$ را ۰ و 2 قرار دهید.

- میانه: یکی دیگر از اندازه‌ها مهم مرکزی است. اندازه‌ای می‌باشد که حداقل نیمی از داده‌ها از آن کمتر و نیمی از آن بیشترند. نحوه محاسبه این اندازه به این صورت است که اگر داده‌ها (تعداد) $x_{(1)}, \dots, x_{(n)}$ به صورت کم باشد ابتدا داده‌ها را به ترتیب صعودی یا نزولی مرتب می‌شوند و چنانچه (تعداد) داده‌ها فرد باشد میانه عددی است که رتبه $\frac{n+1}{2}$ دارد و اگر n زوج باشد میانه به صورت

$$m = \frac{x_{(n/2)} + x_{(\frac{n}{2}+1)}}{2}$$

به عنوان مثال میانه اعداد 79 ,77 , 80 ,87 , 94 ,99 ,86 به این صورت پیدا می‌شود که ابتدا به صورت صعودی داده‌ها را مرتب می‌کنیم 77, 79, 80, 86, 87, 94, 99 و چون تعداد داده‌ها فرد است یعنی $n=7$ لذا $x_{(4)} = 86$ میانه داده‌ها است. اکنون فرض کنید عدد 80 در داده‌ها نباشد و داده‌ها پس از مرتب کردن به صورت 77, 79, 86, 87, 94, 99 باشند چون $n=6$ زوج است میانه به صورت

$$m = \frac{x_{(3)} + x_{(4)}}{2} = \frac{86+87}{2} = 86.5$$

محاسبه می‌شود.

- محاسبه میانه از روی جدول فراوانی

اگر تعداد داده ها زیاد باشد و داده ها به صورت جدول فراوانی طبقه بندی شده باشند برای محاسبه میانه باید ابتدا مقدار $\frac{n}{2}$ را محاسبه نمود و با ستون فراوانی تجمعی جدول مقایسه کرد. اولین طبقه یا رده ای که فراوانی تجمعی آن بزرگتر یا مساوی $\frac{n}{2}$ باشد را یافته و آن طبقه را طبقه میانه نامید. حال فراوانی تجمعی قبل از طبقه میانه را از $\frac{n}{2}$ کم و بر فراوانی طبقه میانه تقسیم و در طول طبقه ضرب می شود و در نهایت میانه از رابطه

$$m = L_m + \frac{\frac{n}{2} - g_{m-1}}{f_m} \times c$$

به دست می آید که در آن L_m حد پایین رده میانه ، g_{m-1} فراوانی تجمعی قبل از طبقه میانه و f_m فراوانی طبقه میانه و c طول طبقه می باشد.

نکته: برای محاسبه میانه از تمام داده ها استفاده نمی شود به همین دلیل شاخص مناسبی برای بسیاری از مطالعات آماری نیست اما چون میانه تحت تأثیر داده های پرت نیست، گاهی در برخی مطالعات آماری نسبت به میانگین ارجح تر است.

مثال ۳: یک سامانه انرژی خورشیدی برای یک کارخانه طراحی و ساخته شد که آب وارده به دیگ های بخار را پیش از به کار انداختن سامانه بخارده‌ی یا رجیستر گرم کند. انرژی تحویل شده به آب وارده برای مدت ۴۳ ساعت هر ساعت یک بار مشاهده و بر حسب واحد انرژی یعنی کیلووات ساعت اندازه گیری شد و داده ها به صورت زیر مشاهده شد. جدول فراوانی مشاهدات را به دست آورید و سپس میانگین، مُد و میانه داده ها را محاسبه کنید.

239	212	249	227	218	310	281	330	226
233	223	161	195	233	249	284	245	174
154	256	196	299	210	301	199	258	205
195	227	244	355	234	195	179	357	282
265	286	286	176	195	163	297		

(۱۲)

برای به دست آوردن جدول فراوانی ابتدا تعداد طبقه را محاسبه می کنیم از رابطه استورجس داریم

$$k = 1 + ۳٫۳ \log ۴۳ \simeq 6$$

پس داده ها را به ۶ رده تقسیم می کنیم. اکنون برد را محاسبه می کنیم

$R = \max(x_i) - \min(x_i) + h$

$= ۳۵۷ - ۱۵۴ + ۱ = ۲۰۴$

و بنابراین طول رده ها $C = \frac{R}{K} = \frac{۲۰۴}{۶} = ۳۴$ به دست می آید. برای تعیین مرز پایین رده اول

$$L_1 = \min(x_i) - \frac{h}{۲} = ۱۵۴ - ۰/۵ = ۱۵۳/۵$$

لذا جدول فراوانی به صورت زیر حاصل می شود:

حدود رده	فراوانی مطلق f_i	فراوانی نسبی r_i	فراوانی تجمعی g_i	فراوانی تجمعی نسبی S_i	نماینده x_i
۱۵۳٫۵ - ۱۸۷٫۵	۶	0.139	۶	۰٫۱۳۹	۱۷۰٫۵
۱۸۷٫۵ - ۲۲۱٫۵	۱۰	0.232	۱۶	۰٫۳۷۲	۲۰۴٫۵
۲۲۱٫۵ - ۲۵۵٫۵	۱۲	0.279	۲۸	۰٫۶۵۱	۲۳۸٫۵
۲۵۵٫۵ - ۲۸۹٫۵	۸	0.186	۳۶	۰٫۸۳۷	۲۷۲٫۵
۲۸۹٫۵ - ۳۲۳٫۵	۴	0.093	۴۰	۰٫۹۳	۳۰۶٫۵
۳۲۳٫۵ - ۳۵۷٫۵	۳	0.07	۴۳	۱	۳۴۰٫۵

برای محاسبه میانگین حسابی از رابطه $\bar{X} = \frac{1}{n} \sum_{i=1}^{k} f_i x_i$ داریم

$$\bar{X} = \frac{1}{۴۳} [۶ \times ۱۷۰٫۵ + ۱۰ \times ۲۰۴٫۵ + ۱۲ \times ۲۳۸٫۵ + ۸ \times ۲۷۲٫۵ + ۴ \times ۳۰۶٫۵ + ۳ \times ۳۴۰٫۵] = ۲۴۰٫۸$$

برای محاسبه میانه چون $\frac{n}{۲} = \frac{۴۳}{۲} = ۲۱٫۵$ پس رده سوم اولین طبقه ای است که فراوانی تجمعی آن بزرگتر از ۲۱٫۵ است و طبقه میانه می باشد و ~~...~~ $L_m = ۲۲۱٫۵$ ، $f_m = ۱۲$ ؛ g_{m-1} ۱۶ و $C = ۳۴$ و با جایگذاری در رابطه میانه:

$$m = ۲۲۱٫۵ + \left(\frac{۲۱٫۵ - ۱۶}{۱۲}\right) \times ۳۴ = ۲۳۷٫۱$$

Scanned with CamScanner

- نما : نما یا مد اندازه‌ای است که در یک مجموعه داده بیشترین فراوانی را داشته باشد. معمولاً با نماد M نشان داده می‌شود. میانگین و میانه حتماً وجود دارند و منحصر به فرد هستند اما مد ممکن است وجود نداشته باشد، یکتا باشد و یا چند مد در یک مجموعه داده وجود داشته باشد. وقتی تعداد داده‌ها کم باشد، تعیین نما به راحتی امکان پذیر است و مقداری که فراوانی آن بیشتر از بقیه مقادیر باشد به عنوان نما شناخته می‌شود. در صورتی که تعداد داده‌ها زیاد باشد و جدول فراوانی در اختیار باشد

- محاسبه مد از روی جدول فراوانی‌ها

ابتدا رده‌ای که بیشترین فراوانی را دارد مشخص می‌کنیم و سپس اختلاف فراوانی نسبی آن رده و رده قبل از آن را محاسبه با d_1 نشان می‌دهند و اختلاف فراوانی نسبی آن رده با رده بعد از آن را با d_2 نشان می‌دهند در این صورت نما از رابطه زیر

$$M = L_M + \frac{d_1}{d_1 + d_2} \cdot c$$

مثال ۴:

به دست می‌آید که در آن c طول رده و L_M مرز پایین رده دارای بیشترین فراوانی است. در مثال ۳ نما را بیابید

رده سوم بیشترین فراوانی را دارد بنابراین $L_M = 221.5$ ، $d_1 = 0.279 - 0.232 = 0.047$

و $d_2 = 0.279 - 0.186 = 0.093$ و نما عبارت است از

$$M = 221.5 + \frac{0.047}{0.047 + 0.093} \times 34 \simeq 232.9$$

- چندک‌ها: چندک‌ها که معمولاً با Q_p ، $0 < p < 1$ نشان داده می‌شوند از معیارهای تمرکز می‌باشند. چندک‌های مرتبه ۰/۲۵ ، ۰/۵ و ۰/۷۵ با Q_1 ، Q_2 و Q_3 بیان می‌شود. منظور از چندک Q_p یا چندک مرتبه p این است که تقریباً $100p$ درصد داده‌ها کوچکتر از آن می‌باشند. برای مثال $Q_{0.15}$ چندک مرتبه ۰/۱۵ ام نامیده می‌شود که ۱۵ درصد داده‌ها کوچکتر از $Q_{0.15}$ باشند. بدیهی است Q_2 یا $Q_{0.5}$ همان میانه است. به Q_1 ، Q_2 و Q_3 به ترتیب چارک اول، چارک دوم و چارک سوم نیز می‌گویند.

برای محاسبه چندک‌ها در صورتی که تعداد داده کم باشد، ابتدا داده‌ها را مرتب کنید، سپس $(n+1)\times p$ را محاسبه کنید، که در آن n تعداد داده‌ها و p مرتبه چندک می‌باشد. پس از محاسبه $(n+1)\times p$ قسمت صحیح آن را r و قسمت اعشار آن را ω بنامید در این صورت چندک مرتبه p ام از رابطه

$$Q_p = (1-\omega)\,x_{(r)} + \omega\, x_{(r+1)}$$

به دست می‌آید.

- محاسبه چندک‌ها برای جدول فراوانی:

اگر تعداد داده‌ها زیاد باشد و داده‌ها به صورت جدول فراوانی طبقه‌بندی شده باشند برای محاسبه چندک ابتدا اولین طبقه‌ای که فراوانی تجمعی آن از $n\times p$ بیشتر است را مشخص کنید و طبقه چندک بنامید. مرز پایین طبقه چندک را با L_p، فراوانی تجمعی رده قبل از طبقه چندک را با g_{p-1} و فراوانی طبقه چندک را با f_p نشان دهید، چندک p ام از رابطه

$$Q_p = L_p + \left(\frac{n\times p - g_{p-1}}{f_p}\right)\times c$$

به دست می‌آید، که در آن c طول رده چندک است.

مثال ۵: برای داده‌های مثال ۳ چندک مرتبه ۲۵٪ و ۵۰٪ و ۷۵٪ را محاسبه کنید.

ابتدا $n\times p = 43\times 0.25 = 10.75$ را محاسبه کنید بنابراین رده دوم رده چندک ۲۵٪ ام یا همان چارک اول است. اکنون داریم:

$$Q_{0.25} = Q_1 = 187.5 + \left(\frac{43\times 0.25 - 6}{10}\right)\times 34 = 203.65$$

به همین ترتیب $43\times 0.5 = 21.5$ و رده سوم رده چندک ۰/۵ است

$$Q_{0.5} = Q_2 = 221.5 + \left(\frac{43\times 0.5 - 16}{12}\right)\times 34 = 237.1$$

و برای محاسبه Q_3، $43\times 0.75 = 32.25$ و رده چهارم، رده چندک ۰/۷۵ است

$$Q_{0.75} = Q_3 = 255.5 + \left(\frac{43\times 0.75 - 28}{8}\right)\times 34 = 273.6$$

۲.۳.۲ شاخص های پراکندگی

اندازه های مرکزی مانند میانگین، میانه و... توصیف کننده وضعیت کامل توزیع داده ها نیستند. به عبارت دیگر دو مجموعه داده که دارای میانگین های یکسان هستند، ممکن است پراکندگی متفاوت داشته باشند و زمانی می توان نحوه توزیع داده ها را دقیقاً توصیف نمود که علاوه بر شناخت معیار مناسبی برای مرکزیت آن، معیاری هم برای پراکندگی آن ها تعیین نمود. معروف ترین معیارهای پراکندگی ~~عبارتند از~~ دامنه تغییرات یا برد، برد چارک ها، میانگین انحراف ها، واریانس و انحراف معیار و ضریب تغییرات می باشند.

- دامنه تغییرات: عبارت است از تفاضل بزرگ ترین و کوچک ترین مقدار در بین داده ها و معمولاً با نماد R نشان می دهند. بنابراین اگر $x_{(1)}$ مینیمم داده ها و $x_{(n)}$ ماکسیمم داده ها در نمونه ای به حجم n باشد، آن گاه

$$R = x_{(n)} - x_{(1)}$$

است. در مثال ۳ دامنه تغییرات داده ها به صورت $R = ۳۵۷ - ۱۵۴ = ۲۰۳$ به دست می آید. چون این معیار در هر نمونه فقط از دو مقدار ماکسیمم و مینیمم استفاده می کند و دیگر داده ها نقشی در محاسبه آن ندارند، بنابراین ~~تصویر~~ توصیفی درست از تغییرات نمونه ارائه نمی کند و همچنین تحت تأثیر داده های پرت است.

- برد چارک ها: این معیار تحت تأثیر داده های پرت نیست و برای محاسبه آن از چارک سوم و چارک اول استفاده می شود و معمولاً با نماد IQR نشان داده می شود

$$IQR = Q_3 - Q_1$$

برای داده های مثال ۳، $Q_1 = 203.65$ و $Q_3 = ۲۷۳.۶$ در مثال ۵ محاسبه شده است و بنابراین IQR برای این داده ها عبارت است از:

$$IQR = ۲۷۳.۶ - ۲۰۳.۶۵ = ۶۹.۹۵$$

این معیار نیز از تمام داده ها استفاده نمی کند و نمی تواند تصویر درستی از تغییرات نمونه ارائه نماید.

- میانگین انحراف ها

یکی دیگر از معیارهای پراکندگی معیار میانگین انحراف ها می باشد که به صورت

$$d = \frac{1}{n}\sum_{i=1}^{k} f_i \,|x_i - \bar{x}|$$

محاسبه می شود. ایراد اصلی ~~این~~ معیار میانگین انحراف ها این است که اثر خطای خیلی بزرگ را در مقابل تعداد بسیار زیادی از خطاهای کوچک منعکس نمی کند. برای داده های مثال ۳ :

$$d = \frac{1}{۴۳}\left[۶\,|۱۷۰٫۱۵ - ۲۴۰٫۱۸| + ۱۰\,|۲۰۴٫۵ - ۲۴۰٫۱۸| + \dots \right] =$$

- مثال ۶ : میانگین قدر مطلق انحرافات از میانگین را برای مشاهدات

۱۰ ۱۵ ۸ ۲۰ ۲۵

محاسبه کنید : $\bar{x} = ۱۵٫۶$

$$d = \frac{1}{۵}\left[|۱۰ - ۱۵٫۶| + |۱۵ - ۱۵٫۶| + |۸ - ۱۵٫۶| + |۲۰ - ۱۵٫۶| + |۲۵ - ۱۵٫۶| \right] = ۵٫۲۵$$

- واریانس و انحراف معیار

یکی از مفیدترین شاخص های اندازه گیری پراکندگی است به طوری که از تمام داده ها برای محاسبه آن استفاده می شود و به جای استفاده از تابع قدر مطلق انحراف از توان دو انحراف ها در محاسبه آن استفاده می شود واریانس یک ~~نمونه~~ نمونه ای به حجم n به صورت

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

محاسبه می شود در صورتی که داده ها در جدول فراوانی خلاصه شوند می توان از

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

استفاده کرد. در صورتی که به جای نمونه، داده های مربوط به جامعه در دسترس باشد واریانس از رابطه

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

یا

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{K} f_i (X_i - \mu)^2$$

استفاده می‌گردد که در آن μ میانگین جامعه و $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ است.

اگر واریانس به عنوان یک معیار مفید در اغلب مطالعات مورد استفاده می‌گیرد، اما واحد اندازه‌گیری این معیار به دلیل استفاده از توان دوم در محاسبه آن با واحد اندازه‌گیری داده‌ها یکی نیست و نسبت به داده‌های بزرگ و کوچک و به عبارت دیگر داده‌های پرت حساس است. برای همسان کردن واحد اندازه‌گیری معیار واریانس با واحد اندازه‌گیری مشاهدات معمولاً از جذر واریانس استفاده می‌شود که به این معیار جدید، انحراف معیار می‌گویند. لذا انحراف معیار در نمونه عبارت است از

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

و هرچه S کوچک‌تر باشد، داده‌ها همگن‌تر می‌باشند. برای حل مشکل حساسیت واریانس به داده‌های پرت می‌توان از معیار استوارتر مثل میانه قدر مطلق انحراف‌ها از میانه نیز محاسبه گردد که به (MAD) در کنار واریانس

صورت

$$MAD = median\{|x_i - m|\}$$

می‌باشد، که در آن m میانه داده‌ها است. یعنی ابتدا داده‌ها از میانه کم و سپس قدر مطلق آن‌ها محاسبه و در نهایت میانه این قدر مطلق ~~میانه~~ محاسبه می‌شود.

- خواص واریانس
- اگر داده‌ها از مقدار ثابتی کم یا ~~اضافه~~ مقدار ثابتی به آن‌ها اضافه گردد، واریانس و انحراف معیار تغییری نمی‌کنند.
- اگر مشاهدات در عدد ثابتی مانند a ضرب یا تقسیم شود، واریانس در مجذور عدد و انحراف معیار در آن عدد ضرب یا تقسیم می‌شود.
- واریانس و انحراف معیار عدد ثابت برابر صفر است.

- مثال ۷ : تعداد کتاب‌های منتشر شده در سال ۱۳۹۴ در ۲۰ انتشارات در شهر تهران به شرح زیر

Median Absolute Deviation (MAD)

(۱۸)

MAD و انحراف معیار

واریانس، این نمونه استخراج شده از انتشارات شهر تهران را محاسبه کنید

x_i: ۲ ۲ ۵ ۳ ۴ ۵ ۷ ۲ ۸ ۹ ۱۰ ۱۱ ۴ ۵ ۶ ۷
۲ ۳ ۵ ۴

حل: $\bar{x} = 4.95$ ، واریانس $S^2 = \frac{1}{20-1}\sum_{i=1}^{20}(x_i-\bar{x})^2$ بنابراین:

$$S^2 = \frac{1}{19}\left[(2-4.95)^2 + (2-4.95)^2 + \dots + (4-4.95)^2\right] = 7.44$$

$$S = 2.73$$

برای محاسبه MAD ابتدا $|x_i - m|$ باید محاسبه شود، برای محاسبه میانه داده‌ها باید مرتب شوند:

$x_{(i)}$: ۲ ۲ ۲ ۲ ۳ ۳ ۴ ۴ ۴ ۵ ۵ ۵ ۵ ۶ ۷ ۷
۸ ۹ ۱۰ ۱۱

$$m = \frac{x_{(10)} + x_{(11)}}{2} = \frac{5+5}{2} = 5$$

و بنابراین

$|x_i - m|$ ۳ ۳ ۰ ۲ ۱ ۰ ۲ ۳ ۳
۴ ۵ ۶ ۱ ۰ ۱ ۲ ۳ ۲
۰ ۱

و مجدد میانه این اعداد محاسبه می‌شود

۰ ۰ ۰ ۰ ۱ ۱ ۱ ۱ ۲ ۲ ۲ ۲ ۳ ۳
۳ ۳ ۳ ۴ ۵ ۶

$$\text{median}\ |x_i - m| = \frac{x_{(10)} - x_{(11)}}{2} = \frac{2+2}{2} = 2$$

- ضریب تغییرات (CV):

گاهی ~~علاقمند~~ مقایسه پراکندگی دو گروه ~~یا~~ دو نمونه ~~که متفاوت باشد~~ یا دو صفت با واحد اندازه‌گیری متفاوت مورد نظر می‌باشد. معیار ضریب تغییرات به واحد اندازه‌گیری بستگی ندارد و به صورت

$$CV = \frac{S}{|\bar{x}|}$$

برای نمونه و $CV = \frac{\sigma}{|\mu|}$ برای جامعه محاسبه می‌گردد.

Coefficient of variation

بنابراین اگر هدف مقایسه دو صفت با واحدهای اندازه گیری متفاوت باشد باید از ضریب تغییرات استفاده کرد.

اگر هر یک از مشاهدات را در عدد ثابتی مانند a ضرب یا تقسیم شود ضریب تغییرات تغییر نمی کند و اگر به مشاهدات عدد ثابتی کم یا اضافه شود در این صورت ضریب تغییرات اضافه یا کم می شود.

مثال ۸ : فشارخون سیستولیک و وزن یک گروه از افراد اندازه گیری شد و مقادیر زیر به دست آمد

صفت	میانگین	انحراف معیار
وزن	۷۰	۷
فشارخون	۱۰۰	۱۰

تغییرات وزن بیشتر است را با تغییرات فشارخون مقایسه کنید.

$$\text{ضریب تغییرات وزن} = \frac{7}{70} = \frac{1}{10} \times 100 = \%10$$

$$\text{ضریب تغییرات فشارخون} = \frac{10}{100} = \frac{1}{10} \times 100 = \%10$$

پس تغییرات یکسان است.

۳.۲.۲ چولگی و برجستگی

برای سنجش تقارن توزیع مشاهدات نمونه از معیار چولگی و برای سنجش بلندای توزیع مشاهدات نمونه در مقایسه با توزیع نرمال از معیار برجستگی یا کشیدگی همان استفاده می شود. ضریب چولگی را می توان از سه طریق و به صورت

① ضریب چولگی اول پیرسن
$$b_1 = \frac{\bar{x} - M}{S}$$

۲ - ضریب چولگی دوم پیرسن
$$b_2 = \frac{3(\bar{x} - m)}{S}$$

۳ - ضریب گشتاوری
$$g_1 = \frac{m_3}{S^3}$$

محاسبه نمود که در آن M مد مشاهدات ، m میانه مشاهدات ، S انحراف معیار و m_3 گشتاور مرکزی مرتبه ۳ می باشد. گشتاور مرکزی مرتبه ۳ ام عبارت است از

$$m_3 = \frac{1}{n} \sum_{i=1}^{k} f_i (x_i - \bar{x})^3$$

Scanned with CamScanner

و گشتاور مرتبه r ام به صورت

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

محاسبه می گردد.

درصورتی که ضریب چولگی مثبت حاصل شود بیانگر آن است که دم راست توزیع مشاهدات نمونه از میانگین بیشتر دور می شود تا دم چپ آن و به همین ترتیب ضریب چولگی اگر منفی شود، دم چپ توزیع مشاهدات نمونه نسبت به دم راست از میانگین بیشتر دور می شود. ضریب چولگی صفر تقارن توزیع مشاهدات را نشان می دهد.

برای محاسبه کشیدگی از رابطه

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{m_4}{s^4} - 3$$

استفاده می شود. کشیدگی مثبت بیانگر کشیده تر بودن توزیع مشاهدات نمونه از توزیع نرمال استاندارد و کشیدگی منفی نشان دهنده پخت تر بودن توزیع مشاهدات نمونه از توزیع نرمال استاندارد است. اگر کشیدگی صفر شود نشان دهنده هم ارتفاع بودن توزیع مشاهدات نمونه با نرمال استاندارد است.

تذکر: برای به دست آوردن معیارهای نااریب از چولگی و کشیدگی می توان از روابط زیر استفاده نمود:

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2}\, g_1 \qquad G_2 = \frac{n-1}{(n-2)(n-3)}\left[(n+1)g_2 + 6\right]$$

مثال ۹: برای داده های مثال ۳، $\bar{x} = 240.18$ ، $m = 237.1$ و $M = 232.9$ و $S \cong 51$ و $m_3 = 57051.25$ و $m_4 = 16652384$ و بنابراین

$$g_1 = \frac{m_3}{s^3} = 0.42 \quad , \quad g_2 = \frac{m_4}{s^4} - 3 = -0.54$$

۲.۳ چند نمودار مفید دیگر

۲.۳.۱ نمودار ساقه و برگ

برای رسم این نمودار از ارقام مشاهدات استفاده می‌شود به طوری که رقم‌ها برای نمایش دامنه تغییرات، شکل توزیع و تراکم داده‌ها به کار برده می‌شود. ابتدا هر مشاهده به رقم‌های اصلی و رقم‌های فرعی تجزیه می‌شود. همه رقم‌های اصلی در سمت چپ یک خط قائم به ترتیب افزایشی مرتب می‌شوند و سپس رقم‌های فرعی مربوط به آن‌ها در سمت راست خط در جای مناسب خود پشت سر هم در یک سطر ثبت می‌شوند. رقم‌های اصلی ساقه و رقم‌های فرعی برگ نامیده می‌شود.

مثال ۱۰: برای داده‌های زیر نمودار ساقه و برگ رسم کنید

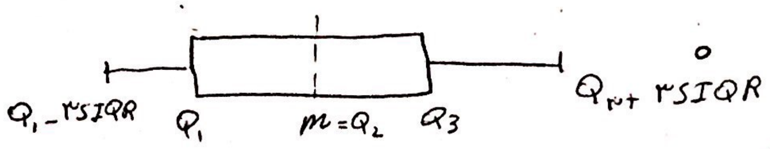
47 ,45 ,45 ,41 ,41 ,35 ,34 ,24 ,26 ,22 ,60 ,54 ,58 و 49

داده‌ها را به دو قسمت تقسیم، رقم دهگان را ساقه و یکان را برگ قرار دهید

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

۲.۳.۲ نمودار جعبه‌ای

با استفاده از نمودار جعبه‌ای می‌توان وصف کل داده‌ها را در یک مستطیل در جهت نشان داد. برای رسم این نمودار ابتدا باید چارک اول، دوم و سوم، و سپس نیم دامنه بین چارکی یعنی (SIQR)[1] $SIQR = \frac{Q_3 - Q_1}{2}$ محاسبه گردد و مستطیلی به طول $2SIQR$ رسم می‌شود، در نهایت یک نمونه نمودار جعبه‌ای به صورت زیر می‌باشد



تذکر: می‌توان میانه را با علامت * درون جعبه مشخص کنید و اگر داده‌ای در محدوده $Q_1 - 3SIQR$ و $Q_3 + 3SIQR$ قرار نگرفت به عنوان داده پرت و با علامت o مشخص می‌گردد.

1 semi-interquartile range

۴.۲. استاندارد کردن مشاهدات

گاهی در عمل برای از بین بردن واحد اندازه گیری مشاهدات و یا مقایسه یک صفت در دو گروه از روش استاندارد کردن مشاهدات استفاده می شود. فرض کنید $x_1 \dots x_k$ به ترتیب دارای فراوانی $f_1, \dots, f_k$ باشند، برای استاندارد کردن x_i ها با مقادیر از میانگین مشاهدات کم و بر انحراف معیار تقسیم گردد. معمولاً استاندارد شده مشاهدات را با z_i، $z_i = \frac{x_i - \bar{x}}{s}$ نشان داده می شود.

مثال ۱۱. فرض کنید علی در درس احتمال در کلاس ۱ نمره ۱۶ و محمد در کلاس ۲، نمره ۱۶ کسب کرده اند. آیا علی و محمد در این درس یکسان عمل کرده اند به طوری که می دانیم میانگین نمره در کلاس ۱ ۱۴ و انحراف معیار ۲ و در کلاس ۲ میانگین ۱۲٫۵ و انحراف معیار ۱٫۵ است.

ابتدا نمره علی و محمد باید استاندارد شوند

$$z_1 = \frac{16 - 14}{2} = 1 \quad و \quad z_2 = \frac{16 - 12.5}{1.5} = 2.3$$

پس محمد از علی در درس احتمال بهتر عمل نموده است.

Subject
Date

داده‌های زیر میزان تصادف منجر به مرگ در ۲۰ منطقه از شهری را نشان می‌دهند. جدول فراوانی داده‌ها را بدست آورید نمودار میله‌ای آن را رسم کنید.

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

x_i	f_i	r_i	g_i	s_i
۲	۱	۰/۰۵	۱	۰/۰۵
۳	۳	۰/۱۵	۴	۰/۲
۴	۳	۰/۱۵	۷	۰/۳۵
۵	۳	۰/۱۵	۱۰	۰/۵۰
۶	۴	۰/۲۰	۱۴	۰/۷
۷	۲	۰/۱۰	۱۶	۰/۸
۸	۱	۰/۰۵	۱۷	۰/۸۵
۹	۳	۰/۱۵	۲۰	۱

$n = 20 \quad (n+1)p = 21 \times 0/25 = 5/25 \quad r = 5$

$w = 0/25$

$Q_1 = (1-w)X_{(r)} + w\,X_{(r+1)}$

$Q_1 = 0/75\,X_{(5)} + 0/25\,X_{(6)} = 0/75 \times 4 + 0/25 \times 4 = 4$

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

میانه $m = 5/5 = Q_2$

$21 \times 0/75 = 15/75 \rightarrow r = 15 \quad w = 0/75$

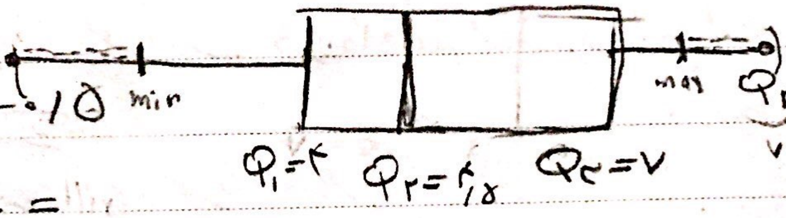
$Q_3 = (1 - 0/75)\,X_{(15)} + 0/75\,x_{(16)}$

$= 0/25 \times 7 + 0/75 \times 7 = 7 \qquad SIQR = \frac{Q_3 - Q_1}{2}$

$= \frac{7-4}{2} = \frac{3}{2}$

Subject

Date



$Q_1 - ۳SIQR = -۰٫۵$

$\underbrace{Q_3}_{۷} + \underbrace{۳SIQR}_{۴\times ۴٫۵} = ۱۱٫۵$

$Q_1 - ۳ \times \frac{۳}{۲} =$

$۴ - ۴٫۵ = -۰٫۵$

داده‌های زیر از یک پژوهش علمی برای مقایسه میزان مصرف بنزین چهار نوع خودرو

به دست آمده‌اند نمودار ساقه و برگ خودرو PG را رسم کنید.

واریانس خودرو PK را به دست آورید. نمودار جعبه‌ای چهار خودرو را مقایسه کنید

آزمایش	MX	PG	PK	PR
۱	۱۰٫۲۳	۱۰٫۲۸	۸٫۰۸	۱۷٫۲۴
۲	۱۰٫۶۲	۱۰٫۰۲	۸٫۱۴	۱۷٫۲۴
۳	۱۰٫۷۹	۱۰٫۳۶	۸٫۲۶	۱۷٫۴۲
۴	۱۰٫۶۲	۱۰٫۰۲	۸٫۲۶	۱۷٫۴۷
۵	۱۰٫۵۴	۱۰٫۴۲	۸٫۳۰	۱۷٫۵۹
۶	۱۰٫۶۶	۱۰٫۶۲	۸٫۴۷	۱۷٫۹۷
۷	۱۰٫۷۵	۱۰٫۷۵	۸٫۴۷	۱۸٫۰۶
۸	۱۰٫۷۵	۱۰٫۷۵	۸٫۵۱	۱۸٫۶۲

PG

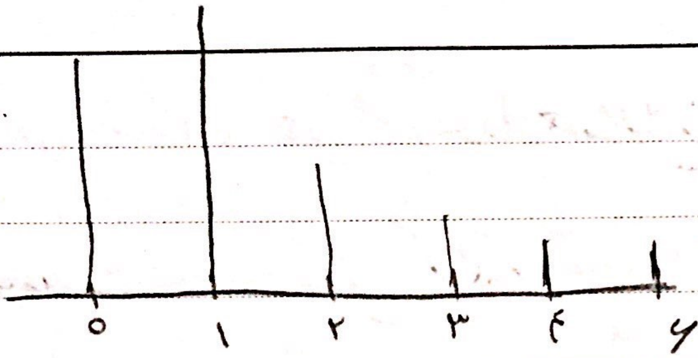
۱۰٫۰	۲	۲
۱۰٫۲	۸	
۱۰٫۳	۶	۶
۱۰٫۶	۲	
۱۰٫۷	۵	۵

PAPCO

Scanned with CamScanner

Subject

Date



$(۳۰+۱) \times ۰/۲۵ = ۳۱ \times ۰/۲۵ = ۷/۷۵$ $\qquad r=۷$, $w=۰/۷۵$

$$Q_1 = (۱-۰/۷۵)\underbrace{X_{(۷)}}_{۰} + ۰/۲۵\underbrace{X_{(۸)}}_{۰} = ۰$$

$$\underbrace{۰\ ۰ \cdots ۰}_{۹}\quad \underbrace{۱ \cdots ۱}_{۱۰}\quad \underbrace{۲ \cdots ۲}_{۵}\quad \underbrace{۳\ ۳\ ۳}_{۳}\quad \underbrace{۴\ ۴}_{۲}\quad \underbrace{۶}_{۱}$$

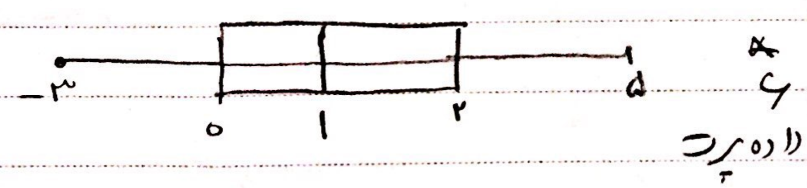
$$M = \frac{x_{۱۵}+x_{۱۶}}{۲} = \frac{۱+۱}{۲} = ۱$$

$(n+1)\times ۰/۷۵ = ۲۳/۲۵ \rightarrow r=۲۳ \quad w=۰/۲۵$

$$Q_3 = (۱-۰/۲۵)\underbrace{x_{(۲۳)}}_{۲} + ۰/۲۵\underbrace{x_{(۲۴)}}_{۲} = ۰/۷۵\times۲ + ۰/۲۵\times۲ = ۲$$

$$SIQR = \frac{Q_3 - Q_1}{۲} = ۱$$

$$Q_1 - ۳SIQR = -۳ \qquad \overbrace{Q_3}^{۲} + ۳\overbrace{SIQR}^{۱} = ۵$$



- معدل ۲۰ دانشجوی دانشگاه به شرح زیر است

۱۲٫۴	۱۲٫۹	۱۱٫۷	۱۱٫۸	۱۲٫۴	۱۲٫۲	۱۲٫۱	۱۲٫۲	۱۱٫۶	۱۲٫۱
۱۲٫۸	۱۱٫۹	۱۲٫۹	۱۲٫۶	۱۲٫۵	۱۲٫۴	۱۲	۱۲٫۴	۱۲٫۸	۱۱٫۹

Subject
Date

نوع متغیر، جدول فراوانی را تشکیل دهید. هیستوگرام و نمودار جعبه‌ای را رسم کنید. ~~[illegible]~~ ...

میانگین ~~[illegible]~~ کشیدگی و چولگی را هم رسم کنید

$\min = 11.6$

$\max = 12.9$ $\quad h = 0.1 \quad 12.9 - 11.6 + 0.1 = 1.4$

← چون یک رقم اعشار

دو رقم، یک رقم، هیچ رقم

$h = 1 \quad 0.1 \quad 0.01 \ldots$

5.29

تعداد رده $1 + 3.3 \log_{10} 20 = 5$

هر دو درست است ۶ یا ۵

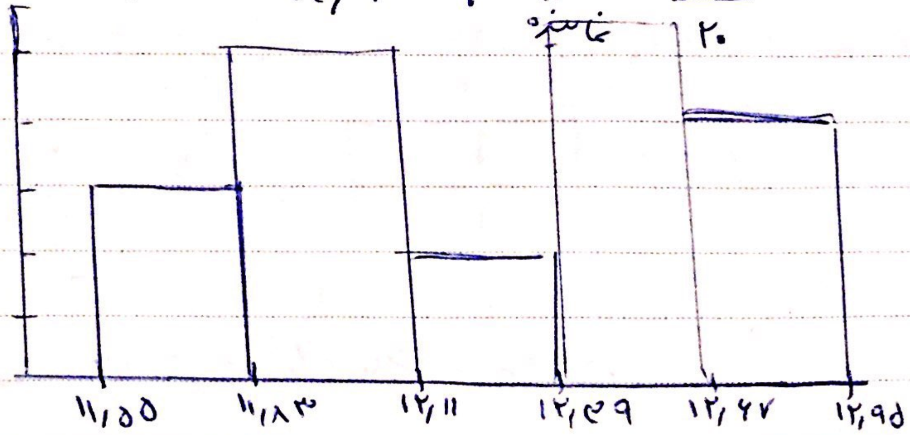
در ۵ رده تقسیم کنید ←

$C_{\text{طول}} = \frac{1.4}{5} = 0.28$

مرز پایین رده اول: $11.6 - \frac{0.1}{2} = 11.55$

طبقات	x_i	f_i	r_i	g_i	s_i
$11.55 - 11.83$	$\frac{11.55+11.83}{2} = 11.69$	3	$3/20$	3	$3/20$
$11.83 - 12.11$	11.97	5	$5/20$	8	$8/20$
$12.11 - 12.39$	12.25	2	$2/20$	10	$10/20$
$12.39 - 12.67$	12.53	6	$6/20$	16	$16/20$
$12.67 - 12.95$	12.81	4	$4/20$	20	1
		20			

نمایش



$Q_1 = L_p + \left(\frac{np - g_{p-1}}{f_p}\right) C = 11.83 + \frac{5 - 3}{5} \times 0.28 = 11.942$

طول طبقه

$np = 20 \times 0.25 = 5$ → طبقه دوم

اولین طبقه‌ای که حداقل فراوانی تجمعی ۵ باشد.

Subject
Date

$$Q_{۳} = ۱۲٫۳۹ + \left(\frac{۱۵-۱۰}{۶}\right)۰٫۱۲۸ = ۱۲٫۶۲$$

$$np = ۲۰ \times ۰٫۱۷۵ = ۱۵ \rightarrow$$ طبقه چهارم

$$m = Q_{۲} = ۱۲٫۱۱ + \left(\frac{۱۰-۸}{۲}\right) \times ۰٫۱۲۸ = ۱۲٫۳۹$$

$$۲۰ \times ۰٫۵ = ۱۰$$

$$M^{\text{مد}} = L_M + \frac{d_1}{d_1+d_2} \times C = ۱۲٫۲۹ + \frac{۴/۲۰}{۴/۲۰ + ۲/۲۰} \times ۰٫۱۲۸ = ۱۲٫۵۸$$

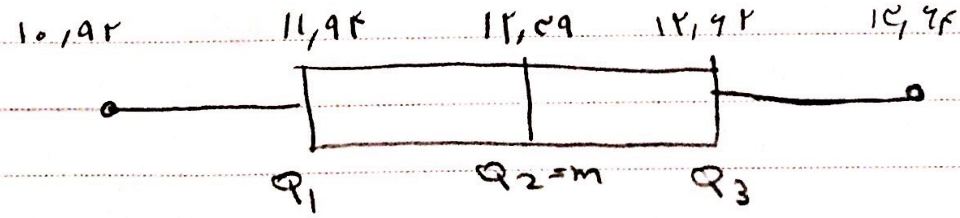
بیشترین فراوانی طبقه چهارم ← طبقه مد

$$d_1 = \underbrace{۶/۲۰}_{\text{فراوانی بیشترین رده مد}} - \underbrace{۲/۲۰}_{\text{رده قبل}} = ۴/۲۰ \qquad d_2 = \frac{۶}{۲۰} - \frac{۴}{۲۰} = \frac{۲}{۲۰}$$

نمودار جعبه‌ای

$$SIQR = \frac{Q_۳ - Q_۱}{۲} = \frac{۱۲٫۶۲ - ۱۱٫۹۴}{۲} = ۰٫۳۴$$

$$Q_۳ + ۳SIQR = ۱۳٫۶۴$$
$$Q_۱ - ۳SIQR = ۱۰٫۹۲$$



چولگی اول پیرسون $$\frac{\bar{X} - M}{S} = ? \quad \frac{۱۲٫۲۹ - ۱۲٫۵۸}{\sqrt{۰٫۱۵}} = -۰٫۷۴$$

$$\bar{X} = \frac{۳ \times ۱۱٫۶۹ + ۵ \times ۱۱٫۹۷ + ۲ \times ۱۲٫۲۵ + ۶ \times ۱۲٫۵۳ + ۴ \times ۱۲٫۸۱}{۲}$$

$$= ۱۲٫۲۹$$

$$S^۲ = \frac{1}{۱۹}\left[(۱۱٫۶۹ - ۱۲٫۲۹)^۲ \times ۳ + \cdots\right] = \cdots$$

$$\cong ۰٫۱۵$$

PAPCO

Subject
Date

کشیدگی $\frac{m'_4}{S^4} - 3 = \frac{0/04}{(0/15)^2} - 3 = -1/4$

$m'_4 = \frac{1}{n} \sum f_i (x_i - \bar{x})^4 = 0/04$ پخ‌تر از توزیع نرمال است ندارد

ضریب چولگی دوم پیرسون $= \frac{3(\bar{x} - m)}{S} = \frac{3(12/29 - 12/39)}{\sqrt{0/15}} = -0/7$

$m'_3 = -0/009$ ، چولگی گشتاوری $= \frac{m'_3}{S^3} = \frac{-0/009}{0/059} = -0/15$

Scanned with CamScanner

Date

$$PK \begin{cases} S^2 = \frac{1}{n-1} \sum_{i=1}^{7} (X_i - \bar{X})^2 = \frac{1}{7} \sum (X_i - 8.31)^2 = 0.0289 \\ \bar{X}_{PK} = 8.31 \end{cases}$$

($n-1$: ۷)

برای تمرین برای همه خودروها بدست آورید

$(n+1)p = 9 \times 0.25 = 2.25$

چون داده‌ها کم است از فرمول

$$PK \begin{cases} Q_1 = 0.75 \times \underset{8.12}{X_{(2)}} + 0.25 \underset{8.26}{X_{(3)}} = 8.16 \\ Q_3 = 8.47 \\ Q_2 = 8.28 \end{cases}$$

$(n+1)p \rightarrow$ صحیح r ، اعشار w

$Q_p = (1-w) X_{(r)} + w X_{(r+1)}$

$$MX \begin{cases} Q_1 = 10.56 \\ Q_2 = 10.64 \\ Q_3 = 10.75 \end{cases}$$

به همین ترتیب بقیه محاسبه نمودار رسم و مقایسه شود

فرض کنید دانشجویان یک کلاس در دو امتحان ۱ و ۲ شرکت کرده‌اند و نتایج به شرح جدول زیر بدست آمده است

	میانگین	~~انحراف~~ واریانس	~~نمره~~	
امتحان ۱	۵۵	۲۵	از ۱۰۰	
امتحان ۲	۶۶۰	۳۶	از ۱۰۰۰	

- با استفاده از ضریب تغییرات این نتایج را مقایسه کنید

اگر دانشجوی A در آزمون اول نمره ۷۵ و در آزمون دوم نمره ۸۳۰ را کسب کرده باشد دانشجو در کدام آزمون عملکرد بهتری داشته است

$CV_1 = \frac{S_1}{\bar{X}_1} = \frac{5}{55} = 0.091 \quad CV_2 = \frac{S_2}{\bar{X}_2} = \frac{6}{660} = 0.0091$

PAPCO

چون $CV_2 < CV_1$ در امتحان دوم تغییرات کمتر و نمره‌ی بهتر است

Scanned with CamScanner

$$Z_1 = \frac{75 - 55}{\sqrt{25}} = 2$$

$$Z_2 = \frac{830 - 660}{\sqrt{36}} = 28{,}3$$

$Z_2 > Z_1$

نمره آزمون دوم بهتر از نمره آزمون اول است.

مثال مقدار آماره ازیر برای ۵ نمونه تصادفی بدست آمده است واریانس و میانگین مشاهدات را بدست آورید

$$\sum_{i=1}^{5} y_i = 26 \qquad \sum_{i=1}^{5} y_i^2 = 158 \longrightarrow \bar{y} = 26/5$$

$$S^2 = \frac{1}{n-1}\left[\sum y_i^2 - n\bar{y}^2\right] = \frac{1}{4}\left[158 - \left(\frac{26}{5}\right)^2 \times 5\right] = 5{,}7$$

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\left[\sum y_i^2 - n\bar{y}^2\right]}$$

مثال. تعداد دفعات توقف کامپیوتر در مرکز محاسبات دانشگاه در یک واحد به مدت یک ماه به قرار زیر است. نوع مقیاس و متغیر را تعیین کنید. جدول فراوانی را به دست آورید. نمودار میله‌ای را رسم کنید. میانه، نما، چارک اول و سوم را تعیین کنید. نمودار جعبه‌ای؟

۱ ۳ ۱ ۱ ۰ ۱ ۰ ۱ ۱ ۰ ۲ ۲ ۰ ۰ ۰ ۱ ۲ ۱ ۲ ۰
۰ ۱ ۶ ۴ ۳ ۳ ۱ ۲ ۴ ۰

رتبه‌ای گسسته

x_i	۰	۱	۲	۳	۴	۵	۶
f_i	۹	۱۰	۵	۳	۲	۰	۱
r_i	۹/۳۰	۱۰/۳۰	۵/۳۰	۳/۳۰	۲/۳۰	۰	۱/۳۰
g_i	۹	۱۹	۲۴	۲۷	۲۹	۳۰	
S_i	۹/۳۰	۱۹/۳۰	۲۴/۳۰	۲۷/۳۰	۲۹/۳۰	۳۰/۳۰	