فصل اول

آمار توصیفی

۱ـ۱ـ تعاریف و کلیات

۱ـ۱ـ۱ـ تعریف علم آمار

منظور از آمار روش علمی جمع‌آوری، تنظیم، تلخیص و تجزیه و تحلیل ارقام و اطلاعات برای رسیدن به نتایج روشن و اتخاذ تصمیمات منطقی می‌باشد.

۱ـ۱ـ۲ـ آمار توصیفی

قسمتی از روشهای آماری که فقط به توصیف و تجزیه و تحلیل گروه معینی بدون تعمیم نتایج حاصله به گروه بزرگتر از آن محدود می‌گردد آمار توصیفی نامیده می‌شود.

۱ـ۱ـ۳ـ آمار استنتاجی

قسمتی از روشهای آماری که می‌تواند نتایج حاصله از تجزیه و تحلیل نمونه را به جامعه تعمیم دهد آمار استنتاجی نامیده می‌شود.

۱ـ۱ـ۴ـ جامعهٔ آماری

با اینکه جامعهٔ آماری جزء مفاهیم تعریف نشده می‌باشد معذالک می‌توان آن را به این صورت بیان کرد که کلیه افراد و اشیائی که دارای لااقل یک صفت مشترک باشند تشکیل یک جامعهٔ آماری را می‌دهند. مانند دانشجویان دانشکدهٔ حسابداری و مدیریت دانشگاه تهران.

۱-۱-۴-۱- جامعه آماری محدود و نامحدود

اگر تعداد افراد جامعه آماری محدود باشد به آن جامعه آماری محدود و اگر تعداد افراد جامعه آماری نامحدود باشد به آن جامعه آماری نامحدود می‌گویند.

به عنوان مثال جامعه ساکنین شهر تهران یک جامعه محدود و جامعه ستارگان جهان یک جامعه نامحدود می‌باشد.

۱-۱-۵- صفت متغیر

خاصیتی است که افراد یک جامعه را از یکدیگر جدا و مشخص می‌سازد. مثلاً در جامعه دانشجویان دانشکده حسابداری و مدیریت می‌توان قد، سن و رشته تحصیلی را صفت متغیر در نظر گرفت.

تبصره: به صفت مشترک افراد یک جامعه صفت مشخصه می‌گویند.

۱-۱-۵-۱- انواع صفات متغیر

الف - صفات متغیر کمّی که با عدد بیان می‌شوند مانند: سن، قد و وزن.

ب - صفات متغیر کیفی که با عدد بیان نمی‌شوند. مانند: رشته تحصیلی، رنگ پوست و مرغوبیت جنس

۱-۱-۵-۲- انواع صفات متغیر کمی

الف - صفات کمی پیوسته - که در فاصله a و b تمام اعداد حقیقی را اختیار می‌نمایند. مانند: قد و وزن

ب - صفات کمی گسسته - که فقط بعضی از مقادیر جدا از هم را قبول می‌کنند. مانند: تعداد فرزندان خانواده‌ها

۱-۱-۶- انواع مقیاس‌ها

انواع مقیاس‌های اندازه‌گیری صفت متغیر x بشرح زیر است.

۱- مقیاس اسمی - که اعداد و علائم فقط برای طبقه‌بندی اشیاء بکار می‌رود مانند بررسی صفتی در

مورد دانشگاههای مختلف. در این مقیاس اندازه‌گیری در ضعیف‌ترین شکل خود می‌باشد. این مقیاس در مورد صفات متغیر کیفی به کار می‌رود.

۲ـ مقیاس رتبه‌ای یا ترتیبی ـ که در صفت متغیر رابطه ترتیبی وجود داشته باشد مثلاً بررسی خصوصیتی در مورد افراد کم درآمد، با درآمد متوسط و پردرآمد. این مقیاس هم در مورد صفات متغیر کیفی به کار می‌رود.

۳ـ مقیاس فاصله‌ای ـ که مقیاسی قوی‌تر از ترتیبی است و می‌توان با عدد فواصل را مشخص نمود و صفر آن قراردادی است مانند مقیاس فارنهایت و سانتیگراد برای اندازه‌گیری درجه حرارت. این مقیاس در مورد صفات متغیر کمی به کار می‌رود.

۴ـ مقیاس نسبی یا نسبتی ـ که در قویترین شکل خود قرار دارد و صفر آن واقعی است مانند مقیاسهائی نظیر متر و اینچ برای اندازه‌گیری طول. این مقیاس هم در مورد صفات متغیر کمّی به کار می‌رود.

۱ـ۱ـ۷ـ علامت مجموع Σ (زیگما یا سیگما)

علامت (Σ) در آمار کاربرد زیادی دارد و برای نشان دادن مجموع اعداد گسسته می‌باشد.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \dots + x_n$$

مثال:

$$\sum_{x=1}^{5} \frac{1}{x} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}$$

۱ـ۱ـ۷ـ۱ـ خواص زیگما

الف ـ

$$\sum_{i=1}^{n} (x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

اثبات:

$$\sum_{i=1}^{n} (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \dots + (x_n + y_n) =$$

$$(x_1 + x_2 + \dots + x_n) + (y_1 + y_2 + \dots + y_n) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

ب ـ

$$\sum_{i=1}^{n} K x_i = k \sum_{i=1}^{n} x_i \qquad K \in R$$

اثبات: $\sum_{i=1}^{n} K x_i = kx_1 + kx_2 + ... + kx_n = k(x_1 + x_2 + ... + x_n) = K \sum_{i=1}^{n} x_i$

ج- $\sum_{i=1}^{n} K = n.k \qquad K \in R$

اثبات: $\sum_{i=1}^{n} K = k + k + ... + k = nk$

مثال: فرض می‌کنیم: $\sum_{i=1}^{3} x_i = 6$ و $\sum_{i=1}^{3} x_i^2 = 14$ حاصل عبارت $\sum_{i=1}^{3} (x_i - 2)^2$ را حساب کنید.

$$\sum_{i=1}^{3} (x_i - 2)^2 = \sum_{i=1}^{3} (x_i^2 - 4x_i + 4) = \sum_{i=1}^{3} x_i^2 - 4\sum_{i=1}^{3} x_i + \sum_{i=1}^{3} 4 = 14 - 4 \times 6 + 3 \times 4 = 2$$

۱-۱-۸- داده‌های آماری

میزان تغییرات صفت متغیر افراد یک جامعه آماری را که با واحد مشخصی اندازه‌گیری شده و یا بوسیله مشاهده و یا آزمایش بدست می‌آیند و معمولاً با یک عدد نشان داده می‌شوند یک داده آماری می‌گویند و مجموعه آنها را مجموعه آماری می‌گویند.

۱-۲- شاخص‌های مرکزی و پراکندگی در سری اعداد طبقه‌بندی نشده

در آمار توصیفی پیدا کردن مرکز و پراکندگی یک سری اطلاعات آماری حائز اهمیت است.
منظور از معیار مرکزی اینست که وسط و مرکز اعداد را تعیین نمائیم و مراد از معیار پراکندگی اینست که میزان انحراف اعداد از مرکز اعداد را مشخص نمائیم.

اگر تعداد داده‌های آماری زیاد نباشد برای پیدا کردن معیارهای مرکزی و پراکندگی و بررسی خواص این داده‌ها احتیاج به طبقه‌بندی آنها نیست ولی اگر تعداد اعداد و داده‌های آماری زیاد باشد لازم است آن‌ها را طبقه‌بندی و دسته‌بندی نمائیم.

ذیلاً نحوه پیدا کردن انواع شاخص‌های مرکزی و شاخص‌های پراکندگی را در سری اعداد طبقه‌بندی نشده شرح می‌دهیم:

۱ـ۲ـ۱ـ شاخص‌ها و معیارهای مرکزی

معیارهای مرکزی در سری اعداد طبقه‌بندی نشده شامل میانگین و مد(نما) و میانه می‌باشد.

۱ـ۲ـ۱ـ۱ـ میانگین (معدل یا متوسط)

الف ـ میانگین حسابی یا عددی

ساده‌ترین نوع میانگین می‌باشد و از فرمول زیر بدست می‌آید

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

مثال: میانگین حسابی اعداد ۱۱، ۹، ۷، ۳: x_i بشرح زیر است:

$$\overline{X} = \frac{۳+۷+۹+۱۱}{۴} = \frac{۳۰}{۴} = ۷/۵$$

تبصره ـ معمولاً میانگین نمونه را با $\overline{x}$ و میانگین جامعه را با μ نشان می‌دهند

ب ـ میانگین هندسی

میانگین هندسی از فرمول زیر بدست می‌آید

$$\overline{X}_G = \sqrt[n]{x_1 . x_2 ..... x_n} \quad \text{یا} \quad \overline{X}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

معمولاً هنگامی از میانگین هندسی استفاده می‌شود که بخواهند میانگین اعداد نسبی را حساب کنند. منظور از اعداد نسبی اعدادی هستند که واحد ندارند و خارج قسمت دو عدد می‌باشند.

مثال: سرمایه شرکتی در سال ۷۷ نسبت به سال ۷۶ دو برابر شده است و سرمایه همین شرکت در سال ۷۸ نسبت به سال ۷۷ هشت برابر شده است. بطور متوسط در هر سال سرمایه این شرکت چند برابر شده است؟

حل:

$$x_1 = ۲ \quad , \quad x_2 = ۸ \qquad \overline{X}_G = \sqrt{x_1 x_2} = \sqrt{۲\times۸} = ۴$$

یعنی در هر سال بطور متوسط سرمایه شرکت چهار برابر شده است

ج ـ میانگین هارمونیک یا توافقی یا همساز و یا معکوس

میانگین هارمونیک از فرمول زیر حساب می‌شود:

$$\overline{X}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}} \quad \text{یا} \quad \overline{X}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

معمولاً وقتی از میانگین هارمونیک استفاده می‌شود که بخواهند میانگین اعداد با واحد ترکیبی را حساب کنند نظیر سرعت که واحد آن متر بر ثانیه یا کیلومتر بر ساعت می‌باشد.

مثال: اتومبیلی فاصله تهران تا قم را با سرعت ۸۰ کیلومتر در ساعت طی کرده است و همان فاصله را با

سرعت ۱۲۰ کیلومتر در ساعت برگشته است. سرعت متوسط رفت و برگشت چقدر است؟

حل: $\overline{X}_H = \frac{۲}{\frac{۱}{x_۱} + \frac{۱}{x_۲}} = \frac{۲}{\frac{۱}{۸۰} + \frac{۱}{۱۲۰}} = ۹۶\text{km/h}$ $\quad x_۲ = ۱۲۰\ \text{km/h} \quad x_۱ = ۸۰\ \text{km/h}$

یعنی سرعت متوسط ۹۶ کیلومتر بر ساعت است و اگر با این سرعت مسیر رفت و برگشت را انجام دهد همان مدتی در راه است که با سرعت ۸۰ کیلومتر بر ساعت مسیر رفت و با سرعت ۱۲۰ کیلومتر بر ساعت مسیر برگشت را انجام دهد.

تبصره: بین میانگین‌های حسابی و هندسی و توافقی نامساوی زیر برقرار است

$$\overline{X} \geq \overline{X}_G \geq \overline{X}_H$$

د ـ میانگین وزنی یا موزون

فرمول میانگین وزنی بشرح زیر است:

$$\overline{X}_w = \frac{\sum_{i=1}^{n} x_i\omega_i}{\sum_{i=1}^{n} \omega_i}$$

معمولاً زمانی از این روش استفاده می‌کنند که بخواهند میانگین اعداد با ارزش‌های مختلف را حساب نمایند

مثال: نمرات یک دانشجو بشرح زیر است. میانگین نمرات این دانشجو را حساب کنید.

حل: بطوری که ملاحظه می‌شود ارزش نمرات براساس ضریب (واحد) هر درس تغییر می‌کند. بنابراین باید از روش میانگین وزنی استفاده نمود.

درس	x_i = نمره	ω_i = ضریب	$x_i\omega_i$
حسابداری	۱۴	۴	۵۶
ریاضیات	۲۰	۳	۶۰
حقوق	۱۶	۳	۴۸
اقتصاد	۱۶	۴	۶۴
مبانی مدیریت	۱۳	۴	۵۲
جمع		۱۸	۲۸۰

بنابراین معدل دانشجو عبارتست از

$$\overline{X}_w = \frac{\Sigma x_i\omega_i}{\Sigma\omega_i} = \frac{۲۸۰}{۱۸} = ۱۵/۵۵$$

پیشـ
کتاب
دانشگاهها:
کتاب
(صنعتی -
هر فص
هم چنین م
این کتا
حسابداری
مربوط به آ
ضمناً راه
نشر هستان
در خاتمه از
دکتر جه

۱-۲-۱-۲- مد یا نما

در یک سری اطلاعات آماری آن عددی که بیش از سایرین تکرار شود مد یا نما نام دارد. به عبارت دیگر مد یا نما مقداری است که فراوانی آن از سایرین بیشتر باشد. مد را با MO نشان می‌دهند.

مثال ۱:	MO = ۳	x_i : ۲-۳-۴-۳-۵-۷
مثال ۲:	MO = ۲ و ۳	x_i : ۲-۳-۴-۳-۵-۲
مثال ۳:	MO = ندارد	x_i : ۲-۳-۴-۳-۴-۲

مثال ۴: سازمان صدا و سیما می‌خواهد برای پخش یک برنامه پربیننده ساعت مناسبی را تعیین کند برای این منظور از بینندگان خود نظرخواهی می‌کند. نتیجه نظرخواهی بشرح زیر است:
۲۰۰۰ نفر زمان مناسب را ساعت ۲۲ و ۱۸۰۰ نفر زمان مناسب را ساعت ۲۰ و ۱۹۰۰ نفر هم زمان مناسب را ساعت ۱۰ تعیین کرده‌اند. برای تعیین زمان مناسب از چه معیار مرکزی باید استفاده شود.
جواب: بهترین معیار استفاده از مد است یعنی ساعت ۲۲ مناسب است.

۱-۲-۱-۳- میانه

در یک سری اطلاعات آماری عددی که از پنجاه درصد اعداد کوچکتر و از پنجاه درصد اعداد بزرگتر باشد میانه نام دارد. میانه را با Md نشان می‌دهند.

برای پیدا کردن میانه ابتدا باید اعداد را به صورت صعودی مرتب نمود سپس به کمک فرمول $C_{Md} = \frac{n}{۲} + \frac{۱}{۲}$ که در آن n تعداد اعداد است محل میانه را بدست آورد و سپس خود میانه را پیدا نمود.
در مثالهای زیر میانه اعداد را پیدا کنید:

مثال ۱: x_i : ۲-۵-۳-۷-۴

اعداد را به صورت صعودی مرتب می‌کنیم: ۲-۳-۴-۵-۷

$C_{Md} = \frac{n}{۲} + \frac{۱}{۲} = \frac{۵}{۲} + \frac{۱}{۲} =$ ۳ محل میانه $\Rightarrow Md =$ ۴ میانه

مثال ۲:

$x_i:$ ۲-۵-۳-۷-۴-۸

اعداد را به صورت صعودی مرتب می‌کنیم:

۲-۳-۴-۵-۷-۸

$$C_{Md} = \frac{n}{۲} + \frac{۱}{۲} = \frac{۶}{۲} + \frac{۱}{۲} = ۳/۵ \text{ محل میانه } \Rightarrow Md = \frac{۴+۵}{۲} = ۴/۵ \text{ میانه}$$

مثال ۳: دریافتی ماهانه ده نفر کارکنان یک شرکت برحسب هزارتومان بشرح زیر است:

$x_i:$ ۲۵ − ۲۵ − ۲۶ − ۲۶/۵ − .... ۳۵ − ۴۰۰

بهترین معیار مرکزی که نشاندهنده متوسط دریافتی ماهانه کارکنان این شرکت باشد کدام است؟

جواب: بهترین معیار مرکزی،میانه است.

۱-۲-۲- شاخص‌های پراکندگی

برای تعیین پراکندگی مقادیر یک متغیر از شاخص‌های زیر استفاده می‌گردد.

۱-۲-۲-۱- طول فاصله تغییرات ـ دامنه تغییرات ـ میدان تغییرات

طول فاصله تغییرات از فرمول $R = x_{max} - x_{min}$ بدست می‌آید.

مثال: دامنه تغییرات اعداد زیر کدام است:

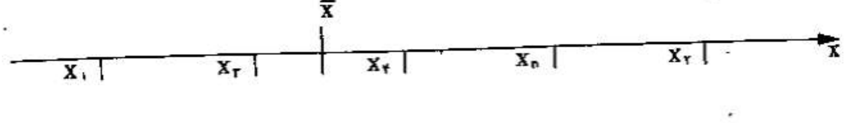
$x_i:$ ۳-۱-۲-۵-۴-۲-۷-۱-۳-۸

حل:

$R =$ ۸ − ۱ = ۷

۱-۲-۲-۲- انحراف متوسط از میانگین

برای تعیین انحراف متوسط از میانگین از فرمول زیر استفاده می‌شود.



$$AD_{\bar{x}} = \frac{\Sigma |x_i - \bar{x}|}{n}$$

مثال: انحراف متوسط از میانگین را برای اعداد ۲ و ۳ و ۷ بدست آورید.

x_i	$x_i-\bar{x}$	$\lvert x_i-\bar{x}\rvert$
۲	۲-	۲
۳	۱-	۱
۷	۳	۳
۱۲	۰	۶

حل: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{12}{3} = 4$

$$AD_{\bar{x}} = \frac{\Sigma\,|x_i-\bar{x}|}{n} = \frac{6}{3} = 2$$

تبصره: $\Sigma(x_i-\bar{x}) = 0$ است

۱-۲-۲-۳- واریانس - پراکندگی - متوسط مجذورات انحرافات از میانگین

چون در ریاضیات عملیات روی قدر مطلق مشکل است بنابراین بهتر است از مجذور کردن اعداد که نظیر قدر مطلق مقادیر مثبت و منفی را مثبت می‌کند استفاده نمود. در نتیجه فرمول واریانس بشرح زیر می‌باشد.

$$\sigma^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$$

تبصره: اگر صورت فرمول فوق را بسط دهیم به فرمولهای مشابه زیر می‌رسیم:

$$\sigma^2 = \frac{1}{n}\left[\Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2\right]$$

$$\sigma^2 = \frac{1}{n}\left[\Sigma x_i^2 - n(\bar{x})^2\right] = \frac{1}{n}\Sigma x_i^2 - \bar{x}^2$$

اثبات: از بسط $\Sigma(x_i-\bar{x})^2$ نتایج زیر حاصل می‌شود:

$$\Sigma(x_i-\bar{x})^2 = \Sigma(x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \Sigma x_i^2 - 2\bar{x}\,\Sigma x_i + \Sigma\bar{x}^2$$
$$= \Sigma x_i^2 - 2\,\frac{\Sigma x_i}{n}\,.\,\Sigma x_i + n\bar{x}^2 = \Sigma x_i^2 - \frac{2}{n}(\Sigma x_i)^2 + n\left(\frac{\Sigma x_i}{n}\right)^2$$
$$= \Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2$$

مثال: واریانس اعداد ۲ و ۳ و ۷ را پیدا کنید:

حل: میانگین این اعداد $\bar{x} = \frac{\Sigma x_i}{n} = ۴$ می‌باشد. اکنون واریانس را از هر سه فرمول فوق حساب می‌کنیم.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
۲	-۲	۴	۴
۳	-۱	۱	۹
۷	۳	۹	۴۹
۱۲	۰	۱۴	۶۲

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{۱۴}{۳} = ۴/۶۶$$

$$\sigma^2 = \frac{1}{n}\left[\Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2\right] = \frac{۱}{۳}\left[۶۲ - \frac{۱}{۳}(۱۲)^2\right] = \frac{۱۴}{۳} = ۴/۶۶$$

$$\sigma^2 = \frac{1}{n}\left[\Sigma x_i^2 - n\bar{x}^2\right] = \frac{۱}{۳}\left[۶۲ - ۳(۴)^2\right] = \frac{۱۴}{۳} = ۴/۶۶$$

تبصره: اگر میانگین اعداد یعنی $\bar{x}$ عدد صحیح نباشد بهتر است از فرمول دوم استفاده شود.

توجه: معمولاً واریانس جامعه را با σ^2 و واریانس نمونه را با s^2 نشان می‌دهند که فرمول محاسبه s^2 بشرح زیر است:

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

۱-۲-۲-۴- انحراف معیار یا انحراف استاندارد

اگر از واریانس جذر بگیریم انحراف معیار یا انحراف استاندارد بدست می‌آید

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

تبصره: واحد واریانس مربع واحد اعداد (یعنی x) می‌باشد. در حالی که انحراف معیار دارای واحدی نظیر اعداد (یعنی x) و میانگین (یعنی $\bar{x}$) و انحراف متوسط از میانگین (یعنی $AD_{\bar{x}}$) می‌باشد و مقدار انحراف معیار یعنی σ به $AD_{\bar{x}}$ نزدیک است و با آن قابل مقایسه می‌باشد.

مثال: انحراف معیار
حل: ابتدا باید و
$\sqrt{۴/۶۶} = ۲/۱۶$

۱-۲-۲-۵- ضریب
برای مقایسه پر
اندازه‌گیری دو جام

تبصره: ضریب وار
جامعه ناهمگن منا
مثال: ضریب واریا

۱-۲-۲-۶- انحراف

۱-۲-۲-۷- انحراف

۱-۲-۲-۸- چارکها

x

x_8

در یک سری

مثال: انحراف معیار اعداد ۲ و ۳ و ۷ را حساب کنید.

حل: ابتدا باید واریانس را حساب نمائیم که مقدار آن $\sigma^2 = ۴/۶۶$ می‌باشد و بنابراین $\sigma = \sqrt{۴/۶۶} = ۲/۱۶$ می‌باشد که به انحراف متوسط از میانگین یعنی ۲ نزدیک است.

۱-۲-۲-۵- ضریب واریانس - ضریب تغییرات - ضریب پراکندگی

برای مقایسه پراکندگی دو جامعه مخصوصاً هنگامی که میانگین آنها برابر نباشد و یا واحد اندازه‌گیری دو جامعه یکسان نباشد می‌توان از ضریب واریانس بشرح زیر استفاده نمود.

$$CV = \frac{\sigma}{\bar{x}} \times ۱۰۰\%$$

تبصره: ضریب واریانس واحد ندارد و یک عدد نسبی است بهمین جهت برای مقایسه پراکندگی دو جامعه ناهمگن مناسب است.

مثال: ضریب واریانس اعداد ۲ و ۳ و ۷ را بدست آورید.

$$CV = \frac{\sigma}{\bar{x}} \times ۱۰۰ = \frac{۲/۱۶}{۴} \times ۱۰۰ = ۵۴\%$$

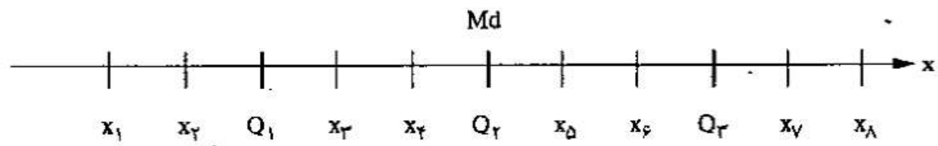
۱-۲-۲-۶- انحراف متوسط از میانه

$$AD_{Md} = \frac{\Sigma |x_i - Md|}{n}$$

۱-۲-۲-۷- انحراف متوسط از مد

$$AD_{Mo} = \frac{\Sigma |x_i - Mo|}{n}$$

۱-۲-۲-۸- چارکها و انحراف چارکها و ضریب انحراف چارکها



در یک سری داده‌های آماری چارک اول عددی است که ۲۵٪ اعداد از آن کمتر و ۷۵٪ اعداد از آن

بیشتر باشد. بهمین ترتیب چارک دوم (میانه) و چارک سوم تعریف می‌شود برای بدست آوردن چارکها ابتدا باید اعداد را بصورت صعودی مرتب نمود سپس طبق فرمول $C_{Q_a} = \frac{a.n}{4} + \frac{1}{2}$ محل چارک aام را پیدا نمود ($a = 1،2،3$) سپس بکمک آن ها خود چارکها پیدا خواهند شد.

انحراف چارکها و ضریب انحراف چارکها از فرمولهای زیر بدست می‌آیند.

$$Q = \frac{Q_3 - Q_1}{2}$$ انحراف چارکها

$$K = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$ ضریب انحراف چارکها

مثال: با فرض داده‌های زیر چارک اول و چارک سوم را محاسبه کنید و انحراف چارکها و ضریب انحراف چارکها را بدست آورید.

x_i : ۱۷ - ۴ - ۴ - ۶ - ۷ - ۱۶ - ۱۲ - ۱۰ - ۸ - ۴

حل: ابتدا اعداد را بصورت صعودی مرتب می‌کنیم.

x_i : ۴ - ۴ - ۴ - ۶ - ۷ - ۸ - ۱۰ - ۱۲ - ۱۶ - ۱۷

اکنون چارکها را پیدا می‌کنیم.

چارک اول
$$C_{Q_1} = \frac{a.n}{4} + \frac{1}{2} = \frac{1 \times 10}{4} + \frac{1}{2} = 3 \quad \Rightarrow Q_1 = 4$$

چارک سوم
$$C_{Q_3} = \frac{a.n}{4} + \frac{1}{2} = \frac{3 \times 10}{4} + \frac{1}{2} = 8 \quad \Rightarrow Q_3 = 12$$

انحراف چارکها
$$Q = \frac{Q_3 - Q_1}{2} = \frac{12-4}{2} = 4$$

ضریب انحراف چارکها
$$K = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{12-4}{12+4} = \frac{1}{2}$$

۱-۲-۲-۹- دهکها و صدکها

اگر تعداد داده‌ها زیاد باشد می‌توان از دهکها و صدکها استفاده نمود برای این منظور نظیر پیدا کردن چارکها ابتدا اعداد را بصورت صعودی می‌نویسند سپس از فرمولهای زیر محل آنها را پیدا می‌کنند و بالاخره خود دهک و صدک پیدا خواهد شد.

آمار توصیفی

۱-۲-۲-۱۰- ض

۱-۲-۲-۱۱- ض

۱-۲-۲-۱۲- ض

۱-۲-۲-۱۳- وا

۱-۲-۳- خو

۱-۲-۳-۱- مج

۱-۲-۳-۲- مج

۱-۲-۳-۳- مج

مثال: حقوق
بهره‌گیری از
بیشتر است

$C_{D_a}= \frac{a.n}{10} + \frac{1}{2}$ محل دهک aام $\Rightarrow D_a$ دهک aام

$C_{P_a}= \frac{a.n}{100} + \frac{1}{2}$ محل صدک aام $\Rightarrow P_a$ صدک aام

۱-۲-۲-۱۰- ضریب پراکندگی از میانگین

$$CD_{\bar{x}} = \frac{AD_{\bar{x}}}{\bar{x}}$$

۱-۲-۲-۱۱- ضریب پراکندگی از میانه

$$CD_{Md} = \frac{AD_{Md}}{Md}$$

۱-۲-۲-۱۲- ضریب پراکندگی از مد

$$CD_{Mo} = \frac{AD_{Mo}}{Mo}$$

۱-۲-۲-۱۳- واریانس نسبی

$$VR = \frac{\sigma^2}{\bar{x}^2}$$

۱-۲-۳- خواص انحرافات

۱-۲-۳-۱- مجموع مقادیر انحرافات از میانگین برابر صفر است.

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

۱-۲-۳-۲- مجموع مربعات انحرافات از میانگین کوچکتر از مجموع مربعات انحرافات از هر عدد دیگری است.

$$\Sigma(x_i - \bar{x})^2 < \Sigma(x_i - a)^2 \qquad a \neq \bar{x}$$

۱-۲-۳-۳- مجموع قدر مطلق انحرافات از میانه کوچکتر از مجموع قدر مطلق انحرافات از هر عدد دیگری است

$$\Sigma\, |x_i - Md| < \Sigma\, |x_i - a| \qquad a \neq Md$$

مثال: حقوق ماهانه هشت نفر از کارکنان شرکتی برحسب هزارتومان در جدول زیر منعکس است با بهره‌گیری از واریانس و انحراف معیار و ضریب واریانس تحقیق نمائید که ثبات درآمد در بین کارگران بیشتر است یا در بین کارمندان این شرکت.

کارگران = x_i	کارمندان = y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
۸۰	۴۰	۳۰	-۱۰	۹۰۰	۱۰۰
۳۰	۶۰	-۲۰	۱۰	۴۰۰	۱۰۰
۵۰	۳۰	۰	-۲۰	۰	۴۰۰
۴۰	۷۰	-۱۰	۲۰	۱۰۰	۴۰۰
۲۰۰	۲۰۰	۰	۰	۱۴۰۰	۱۰۰۰

حل: میانگین حقوق کارگران
$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{۲۰۰}{۴} = ۵۰$$

واریانس حقوق کارگران
$$\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{۱۴۰۰}{۴} = ۳۵۰$$

انحراف معیار حقوق کارگران
$$\sigma_x = \sqrt{۳۵۰} = ۱۸/۷۱$$

ضریب واریانس حقوق کارگران
$$(CV)_x = \frac{\sigma_x}{\bar{x}} \times ۱۰۰ = \frac{۱۸/۷۱}{۵۰} = \%۳۷/۴۲$$

میانگین حقوق کارمندان
$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{۲۰۰}{۴} = ۵۰$$

واریانس حقوق کارمندان
$$\sigma_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n} = \frac{۱۰۰۰}{۴} = ۲۵۰$$

انحراف معیار حقوق کارمندان
$$\sigma_y = \sqrt{۲۵۰} = ۱۵/۸۱$$

ضریب واریانس حقوق کارمندان
$$(CV)_y = \frac{\sigma_y}{\bar{y}} \times ۱۰۰ = \frac{۱۵/۸۱}{۵۰} = \%۳۱/۶۲$$

چون پراکندگی حقوق کارمندان کمتر است بنابراین ثبات درآمد کارمندان بیشتر است.

۱-۳- جمع‌آوری اطلاعات آماری و طبقه‌بندی آنها

اطلاعات آماری همواره بصورت کمی و به صورت ارقام بیان می‌شوند و این ارقام یا از طریق اندازه‌گیری و یا از طریق شمارش حاصل می‌شوند. این اطلاعات را آمارگران جمع‌آوری و آنها را دسته‌بندی یا طبقه‌بندی می‌کنند.

۱ـ۳ـ۱ـ جدول توزیع فراوانی نیمه طبقه‌بندی (جدول نوع اول)

اگر مقادیر مشاهده شده برای متغیر زیاد متنوع نباشد از جدول توزیع فراوانی نیمه طبقه‌بندی استفاده می‌شود.

مثال: داده‌های زیر مفروضند آنها را در جدول توزیع فراوانی نیمه طبقه‌بندی تنظیم نمائید.

x_i : ۷-۱-۱۱-۴-۷-۱۳-۵-۷-۱۱-۵-۱۳-۵-۷-۱۱-۱۵-۴-۷-۵-۱۳-۴

حل: جدول توزیع فراوانی را بشرح زیر تنظیم می‌نمائیم.

x_i	خط و نشان $\bcancel{////}$	فراوانی مطلق F_i
۱	/	۱
۴	///	۳
۵	////	۴
۷	$\bcancel{////}$	۵
۱۱	///	۳
۱۳	///	۳
۱۵	/	۱
		$N = \Sigma F_i = ۲۰$

۱ـ۳ـ۲ـ جدول توزیع فراوانی طبقه‌بندی کامل (جدول نوع دوم)

هرگاه تعداد اطلاعات زیاد باشد آنها را در چند طبقه دسته‌بندی می‌کنند.

۱ـ۳ـ۲ـ۱ـ انتخاب تعداد و عرض طبقات

تعداد طبقات نباید آنقدر زیاد باشد که حجم کار زیاد شود و نه آنقدر کم باشد که دقت را کم کند. معمولاً تعداد طبقات نباید از ۵ کمتر و از ۲۵ بیشتر شود. برای تعیین طبقات فرمولها و دستورالعملهای مختلفی پیشنهاد شده است که از آن جمله فرمول $k = ۱ + ۳/۳۲\log N$ می‌باشد که در آن N تعداد داده‌ها و

K تعداد طبقات می‌باشد.

پس از تعیین تعداد طبقات عرض طبقات از فرمولهای زیر بدست می‌آید:

دامنه تغییرات $= R = x_{max} - x_{min}$ عرض طبقات $= I = \frac{R}{K}$

مثال: اطلاعات حاصل از یک نمونه ۲۰ تایی از مصنوعات یک کارخانه که به منظور تحقیق در مورد وزن مصنوعات اخذ شده است برحسب کیلوگرم به شرح زیر است.

۲۰/۴-۲۷/۶-۲۲/۲-۲۷/۱-۲۰/۵-۲۵/۰-۲۳/۷-۲۱/۵-۲۶-۲۴/۲-۲۱/۱-۲۱/۵-۲۸/۵-۲۲/۲-۲۳/۱-۲۸/۰-۲۸/۴-۲۰/۱-۲۶/۱-۲۹/۹

جدول توزیع فراوانی مطلق و نسبی و فراوانی تجمعی مطلق و فراوانی تجمعی نسبی را برای اطلاعات فوق تنظیم نمایید.

حل: برای دقت بیشتر ابتدا تعداد طبقات را k = ۱۰ فرض می‌کنیم و سپس عرض طبقات را حساب می‌کنیم.

$K = ۱۰ \quad R = ۲۹/۹ - ۲۰/۱ = ۹/۸ \quad I = \frac{۹/۸}{۱۰} = ۰/۹۸ \simeq ۱$

طبقات CL	فراوانی مطلق F_i	نماینده طبقه x_i	فراوانی نسبی $f_i = \frac{F_i}{N}$	فراوانی تجمعی مطلق $F_c = \Sigma F_i$	فراوانی تجمعی نسبی $f_c = \Sigma f_i$
۲۰-۲۱	۳	۲۰/۵	۰/۱۵	۳	۰/۱۵
۲۱-۲۲	۳	۲۱/۵	۰/۱۵	۶	۰/۳۰
۲۲-۲۳	۲	۲۲/۵	۰/۱۰	۸	۰/۴۰
۲۳-۲۴	۲	۲۳/۵	۰/۱۰	۱۰	۰/۵۰
۲۴-۲۵	۱	۲۴/۵	۰/۰۵	۱۱	۰/۵۵
۲۵-۲۶	۱	۲۵/۵	۰/۰۵	۱۲	۰/۶۰
۲۶-۲۷	۲	۲۶/۵	۰/۱۰	۱۴	۰/۷۰
۲۷-۲۸	۲	۲۷/۵	۰/۱۰	۱۶	۰/۸۰
۲۸-۲۹	۳	۲۸/۵	۰/۱۵	۱۹	۰/۹۵
۲۹-۳۰	۱	۲۹/۵	۰/۰۵	۲۰	۱
	$N = \Sigma F_i = ۲۰$		$\Sigma f_i = ۱$		

۱ـ۳ـ۳ـ نمایش نموداری اطلاعات طبقه‌بندی شده

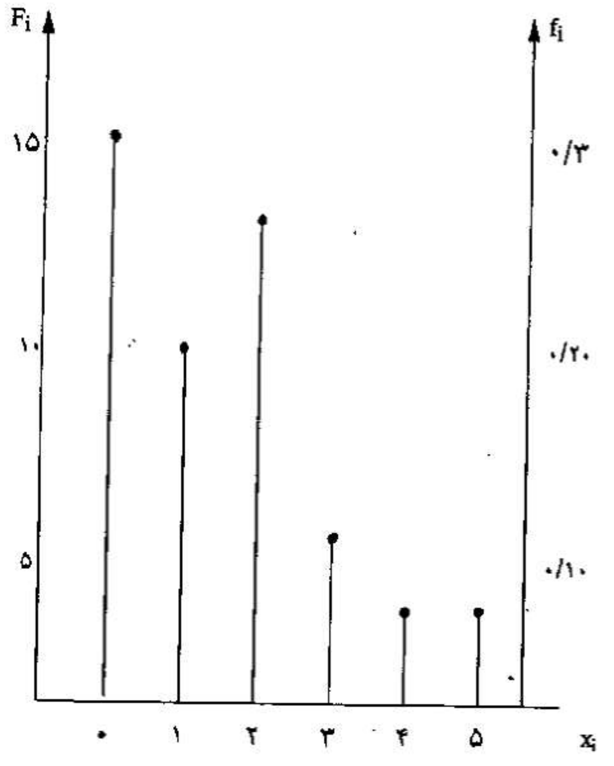
اگر اطلاعات طبقه‌بندی شده را بصورت نمودار رسم کنیم، نمودار بهتر از جدول اطلاعات را نشان می‌دهد. بعضی از نمودارهای مهم در زیر شرح داده می‌شود.

۱ـ۳ـ۳ـ۱ـ نمودار میله‌ای

این نمودار در مواردی به کار می‌رود که متغیر تصادفی گسسته بوده و مقادیر مشاهده شده برای متغیر نیز زیاد متنوع نباشد.

مثال: فراوانی تعداد فرزندان ۵۰ خانواده در جدول زیر منعکس است نمودار میله‌ای آن را رسم کنید.

تعدادفرزندان x_i	F_i	f_i
۰	۱۵	۰/۳۰
۱	۱۰	۰/۲۰
۲	۱۳	۰/۲۶
۳	۶	۰/۱۲
۴	۳	۰/۰۶
۵	۳	۰/۰۶
	N=۵۰	۱



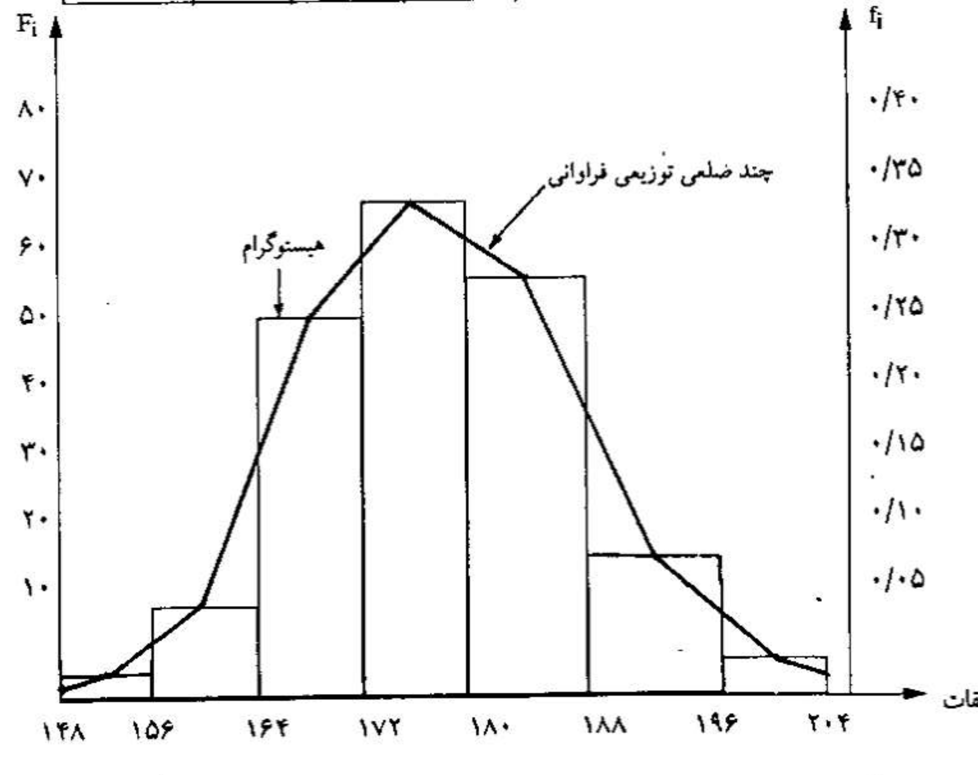
تبصره: نمودار میله‌ای برای رسم نمودار جداول نیمه طبقه‌بندی مناسب است.

۱-۳-۳-۲- هیستوگرام

در مواردی که متغیر پیوسته و یا تعداد مشاهدات زیاد باشد از این نمودار استفاده می‌گردد.

مثال: نمودار هیستوگرام جدول طبقه‌بندی شده زیر را رسم کنید.

CL	x_i	F_i	f_i
۱۴۸_۱۵۶	۱۵۲	۲	۰/۰۱
۱۵۶_۱۶۴	۱۶۰	۱۰	۰/۰۵
۱۶۴_۱۷۲	۱۶۸	۴۸	۰/۲۴
۱۷۲_۱۸۰	۱۷۶	۶۴	۰/۳۲
۱۸۰_۱۸۸	۱۸۴	۵۶	۰/۲۸
۱۸۸_۱۹۶	۱۹۲	۱۶	۰/۰۸
۱۹۶_۲۰۴	۲۰۰	۴	۰/۰۲
		N=۲۰۰	۱



تبصره: هیستوگرام برای نمایش نموداری جداول طبقه‌بندی کامل مناسب است

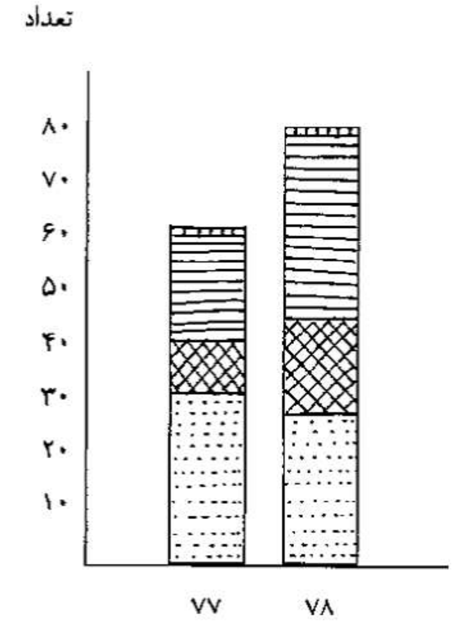
۱-۳-۳-۳- نمودار ستونی

نمودار ستونی بر دو نوع است

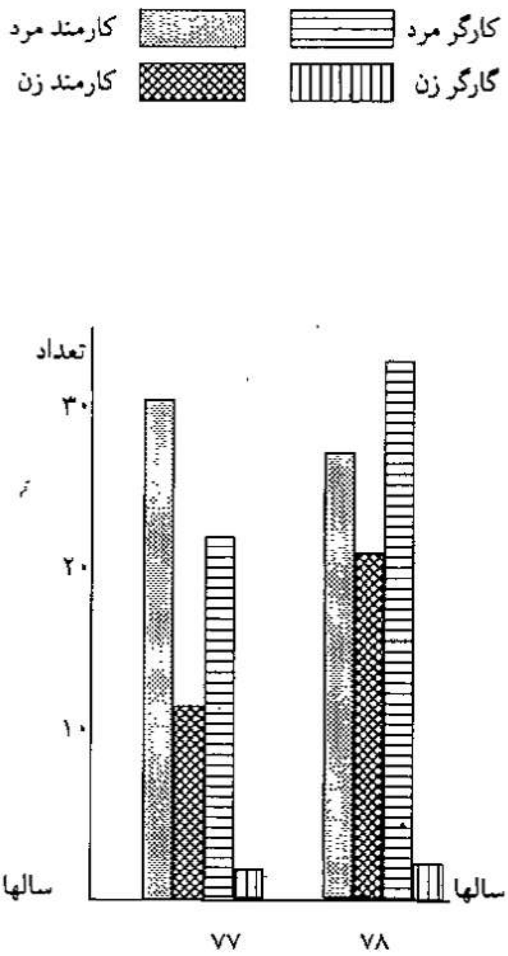
الف) نمودار ستونی مرکب   ب) نمودار ستونی متعدد

مثال: تعداد کارمندان و کارگران زن و مرد شرکتی در سالهای ۷۷ و ۷۸ به شرح زیر بوده است. نمودار آنها را رسم کنید.

سال	کارمندان		کارگران		جمع
	مرد	زن	مرد	زن	
۷۷	۳۰	۱۰	۲۰	۱	۶۱
۷۸	۲۵	۱۸	۳۶	۲	۸۱



نمودار ستونی مرکب



نمودار ستونی متعدد

۱-۳-۳-۴- نمودار دایره‌ای

در نمودار دایره‌ای هر قسمت از دایره از فرمول زیر حساب می‌شود

$$\text{قسمتی از دایره} = \frac{\text{قسمتی از جامعه}}{\text{کل جامعه}} \times ۳۶۰^\circ$$

در مثال زیر نحوه عمل شرح داده می‌شود.

مثال: در امتحان ورودی یک دانشگاه از شهرستانهای مختلف به تعداد زیر پذیرفته شده‌اند. نمودار دایره‌ای آن را رسم کنید.

شیراز ۴۰ نفر - اصفهان ۵۰ نفر - مشهد ۶۰ نفر - تبریز ۳۰ نفر

حل: کل جامعه برابر است با:

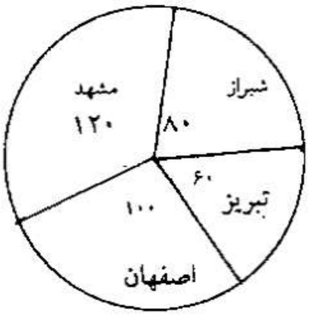
$$۳۰ + ۶۰ + ۵۰ + ۴۰ = ۱۸۰$$

$$\text{شیراز} = \frac{۴۰}{۱۸۰} \times ۳۶۰^\circ = ۸۰^\circ$$

$$\text{مشهد} = \frac{۶۰}{۱۸۰} \times ۳۶۰^\circ = ۱۲۰^\circ$$

$$\text{اصفهان} = \frac{۵۰}{۱۸۰} \times ۳۶۰^\circ = ۱۰۰^\circ$$

$$\text{تبریز} = \frac{۳۰}{۱۸۰} \times ۳۶۰^\circ = ۶۰^\circ$$



تبصره: برای رسم نمودار متغیرهای کیفی می‌توان از نمودارهای ستونی و دایره‌ای استفاده نمود.

۱-۴- شاخص‌های مرکزی و پراکندگی در سری اعداد طبقه‌بندی شده

۱-۴-۱- محاسبه شاخص‌های مرکزی

۱-۴-۱-۱- محاسبه میانگین‌ها

۱) میانگین حسابی

الف) روش فراوانی مطلق

میانگین از فرمول مقابل محاسبه می‌شود

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

در این فرمول n تعداد طبقات و N تعداد کل داده‌ها می‌باشد

ب) به روش فراوانی نسبی

میانگین از فرمول مقابل محاسبه می‌گردد

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

مثال: در جدول طبقه‌بندی زیر میانگین را بروشهای فراوانی مطلق و فراوانی نسبی بدست آورید.

CL	x_i	F_i	x_iF_i	f_i	f_ix_i
۲_ ۴	۳	۳	۹	۰/۳	۰/۹
۴_ ۶	۵	۱	۵	۰/۱	۰/۵
۶_ ۸	۷	۲	۱۴	۰/۲	۱/۴
۸_۱۰	۹	۴	۳۶	۰/۴	۳/۶
Σ		۱۰	۶۴	۱	۶/۴

$$\bar{x} = \frac{\Sigma F_i x_i}{N} = \frac{۶۴}{۱۰} = ۶/۴$$

$$\bar{x} = \Sigma f_i x_i = ۶/۴$$

۲) میانگین هندسی

از فرمول مقابل محاسبه می‌شود

$$\bar{x}_G = \sqrt[N]{x_1^{F_1} \cdot x_2^{F_2} \cdot x_3^{F_3} \ldots\ldots x_n^{F_n}} \qquad \overline{x_G} = \sqrt[N]{\prod_{i=1}^{n} x_i^{F_i}}$$

۳) میانگین هارمونیک

از فرمولهای مقابل محاسبه می‌شود

$$\bar{x}_H = \frac{N}{\frac{F_1}{x_1} + \frac{F_2}{x_2} + \ldots\ldots + \frac{F_n}{x_n}} \qquad \bar{x}_H = \frac{N}{\sum_{i=1}^{n} \frac{F_i}{x_i}}$$

تبصره: هرگاه یکی از داده‌ها قابل مقایسه با سایر داده‌ها نباشد می‌توان آن را حذف کرد. همینطور اگر فراوانی یکی از داده‌ها خیلی کمتر از فراوانی بقیهٔ داده‌ها باشد می‌توان آن را در نظر نگرفت.

۱ـ۴ـ۱ـ۲ـ محاسبه میانه

میانه را می‌توان به روش‌های زیر محاسبه نمود که نحوه عمل در مثال زیر شرح داده می‌شود

مثال: میانه را در جدول طبقه‌بندی زیر بدست آورید

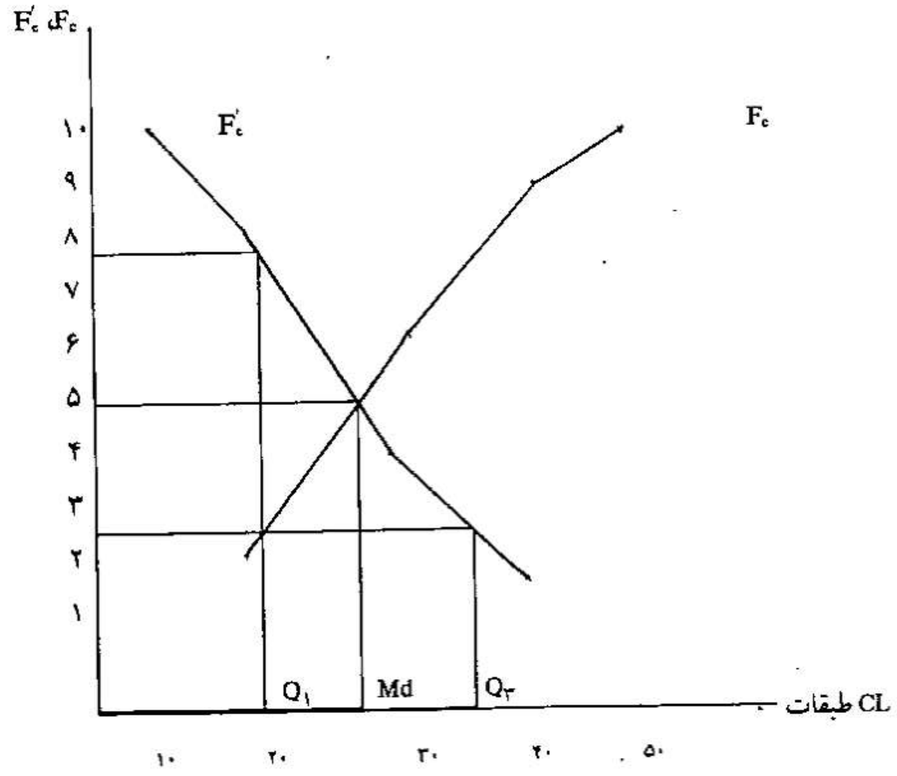
CL	F_i	F_c	F'_c
۱۰_۲۰	۲	۲	۱۰
۲۰_۳۰	۴	۶	۸
۳۰_۴۰	۳	۹	۴
۴۰_۵۰	۱	۱۰	۱
	N=۱۰		

در مثال فوق منظور از F_c و F'_c بترتیب فراوانی تجمعی کمتر از و فراوانی نسبی بیشتر از می‌باشد.

الف ـ نمودار اجایو

ابتدا نمودارهای فراوانی تجمعی کمتر از (F_c) و فراوانی تجمعی بیشتر از (F'_c) را رسم می‌کنیم. محل برخورد این دو نمودار میانه را در اختیار قرار می‌دهد.

در جدول طبقه‌بندی محل میانه از فرمول $C_{Md} = \frac{N}{2}$ بدست می‌آید. در این جدول محل میانه $C_{Md} = \frac{10}{2} = 5$ می‌باشد.



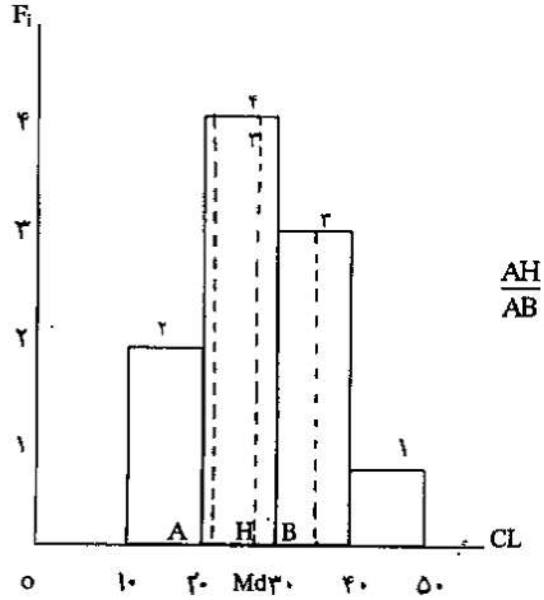
تبصره: از نمودار اجایو می‌توان چارکهای اول و سوم را هم محاسبه نمود

در جداول طبقه‌بندی محل چارک اول و سوم بترتیب از فرمولهای $C_{Q_1} = \frac{N}{4}$ و $C_{Q_3} = \frac{3N}{4}$ بدست می‌آید. در این جدول محل چارکهای اول و سوم بترتیب $C_{Q_1} = \frac{10}{4} = 2/5$ و $C_{Q_3} = \frac{30}{4} = 7/5$ می‌باشد.

ب) نمودار هیستوگرام

ابتدا نمودار هیستوگرام را رسم می‌کنیم سپس محل میانه را حساب می‌کنیم و در روی هیستوگرام محل میانه را پیدا می‌کنیم. در روی محور طبقات میانه مشخص خواهد شد.

در مثال فوق میانه از هیستوگرام بشرح زیر بدست می‌آید



$$C_{Md} = \frac{N}{۲} = \frac{۱۰}{۲} = ۵$$

$$M_d = OH = OA = AH$$

$$\frac{AH}{AB} = \frac{۳}{۴} \Rightarrow \frac{AH}{۱۰} = \frac{۳}{۴} \Rightarrow AH = ۷/۵$$

$$Md = ۲۰ + ۷/۵ = ۲۷/۵$$

تبصره: از هیستوگرام می‌توان چارکهای اول و سوم را نظیر میانه پیدا نمود.

در مثال فوق محل چارکهای اول و سوم بترتیب $C_{Q_1} = ۲/۵$ و $C_{Q_3} = ۷/۵$ می‌باشد.

ج ـ جدول توزیع فراوانی

از فرمول $C_{Md} = \frac{N}{۲}$ محل میانه را پیدا می‌کنیم و در ستون تجمعی F_c طبقه‌ای را پیدا می‌کنیم که میانه در آن قرار بگیرد و به آن طبقه میانه‌دار می‌گوئیم سپس از فرمول زیر استفاده می‌کنیم:

$$Md = L_{Md} + \frac{\frac{N}{۲} - F_c}{F_i} I$$

در این فرمول داریم:

$L_{Md} =$ حد تحتانی طبقه میانه‌دار — حد پایین طبقه میانه

$F_c =$ فراوانی تجمعی طبقه ماقبل — خود طبقه در آن حساب می‌شود

$F_i =$ فراوانی مطلق طبقه میانه‌دار

$I =$ عرض طبقه میانه‌دار

در مثال فوق محاسبه میانه از جدول بشرح زیر است.
محل میانه $C_{Md} = \frac{۱۰}{۲} = ۵$ می‌باشد بنابراین میانه در ستون فراوانی تجمعی F_c در طبقه دوم (جدول مثال ۱-۴-۱-۲-) قرار می‌گیرد و طبقه دوم میانه‌دار است. سپس با توجه به فرمول اخیر و جدول مورد اشاره میانه بشرح زیر حساب می‌شود.

$$Md = ۲۰ + \frac{\frac{۱۰}{۲} - ۲}{۴} \times ۱۰ = ۲۷/۵$$

تبصره: چارکهای اول و سوم را می‌توان از جدول بشرح زیر پیدا نمود.
ابتدا محل چارکهای اول و سوم را از فرمولهای $C_{Q_1} = \frac{N}{۴}$ و $C_{Q_3} = \frac{۳N}{۴}$ پیدا می‌کنیم سپس طبقه‌ای را که چارکها در آن قرار دارند بکمک ستون فراوانی تجمعی F_c پیدا می‌کنیم و از فرمولهای زیر خود چارکهای اول و سوم محاسبه می‌شوند.

$$Q_1 = L_{Q_1} + \frac{\frac{N}{۴} - F_c}{F_i} I \quad \text{چارک اول}$$

$$Q_3 = L_{Q_3} + \frac{\frac{۳N}{۴} - F_c}{F_i} I \quad \text{چارک سوم}$$

در مثال فوق با توجه به اینکه محل چارکهای اول و سوم بترتیب $C_{Q_1} = \frac{۱۰}{۴} = ۲/۵$ و $C_{Q_3} = \frac{۳۰}{۴} = ۷/۵$ می‌باشند و به ترتیب در طبقات دوم و سوم قرار می‌گیرند، مقدار چارکها بشرح زیر می‌باشد.

$$Q_1 = ۲۰ + \frac{۲/۵-۲}{۴} \times ۱۰ = ۲۱/۲۵ \quad \text{چارک اول}$$

$$Q_3 = ۳۰ + \frac{۷/۵-۶}{۳} \times ۱۰ = ۳۵ \quad \text{چارک سوم}$$

۱-۴-۱-۳- محاسبه مد (نما)

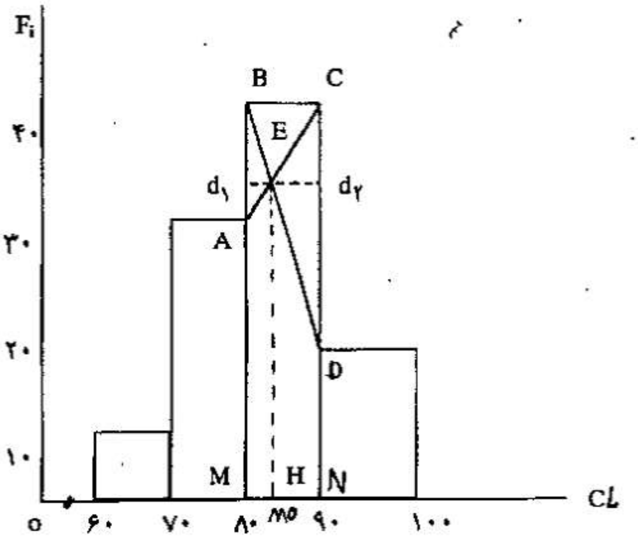
مد را در سری اعداد طبقه‌بندی شده به روش‌های زیر می‌توان محاسبه نمود که در مثال زیر شرح داده می‌شود.

مثال: مد یا نما را در جدول طبقه‌بندی شده زیر پیدا کنید.

CL	F_i	
۶۰_۷۰	۱۰	
۷۰_۸۰	۳۰	
۸۰_۹۰	۴۰	طبقه مد دار ←
۹۰_۱۰۰	۲۰	
Σ	N=۱۰۰	

الف ـ هیستوگرام

ابتدا نمودار هیستوگرام را رسم می‌کنیم سپس در طبقه‌ای که بیشترین فراوانی (ارتفاع) را دارد نظیر شکل زیر خطوط مورب را رسم می‌کنیم محل برخورد این دو خط مورب مد را نشان می‌دهد.



مقدار مد را می‌توان با توجه به تشابه مثلثهای ABE و CDE پیدا نمود. در این دو مثلث نسبت اضلاع و ارتفاعات متناظر برابرند داریم:

$$\frac{AB}{CD} = \frac{MH}{HN}$$

اگر $AB = d_1$ و $CD = d_2$ در نظر گرفته شود: داریم

$$\frac{d_1}{d_2} = \frac{MH}{HN}$$

ترکیب نسبت در مخرج می‌نمائیم

$$\frac{d_1}{d_1+d_2} = \frac{MH}{MH+HN} \Rightarrow \frac{d_1}{d_1+d_2} = \frac{MH}{MN}$$

MN=I عرض طبقه می‌باشد.

$$MH = \frac{d_1}{d_1+d_2}\ MN = \frac{d_1}{d_1+d_2}\ I$$

اکنون داریم:

$$Mo = OH = OM + MH = OM + \frac{d_1}{d_1+d_2}\ I$$

اگر بجای آنها مقادیرشان را قرار دهیم مد بدست می‌آید.

$$Mo = ۸۰ + \frac{۱۰}{۱۰+۲۰} \times ۱۰ = ۸۳/۳۳$$

ب ـ جدول توزیع فراوانی

ابتدا در ستون فراوانی مطلق طبقه‌ای را پیدا می‌کنیم که بیشترین فراوانی را داشته باشد و بآن طبقه مدار می‌گوئیم سپس با توجه به محاسبات فوق مدرا از فرمول زیر بدست می‌آوریم.

$$Mo = L_{Mo} + \frac{d_1}{d_1+d_2}\ I$$

در این فرمول داریم:

L_{Mo} = حد تحتانی طبقه مدار

تفاوت فراوانی طبقه مدار با طبقه قبل از آن

d_1 = مابه التفاوت فراوانی مطلق طبقه مدار با طبقه ماقبل

d_2 = مابه‌التفاوت فراوانی مطلق طبقه مدار با طبقه مابعد تفاوت فراوانی طبقه مدار با طبقه بعد از آن

I = عرض طبقه مدار

در مثال فوق طبقه مدار طبقه سوم می‌باشد در نتیجه داریم:

$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} I$$

$$Mo = ۸۰ + \frac{۱۰}{۱۰ + ۲۰} \times ۱۰ = ۸۳/۳۳$$

۱-۴-۲- شاخص‌های پراکندگی

شاخص‌های پراکندگی در سری اعداد طبقه‌بندی شده نظیر شاخص‌های پراکندگی در سری اعداد طبقه‌بندی نشده می‌باشند فقط باید در فرمولها فراوانی مطلق (یا فراوانی نسبی) را مورد توجه قرار داد. شاخص‌های پراکندگی در سری اعداد طبقه‌بندی شده بشرح زیر است.

۱-۴-۲-۱- انحراف متوسط از میانگین

$$AD_{\bar{x}} = \frac{\Sigma F_i |x_i - \bar{x}|}{N}$$

۱-۴-۲-۲- انحراف متوسط از میانه

$$AD_{Md} = \frac{\Sigma F_i |x_i - Md|}{N}$$

۱-۴-۲-۳- انحراف متوسط از مد

$$AD_{Mo} = \frac{\Sigma F_i |x_i - Mo|}{N}$$

۱-۴-۲-۴- واریانس

$$\sigma^2 = \frac{\Sigma F_i (x_i - \bar{x})^2}{N}$$

$$\sigma^2 = \frac{1}{N} \left[\Sigma F_i x_i^2 - \frac{1}{N} (\Sigma F_i x_i)^2\right]$$

$$\sigma^2 = \frac{1}{N} \left[\Sigma F_i x_i^2 - N\bar{x}^2\right]$$

سایر فرمولهای پراکندگی نظیر فرمولهای پراکندگی سری اعداد طبقه‌بندی نشده می‌باشد.

تبصره: در مواردی که صفت متغیر پیوسته باشد و یا مقادیر بسیار نزدیک بهم را قبول کنند حدود طبقات نظیر مثالهای فوق انتخاب می‌شود یعنی حد انتهائی هر طبقه عیناً حد ابتدائی طبقه بعد می‌باشد.

ولی در مواردی که مناسب باشد می‌توان حدود طبقات را نظیر مثال زیر انتخاب نمود در این صورت اختلاف حد انتهائی هر طبقه و حد ابتدائی طبقه بعد برابر یک واحد می‌باشد (یا بستگی به دقت اندازه اعداد دارد) در این حالت می‌توان به سهولت حدود واقعی طبقات را هم نوشت:

مثال:

فراوانی مطلق F_i	حدود واقعی طبقات	حدود طبقات
۱۵	[۹/۵_۱۹/۵)	۱۰_۱۹
۲۵	[۱۹/۵_۲۹/۵)	۲۰_۲۹
۳۰	[۲۹/۵_۳۹/۵)	۳۰_۳۹
۲۰	[۳۹/۵_۴۹/۵)	۴۰_۴۹
۱۰	[۴۹/۵_۵۹/۵)	۵۰_۵۹
N=۱۰۰		

۱ـ۴ـ۳ـ دو مثال نمونه برای محاسبه معیارهای مرکزی و پراکندگی در سری اعداد طبقه‌بندی شده

۱ـ۴ـ۳ـ۱ـ مثال برای محاسبه معیارهای مرکزی و پراکندگی در جداول نیمه طبقه‌بندی.

در یک مؤسسه بازرگانی بمنظور کنترل صحت اسناد صادره، تعداد n=۱۰۰ سند انتخاب و مورد بررسی قرار گرفت. صفت متغیر X تعداد اشتباه در هر سند می‌باشد که جدول توزیع فراوانی آن بصورت زیر تنظیم شده است.

X_i	۰	۱	۲	۳	۴
F_i	۱۰	۴۰	۱۰	۲۰	۲۰

مطلوبست

الف ـ محاسبه میانگین ـ واریانس و انحراف معیار X یعنی تعداد اشتباهات در هر سند

ب ـ محاسبه مد و میانه و چارکهای اول و سوم

ج ـ ضریب واریانس

د ـ اگر سندی را به تصادف انتخاب کنیم چقدر احتمال دارد که تعداد اشتباهات آن سند یک و یا دو اشتباه باشد.

حل: الف

x_i	F_i	F_ix_i	$x_i-\bar{x}$	$(x_i-\bar{x})^۲$	$F_i(x_i-\bar{x})^۲$
۰	۱۰	۰	۲-	۴	۴۰
۱	۴۰	۴۰	۱-	۱	۴۰
۲	۱۰	۲۰	۰	۰	۰
۳	۲۰	۶۰	۱	۱	۲۰
۴	۲۰	۸۰	۲	۴	۸۰
Σ	N=۱۰۰	۲۰۰			۱۸۰

$$\bar{x} = \frac{\Sigma F_i x_i}{N} = \frac{۲۰۰}{۱۰۰} = ۲$$

$$\sigma^۲ = \frac{\Sigma F_i(x_i-\bar{x})^۲}{N} = \frac{۱۸۰}{۱۰۰} = ۱/۸ \qquad \sigma = \sqrt{۱/۸} = ۱/۳۴$$

توجه: اگر بخواهیم نتایج این نمونه را به جامعه تعمیم دهیم از فرمول $s^۲ = \frac{\Sigma F_i (x_i-\bar{x})^۲}{N-۱}$ استفاده می‌کنیم.

ب ـ بیشترین فراوانی مربوط به ۱ اشتباه است که در ۴۰ سند اتفاق افتاده است بنابراین Mo = ۱

محل میانه برابر $C_{Md} = \frac{N}{۲} + \frac{۱}{۲} = \frac{۱۰۰}{۲} + \frac{۱}{۲} = ۵۰/۵$ است میانه بین ۱ و ۲ قرار می‌گیرد بنابراین $Md = \frac{۱+۲}{۲} = ۱/۵$ است.

محل چارک اول $۲۵/۵ = \frac{۱}{۲} + \frac{۱۰۰}{۴} = \frac{۱}{۲} + \frac{N}{۴} = C_{Q_1}$ بنابراین $Q_1 = ۱$

محل چارک سوم $۷۵/۵ = \frac{۱}{۲} + \frac{۳۰۰}{۴} = \frac{۱}{۲} + \frac{۳N}{۴} = C_{Q_3}$ بنابراین $Q_3 = ۳$

ج - $CV = \frac{\sigma}{\bar{X}} \times ۱۰۰ = \frac{۱/۳۴}{۲} \times ۱۰۰ = \%۶۷$

د - تعداد سندهائی که دارای یک و یا دو اشتباه می‌باشند به ترتیب ۱۰ و ۴۰ سند می‌باشد بنابراین:

$P(x= ۱ \text{ یا } ۲) = \frac{۱۰ + ۴۰}{۱۰۰} = ۰/۵$

۱-۴-۳-۲- مثال برای محاسبه معیارهای مرکزی و پراکندگی در جداول طبقه‌بندی کامل

جدول زیر توزیع فراوانی حقوق ماهانه ۶۰ کارمند یک شرکت را نشان می‌دهد.

CL - برحسب هزار تومان	۳۰-۵۰	۵۰-۷۰	۷۰-۹۰	۹۰-۱۱۰	۱۱۰-۱۳۰
F_i - فراوانی مطلق	۸	۱۵	۲۵	۸	۴

مطلوبست:

الف - محاسبه میانگین و واریانس و انحراف معیار حقوق ماهانه کارمندان شرکت

ب - رسم هیستوگرام و تعیین مد و میانه و چارکها روی آن

ج - محاسبه مد و میانه و چارکها از جدول

د - محاسبه ضریب واریانس

ه - اگر یک کارمند را به تصادف انتخاب کنیم چقدر احتمال دارد که حقوق ماهانه او بین حداقل ۵۰ ولی کمتر از ۹۰ هزار تومان باشد.

حل: الف

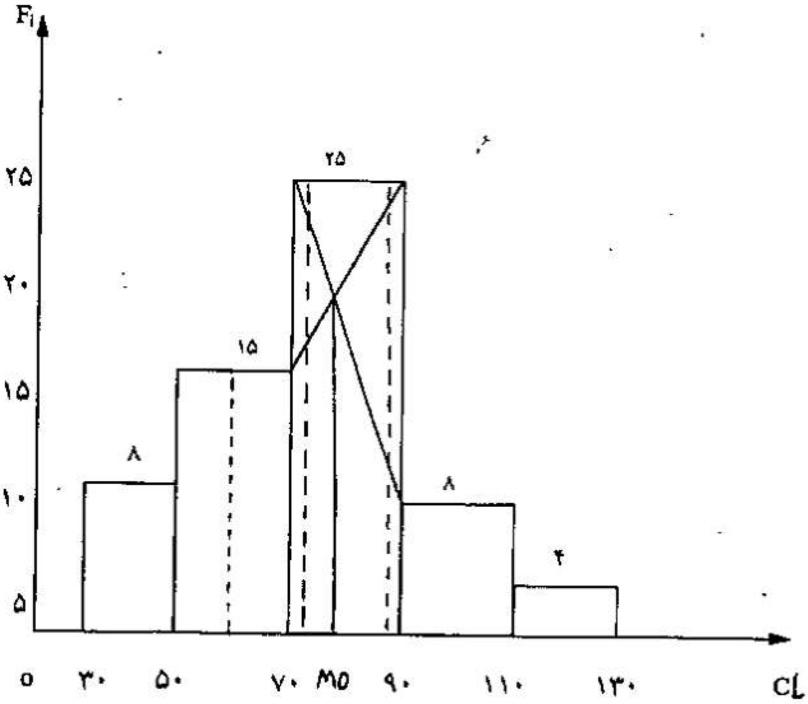
CL	F_i	x_i	F_ix_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$F_i(x_i-\bar{x})^2$	F_c	
۳۰-۵۰	۸	۴۰	۳۲۰	-۳۵	۱۲۲۵	۹۸۰۰	۸	طبقه چارک
۵۰-۷۰	۱۵	۶۰	۹۰۰	-۱۵	۲۲۵	۳۳۷۵	۲۳	اول ←
۷۰-۹۰	۲۵ طبقه مددار	۸۰	۲۰۰۰	۵	۲۵	۶۲۵	۴۸	← طبقه میانه‌دار
۹۰-۱۱۰	۸	۱۰۰	۸۰۰	۲۵	۶۲۵	۵۰۰۰	۵۶	وطبقه چارک سوم
۱۱۰-۱۳۰	۴	۱۲۰	۴۸۰	۴۵	۲۰۲۵	۸۱۰۰	۶۰	
Σ	N=۶۰		۴۵۰۰			۲۶۹۰۰		

$\bar{x} = \frac{\Sigma F_ix_i}{N} = \frac{۴۵۰۰}{۶۰} = ۷۵$ میانگین

$\sigma^2 = \frac{\Sigma F_i(x_i-\bar{x})^2}{N} = \frac{۲۶۹۰۰}{۶۰} = ۴۴۸/۳۳$ واریانس

$\sigma = \sqrt{۴۴۸/۳۳} = ۲۱/۱۷$ انحراف معیار

ب ـ



$C_{Md} = \frac{N}{2} = \frac{60}{2} = 30$ محل میانه

$C_{Q_1} = \frac{N}{4} = \frac{60}{4} = 15$ محل چارک اول

$C_{Q_3} = \frac{3N}{4} = \frac{180}{4} = 45$ محل چارک سوم

ج -

مد
$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} \times I = 70 + \frac{10}{10+17} \times 20 = 77/4$$

میانه
$$Md = L_{Md} + \frac{\frac{N}{2} - F_c}{F_i} \times I = 70 + \frac{30-23}{25} \times 20 = 75/6$$

چارک اول
$$Q_1 = L_{Q_1} + \frac{\frac{N}{4} - F_c}{F_i} \times I = 50 + \frac{15-8}{15} \times 20 = 59/33$$

چارک سوم
$$Q_3 = L_{Q_3} + \frac{\frac{3N}{4} - F_c}{F_i} \times I = 70 + \frac{45-23}{25} \times 20 = 87/6$$

د -
$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{21/17}{75} \times 100 = \%28/23$$

هـ -
$$P(50 \le x < 90) = \frac{15+25}{60} = \frac{40}{60} = 0/67$$