درس سوم : معیارهای (شاخص های) پراکندگی

شاخص های پراکندگی میزان پراکندگی یا تفسیر داده ها را نشان می دهد.

دامنه تغییرات ، انحراف از میانگین ، انحراف معیار (انحراف استاندارد) ، واریانس و ضریب تغییرات داده ها ، چهار معیار پراکندگی هستند.

الف) دامنهٔ تغییرات : ساده ترین شاخص پراکندگی ، دامنهٔ تغییرات است . دامنهٔ تغییرات عبارتست از تفاضل بزرگترین و کوچکترین داده ، که با نماد R نمایش می دهیم.

اگر α کوچکترین داده و β بزرگترین داده فرض شود آنگاه : $R = \beta - \alpha$

مثال : دامنهٔ تغییرات داده های ۱۸ ، ۱۶ ، ۱۷ ، ۱۴ ، ۱۲ ، ۱۱ ، ۹ ، ۷ ، ۸ ، ۷ ، ۱۶ ، ۱۶ ، ۵ و ۱۲ را بیابید.

کوچکترین داده: $\alpha = ۵$ و بزرگترین داده: $\beta = ۱۸ \implies R = ۱۸ - ۵ = ۱۳$

توجه : بزرگی دامنهٔ تغییرات ، نشان دهندهٔ تفاوت زیاد در جامعه می باشد ، هرچقدر این دامنه کمتر باشد ، افراد جامعه از لحاظ این متغیر به هم نزدیکترند. اگر دامنهٔ تغییرات صفر باشد ، تمام داده ها با هم برابرند.
دامنهٔ تغییرات خیلی سریع و به سهولت محاسبه می شود ، اما چون در محاسبهٔ آن فقط از دو عدد (بزرگترین و کوچکترین) استفاده می شود قادر به توصیف توزیع داده ها به صورت حقیقی نیست.

نکته : اگر داده ها را با مقدار ثابت b جمع کنیم ، تاثیری بر روی دامنهٔ تغییرات ندارد یعنی : $R_{x+b} = R_x$
اما اگر داده ها را در عدد ثابت a ضرب کنیم ، آنگاه دامنهٔ تغییرات در $|a|$ ضرب می شود.

یعنی : $R_{ax} = |a| R_x$

بنابراین در حالت کلی می توان نوشت : $R_{ax+b} = |a| R_x$

مثال : اگر دامنهٔ تغییرات داده های x_1 و x_2 و ... و x_n برابر ۲ باشد ، آنگاه دامنهٔ تغییرات داده های $-5x_1+3$ و $-5x_2+3$ و ... و $-5x_n+3$ را تعیین کنید.

$R_{-5x+3} = |-5| R_x = 5 \times 2 = 10$

ب) **انحراف از میانگین:** اگر $\bar{x}$ میانگین داده های x_1 و x_2 و ... و x_n باشد آنگاه برای هر داده x_i مقدار $x_i - \bar{x}$ را انحراف از میانگین آن داده گوییم.

مثال: انحراف از میانگین داده های ۱۲ و ۷ و ۵ و ٤ و ۲ را در یک جدول بنویسید.

برای این منظور ابتدا میانگین داده ها را حساب می کنیم:

$$\bar{x} = \frac{2+4+5+7+12}{5} = 6$$

داده	۱۲	۷	۵	۴	۲
انحراف از میانگین داده	$12-6=6$	$7-6=1$	$5-6=-1$	$4-6=-2$	$2-6=-4$

توجه: مجموع انحراف از میانگین داده ها برابر صفر است. به عنوان نمونه در مثال فوق داریم:

$$(-4)+(-2)+(-1)+(1)+(6)=0$$

مثال: به ازای چه مقدار از a، انحراف از میانگین چند داده آماری به صورت ۲- و ۲- و ۴ و a و ۳ و ۱ می باشد؟

$$\text{مجموع انحراف از میانگین} = 0 \Rightarrow 1+3+a+4+(-2)+(-2)=0 \Rightarrow a=-4$$

تمرین: ثابت کنید مجموع انحراف از میانگین داده ها برابر صفر است.

فرض کنید $\bar{x}$ میانگین داده های x_1 و x_2 و ... و x_n باشد بنابراین: $\quad ①\ x_1+x_2+\dots+x_n = n\bar{x}$

$$\text{مجموع انحراف از میانگین داده ها} = (x_1-\bar{x})+(x_2-\bar{x})+\dots+(x_n-\bar{x})$$

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

$$\overset{①}{=} n\bar{x}-n\bar{x}=0$$

مثال: با توجه به جدول روبه رو، حاصل $a+b+c+d$ را بدست آورید.

x_i	1	a	b	c
$x_i-\bar{x}$	d	2	1	5

$$\text{مجموع انحراف از میانگین} = 0 \Rightarrow d+2+1+5=0 \Rightarrow d=-8$$

$$x_1-\bar{x}=-8 \Rightarrow 1-\bar{x}=-8 \Rightarrow \bar{x}=9$$

$$x_2-\bar{x}=2 \Rightarrow a-9=2 \Rightarrow a=11 \quad ; \quad x_3-\bar{x}=1 \Rightarrow b-9=1 \Rightarrow b=10$$

$$x_4-\bar{x}=5 \Rightarrow c-9=5 \Rightarrow c=14 \quad \Longrightarrow a+b+c+d=11+10+14-8=27$$

پ) انحراف معیار و واریانس داده ها :

در آمار، یک معیار سنجش برای میزان پراکندگی داده ها حول میانگینشان انحراف معیار است که با نماد σ نمایش می دهیم.

مربع انحراف معیار داده ها را واریانس داده ها نامیده و با نماد σ^2 نشان می دهیم.

برای محاسبهٔ واریانس از فرمول زیر استفاده می کنیم:

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

طبق تعریف واضح است که برای به دست آوردن انحراف معیار کافیست جذر واریانس را حساب کنیم.

سؤال: روابط محاسبهٔ واریانس و انحراف معیار را با نماد سیگما بنویسید.

$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$: واریانس ؛ انحراف معیار : $\sigma=\sqrt{\sigma^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

مثال: واریانس و انحراف معیار داده های ۴ ، ۱ و ۰ ، ۲ ، ۳ را تعیین کنید.

ابتدا میانگین داده ها را حساب می کنیم: $\bar{x}=\frac{3+2+0+1+4}{5}=2$

$$\sigma^2=\frac{1}{5}\left(1^2+0^2+(-2)^2+(-1)^2+2^2\right)=\frac{1}{5}\times 10=2 \xrightarrow{\text{جذر}} \sigma=\sqrt{2}$$

مثال: انحراف از میانگین داده هایی برابر ۱ ، ۱- و ۳ ، ۲ ، a است. واریانس داده ها را بدست آورید.

$\text{مجموع انحراف از میانگین}=0 \Rightarrow a+2+3+(-1)+1=0 \Rightarrow a=-5$

باتوجه به این که در فرمول محاسبهٔ واریانس، عبارت صورت، همان مجموع مربعات انحراف از میانگین است

واریانس را محاسبه می کنیم: $\sigma^2=\frac{(-5)^2+2^2+3^2+(-1)^2+1^2}{5}=\frac{40}{5}=8$

نکته تستی: اگر n داده با هم تشکیل دنبالهٔ حسابی با قدر نسبت d بدهند، آنگاه $\sigma^2=\frac{n^2-1}{12}d^2$ است.

مثال: انحراف معیار داده های ۱۴ ، ۱۱ ، ۸ ، ۵ ، ۲ را محاسبه نمایید.

۵ داده، تشکیل دنبالهٔ حسابی با قدر نسبت $d=3$ داده اند بنابراین: $\sigma^2=\frac{25-1}{12}\times 3^2=2\times 9 \Rightarrow \sigma=3\sqrt{2}$

نکته مهم:

درصورتی که عدد ثابتی به همه داده ها اضافه کنیم (یا کم کنیم) واریانس تغییری نمی کند.

اما اگر تمام داده ها را در عدد ثابت ضرب کنیم (یا تقسیم کنیم) واریانس در مربع آن عدد ضرب (یا تقسیم) خواهد شد.

به عبارت دیگر، اگر واریانس داده های x_1 و ... و x_n برابر σ_x^2 باشد، آنگاه:

واریانس داده های x_1+b و ... و x_n+b برابر σ_x^2 است و می نویسیم: $\sigma_{x+b}^2 = \sigma_x^2$

همچنین واریانس داده های ax_1 و ... و ax_n برابر $a^2\sigma_x^2$ است و می نویسیم: $\sigma_{ax}^2 = a^2\sigma_x^2$

بنابراین: $\sigma_{ax+b}^2 = a^2\sigma_x^2$

پس اگر داده ها بزرگ باشند بهتر است مقدار ثابتی را از همه کم کرده سپس واریانس را محاسبه نمود.

مثال: انحراف معیار داده های ۱۰۱ و ۱۰۲ و ۹۹ و ۹۸ را بدست آورید.

برای سهولت در محاسبه. ابتدا تمام داده ها را منهای ۱۰۰ می کنیم بنابراین داده ها ۱ و ۲ و ۱- و ۲- خواهند شد:

$\bar{x} = \frac{(-۲)+(-۱)+۲+۱}{۴} = ۰ \Rightarrow \sigma^2 = \frac{(-۲)^2+(-۱)^2+(۲)^2+(۱)^2}{۴} = \frac{۱۰}{۴} \xRightarrow{\text{جذر}}$ انحراف معیار: $\sigma = \frac{\sqrt{۱۰}}{۲}$

مثال: به ازای چه مقدار از a، داده های $a+۹$, $a+۳$, $a+۱$ و $a-۱$ دارای میانگین و واریانس یکسان هستند؟

$\bar{x} = \frac{(a-۱)+(a+۱)+(a+۳)+(a+۹)}{۴} = a+۳$

برای محاسبهٔ واریانس مقدار a واحد از همه داده ها کم می کنیم بنابراین داده ها ۹ و ۳ و ۱ و ۱- خواهند شد:

$\bar{x} = ۳ \Rightarrow \sigma^2 = \frac{(-۴)^2+(-۲)^2+۰^2+۶^2}{۴} = ۱۴ \Rightarrow a+۳ = ۱۴ \Rightarrow a = ۱۱$

سؤال: اگر تمام داده ها را با عدد ثابت b جمع کنیم یا تمام داده ها را در عدد ثابت a ضرب کنیم، انحراف معیار چه تغییری می کند؟

با توجه به این که انحراف معیار جذر واریانس است داریم:

$\sigma_{x+b} = \sigma_x$ و $\sigma_{ax} = |a|\sigma_x \Rightarrow \sigma_{ax+b} = |a|\sigma_x$

توجه: اگر واریانس (انحراف معیار) مجموعه داده ها عدد کوچکی باشد، بدین معناست که پراکندگی داده ها حول میانگینشان کم و درنتیجه داده ها به هم نزدیکتر است و در صورتی که واریانس عددی بزرگ باشد، پراکندگی داده ها حول میانگینشان زیاد است و درنتیجه داده ها از هم دورتر است.

تست (ریاضی ۹۳): نمرات آزمون مهارت فنی دو کارگر A و B به صورت مقابل است. دقت عمل کدام بیشتر بوده است؟

A : ۱۵ ، ۱۴ ، ۱۵ ، ۱۶ ، ۱۷ ، ۱۹

B : ۱۶ ، ۱۴ ، ۱۷ ، ۱۴ ، ۱۷ ، ۱۸

۱) A ۲) B ۳) یکسان ۴) غیرقابل پیش بینی

$$A \xrightarrow{-15} 0 , -1 , 0 , 1 , 2 , 4 \Rightarrow \bar{x} = 1 \Rightarrow \sigma^2 = \frac{1+4+1+0+1+9}{6} = \frac{16}{6}$$

$$B \xrightarrow{-14} 2 , 0 , 3 , 0 , 3 , 4 \Rightarrow \bar{x} = 2 \Rightarrow \sigma^2 = \frac{0+4+1+4+1+4}{6} = \frac{14}{6}$$

واریانس B کمتر است، پس پراکندگی داده های B کمتر است درنتیجه دقت عمل آن بیشتر است ⟵ گ ۲

نکته مهم: برای محاسبهٔ واریانس داده های $x_1 , \dots , x_n$ می توان از رابطهٔ زیر نیز کمک گرفت:

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

مثال: مجموع ۱۸ دادهٔ آماری ۵۴ و مجموع مجذورات آنها ۱۸۰ می باشد، انحراف معیار آنرا حساب کنید.

$$x_1 + \dots + x_{18} = 54 \Rightarrow \bar{x} = \frac{54}{18} = 3 \quad , \quad x_1^2 + \dots + x_{18}^2 = 180$$

$$\Rightarrow \sigma^2 = \frac{180}{18} - 3^2 = 10 - 9 = 1 \xRightarrow{\text{جذر}} \sigma = 1$$

تست: مجموع و انحراف معیار ده دادهٔ آماری به ترتیب ۴۰ و $\sqrt{2}$ است. مجموع مربعات این ده داده کدام است؟

۱) ۱۶۰ ۲) ۱۸۰ ۳) ۱۹۰ ۴) ۲۰۰

$$x_1 + \dots + x_{10} = 40 \Rightarrow \bar{x} = \frac{40}{10} = 4 \quad , \quad \sigma = \sqrt{2} \Rightarrow \sigma^2 = 2$$

می دانیم: $\sigma^2 = \frac{x_1^2 + \dots + x_{10}^2}{10} - (\bar{x})^2 \Rightarrow 2 = \frac{x_1^2 + \dots + x_{10}^2}{10} - 16 \Rightarrow x_1^2 + \dots + x_{10}^2 = 180$ ⟵ گ ۲

ج) ضریب تغییرات داده ها :

معیاری است که از تقسیم انحراف معیار داده ها (σ) به میانگین داده ها ($\bar{x}$) به دست می آید و آن را با نماد CV نشان می دهند : $CV=\frac{\sigma}{\bar{x}}$

به عنوان نمونه اگر بخواهیم پراکندگی نمرات دو کلاس، در درس ریاضی را با هم مقایسه کنیم، واریانس داده ها را حساب می کنیم، هر کدام از کلاس ها که واریانس کمتری داشته باشد، پراکندگی کمتری دارد ولی در صورت برابر بودن واریانس آنها، از ضریب تغییرات استفاده می کنیم.

همچنین اگر بخواهیم پراکندگی دو دسته از داده ها را که هم واحد نیستند با هم مقایسه کنیم، از ضریب تغییرات استفاده می کنیم (زیرا ضریب تغییرات به واحد بستگی ندارد) هرچه ضریب تغییرات داده ها کمتر باشد، پراکندگی داده ها کمتر است.

مثال : ضریب تغییرات داده های ۱۴ و ۱۴ و ۱۳ و ۱۲ و ۱۲ و ۱۲ و ۱۱ و ۱۱ و ۹ را محاسبه نمایید. (انسانی ۸۹)

$$\bar{x}=\frac{\text{مجموع داده ها}}{\text{تعداد داده ها}}=\frac{96}{8}=12 \rightsquigarrow \sigma^2=\frac{(-3)^2+(-1)^2+(-1)^2+0+0+(1)^2+(2)^2+(2)^2}{8}=\frac{10}{4}\Rightarrow\sigma=\frac{\sqrt{10}}{2}$$

$$\Rightarrow CV=\frac{\sigma}{\bar{x}}=\frac{\frac{\sqrt{10}}{2}}{12}=\frac{\sqrt{10}}{24}\simeq 0/13$$

تست (ریاضی ۹۳) : در ۱۲ داده آماری مجموع تمام داده ها ۷۲ و مجموع مجذورات آنها ۴۸۰ می باشد ضریب تغییرات این داده ها کدام است؟

$\frac{1}{4}$ $\frac{2}{9}$ $\frac{1}{3}$ $\frac{2}{5}$

$$\bar{x}=\frac{72}{12}=6 \quad \text{و} \quad \sigma^2=\frac{\sum x_i^2}{n}-(\bar{x})^2=\frac{480}{12}-(6)^2=4\Rightarrow\sigma=2\Rightarrow CV=\frac{2}{6}=\frac{1}{3}$$

تست (ریاضی ۹۲) : اگر میانگین و ضریب تغییرات اندازهٔ اضلاع مربع هایی ۱۵ و ۰/۲ باشد، میانگین مساحت این مربع ها کدام است؟

۲۲۷ ۲۲۹ ۲۳۲ ۲۳۴

ضلع مربع ها را برابر x_i در نظر می گیریم، پس مساحت x_i^2 می باشد. میانگین اضلاع ها $\bar{x}$ و میانگین مساحت ها همان $\frac{\sum x_i^2}{n}$ است. بنابراین :

$$CV=\frac{\sigma}{\bar{x}}\xrightarrow{\bar{x}=15\ ,\ CV=0/2}0/2=\frac{\sigma}{15}\Rightarrow\sigma=3$$

$$\sigma^2=\frac{\sum x_i^2}{n}-(\bar{x})^2\Rightarrow 9=\text{میانگین مساحت ها}-(15)^2\Rightarrow\text{میانگین مساحت ها}=9+225=234$$

مثال : در ۱۰۰ داده آماری با میانگین ۱۸ و انحراف معیار ۲ ، تمام داده ها را در ۴ ضرب و ۸ واحد به آن ها اضافه می کنیم. ضریب تغییرات جدید را به دست آورید.

میانگین جدید $= 4 \times$ (میانگین قدیم) $+ 8 = 4 \times 18 + 8 = 80$

انحراف معیار جدید $= |4| \times$ (انحراف معیار قدیم) $= 4 \times 2 = 8$

$\Rightarrow$ CV جدید $= \frac{8}{80} = 0/1$

تست (انسانی خارج ۹۴) : ضریب تغییرات داده های آماری ۱/۳۵ می باشد. به دو برابر این داده های آماری، $\frac{1}{4}$ میانگین آنها افزوده شده است. ضریب تغییرات داده های جدید کدام است؟ ۰/۹۶ ۱/۰۸ ۱/۱۵ ۱/۲

$\bar{x}_{\text{جدید}} = 2\bar{x} + \frac{1}{4}\bar{x} = \frac{9}{4}\bar{x}$ و $\sigma_{\text{جدید}} = 2\sigma$ و $CV_{\text{قدیم}} = \frac{\sigma}{\bar{x}} = 1/35$

$$\Longrightarrow CV_{\text{جدید}} = \frac{\sigma_{\text{جدید}}}{\bar{x}_{\text{جدید}}} = \frac{2\sigma}{\frac{9}{4}\bar{x}} = \frac{8}{9} \cdot \frac{\sigma}{\bar{x}} = \frac{8}{9} \times 1/35 = 1/2$$

نکته : اگر هر کدام از شاخص های پراکندگی (انحراف معیار، واریانس و ضریب تغییرات) برابر صفر باشد، همهٔ داده ها با هم مساوی اند و برعکس.

مثال : در صورتی که انحراف معیار داده های ۳ ، $c-2$ ، $b+1$ و $a+1$ برابر صفر باشد، ضریب تغییرات داده های a ، b و c را به دست آورید.

داده ها باهم مساوی اند $\Rightarrow a+1 = b+1 = c-2 = 3 \Rightarrow a = b = 2 \ , \ c = 5$

حال باید ضریب تغییرات داده های ۲ و ۲ و ۵ را حساب کرد:

$\bar{x} = \frac{2+2+5}{3} = 3$ و $\sigma^2 = \frac{2^2 + (-1)^2 + (-1)^2}{3} = 2 \Rightarrow \sigma = \sqrt{2} \Rightarrow CV = \frac{\sqrt{2}}{3}$

تست (ریاضی خارج ۹۱) : واریانس ۱۱ داده آماری برابر صفر است. اگر داده های ۲۴ و ۱۶ و ۲۶ به آن ها اضافه شود میانگین داده ها تغییری نمی کند. انحراف معیار ۱۴ داده حاصل کدام است؟ ۰/۷۵ ۱/۲۵ ۱/۵ ۲

داده ها باهم مساوی و به صورت a ، a ، ... ، a در نظر می گیریم که میانگین آنها $\bar{x} = a$ خواهد بود.

بنابراین داده های جدید $26, 24, 16, \underbrace{a, a, \dots, a}_{11\ \text{تا}}$ بوده که طبق فرض سؤال، میانگین آنها با میانگین داده های قبل برابر است:

$$\frac{11a+16+24+26}{14}=a \Rightarrow a=22$$

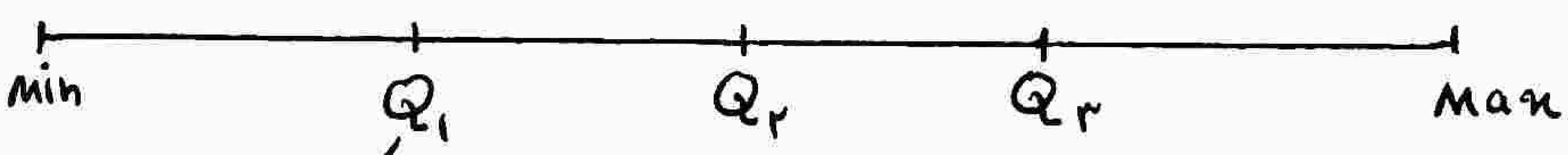
$\Rightarrow$ داده های جدید : ۲۲ و ... و ۲۲ ، ۱۶ ، ۲۴ ، ۲۶

$\Rightarrow \bar{x}=22$ ، $\sigma^2=\frac{0+0+\dots+0+4+36+16}{14}=4 \Rightarrow \sigma=2$

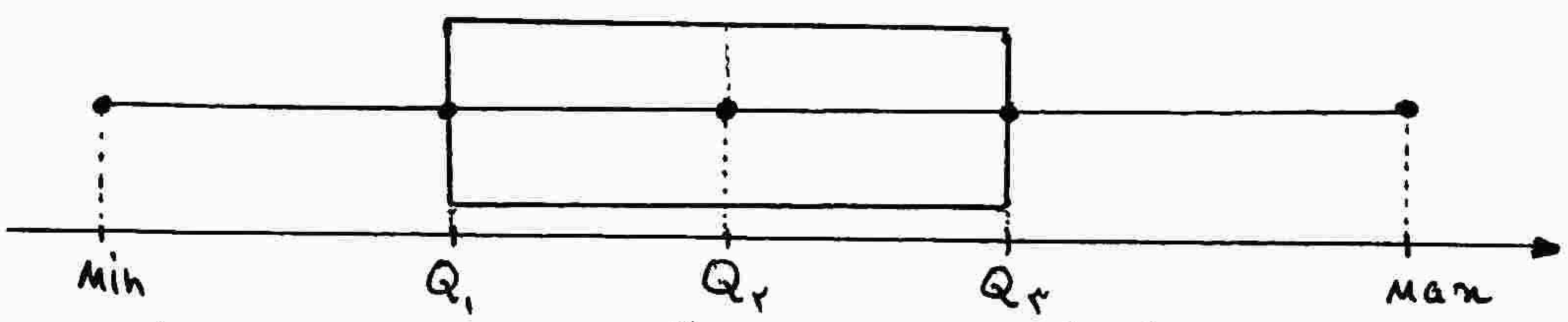
نمودار جعبه ای (نمایش معیارهای پراکندگی داده ها به صورت تصویری) :

نمودار جعبه ای نموداری تصویری است که داده ها را براساس پنج مقدار نمایش می دهد. این مقادیر عبارتند از : کوچکترین داده (Min) ، چارک اول (Q_1) ، میانه (Q_2) ، چارک سوم (Q_3) و بزرگ ترین داده (Max)

ابتدا مقادیر فوق را روی یک محور نمایش می دهیم :



حال با رسم یک جعبه از Q_1 تا Q_3 به عرض دلخواه، حدود دامنه ی میان چارکی ($IQR=Q_3-Q_1$) را مشخص می کنیم، سپس با استفاده از یک خط، میانه را در جعبه مشخص می کنیم و در انتها، از دو طرف جعبه به کمترین و بیشترین مقدار داده ها دو خط رسم می کنیم (شکل زیر) :

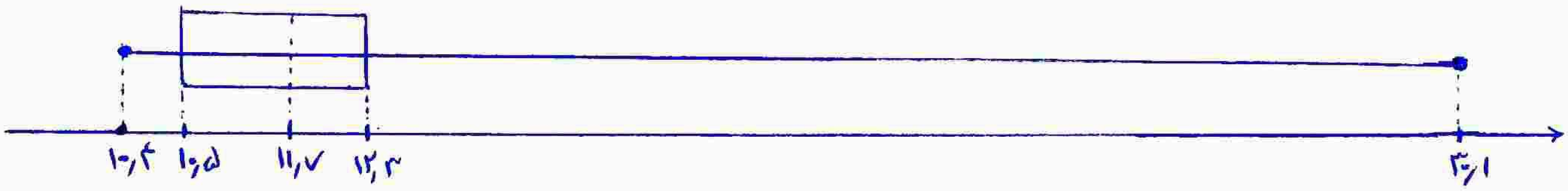


مثال : داده های زیر، مربوط به نرخ بیکاری یک کشور در ده سال گذشته است. نمودار جعبه ای آن را رسم کنید.

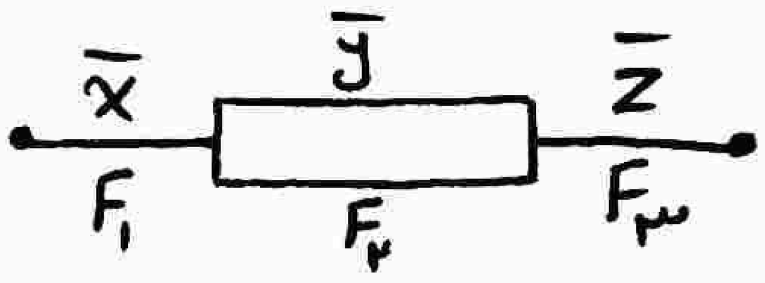
۱۱/۵ ، ۱۱/۳ ، ۱۰/۵ ، ۱۰/۴ ، ۱۱/۹ ، ۱۳/۵ ، ۱۲/۳ ، ۱۲/۲ ، ۱۰/۴ ، ۳۰/۱

داده های مرتب شده : ۱۰/۴ ، ۱۰/۴ ، ۱۰/۵ ، ۱۱/۳ ، ۱۱/۵ ، ۱۱/۹ ، ۱۲/۲ ، ۱۲/۳ ، ۱۳/۵ ، ۳۰/۱

Min = ۱۰/۴ ، Q_1 = ۱۰/۵ ، Q_2 = ۱۱/۷ ، Q_3 = ۱۲/۳ ، Max = ۳۰/۱



نکته: اگر فراوانی نسبی داده ها در طرف چپ جعبه، درون و روی جعبه و طرف راست جعبه به ترتیب F_1 و F_2 و F_3 بوده و میانگین داده های آن ها به ترتیب $\bar{x}$ و $\bar{y}$ و $\bar{z}$ باشد، آنگاه میانگین کل داده ها برابر است با:



میانگین کل $= F_1\bar{x} + F_2\bar{y} + F_3\bar{z}$

تست: در نمودار جعبه ای ۲۳ داده آماری، میانگین داده های دنباله های سمت چپ و سمت راست جعبه به ترتیب ۲۱٫۶ و ۳۳ و میانگین داده های داخل و روی جعبه ۲۵ می باشد. میانگین کل این داده کدام است؟ (تجربی ۹۵ خارج)

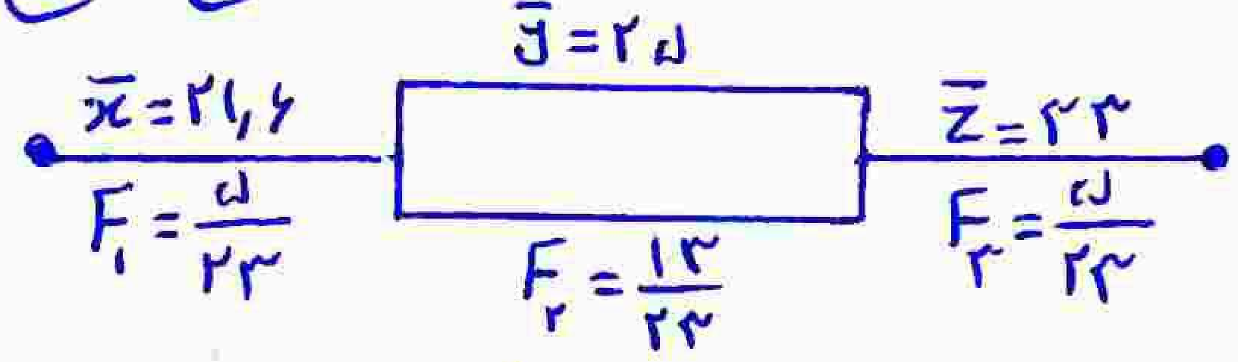
۱) ۲۵٫۸ ۲) ۲۶ ۳) ۲۶٫۱ ۴) ۲۶٫۲

۲۳ داده وجود دارد، پس داده $12 = \frac{23+1}{2}$ اُم میانه است.

۱۱ داده قبل میانه است، لذا داده $6 = \frac{11+1}{2}$ اُم چارک اول است، یعنی ۵ داده در دنباله سمت چپ وجود دارد.

همچنین ۵ داده در دنباله سمت راست جعبه بوده و در نتیجه $13 = 5 - 5 - 23$ داده داخل و روی جعبه واقع است.

لذا می توان نمودار جعبه ای را برحسب میانگین و فراوانی نسبی رسم کرد:

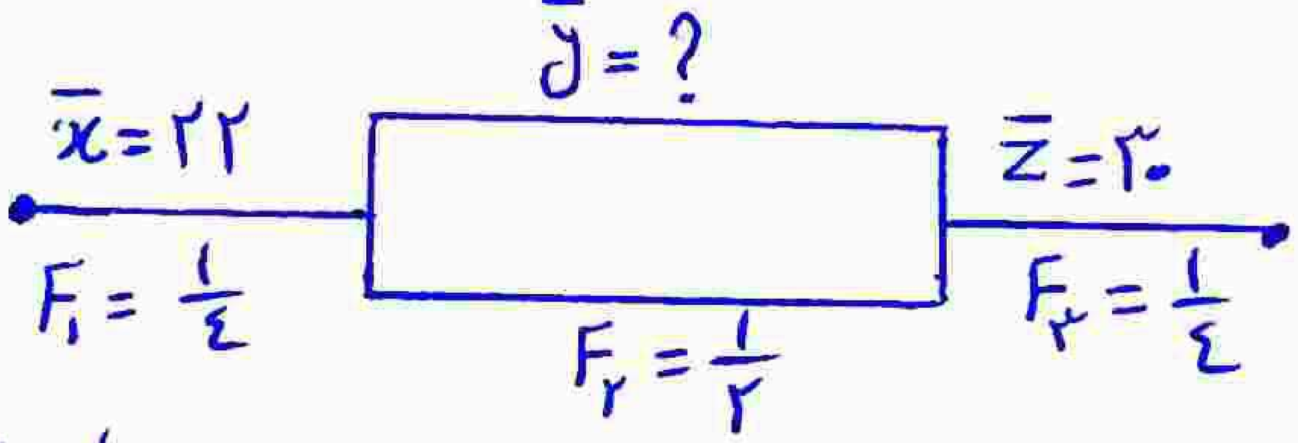


$\Rightarrow$ میانگین کل $= \frac{5}{23}\times 21.6 + \frac{13}{23}\times 25 + \frac{5}{23}\times 33 = 26 \Longrightarrow$ گزینه ۲

تست: در نمودار جعبه ای ۳۶ داده آماری میانگین داده های دو طرف جعبه جداگانه به ترتیب ۲۲ و ۳۰ می باشد. اگر میانگین تمام داده ها ۲۷٫۵ باشد میانگین داده های داخل جعبه کدام است؟ (ریاضی ۹۰)

۱) ۲۸ ۲) ۲۸٫۵ ۳) ۲۹ ۴) ۲۹٫۵

توجه: اگر تعداد داده ها مضرب ۴ باشد، دقیقاً $\frac{1}{2}$ داده ها داخل جعبه و در هر یک از دو طرف جعبه $\frac{1}{4}$ داده ها قرار دارند و می دانیم ۳۶ مضرب ۴ است، بنابراین طبق شرایط تست داریم:



میانگین کل: $27.5 = \frac{1}{4}\times 22 + \frac{1}{2}\bar{y} + \frac{1}{4}\times 30 \Longrightarrow \bar{y} = 29 \longrightarrow$ گزینه ۳

حل چند نمونه سؤال : (مخصوص علاقه‌مندان)

۱- در تعدادی داده آماری به بزرگترین داده $2x+1$ واحد اضافه و از کوچکترین داده $x+3$ واحد کم می‌کنیم. اگر دامنه داده ۲۲ واحد اضافه شود، مقدار x را تعیین کنید. $(x>0)$

داده های قدیم : کوچکترین داده $=\alpha$ و بزرگترین داده $=\beta \Rightarrow R=\beta-\alpha$

داده های جدید : کوچکترین داده $=\alpha-(x+3)$ و بزرگترین داده $=\beta+(2x+1) \Rightarrow R'=\beta+(2x+1)-\alpha+(x+3)$

$$\xrightarrow{R'=R+22} \beta+2x+1-\alpha+x+3=\beta-\alpha+22 \Rightarrow 3x=18 \Rightarrow x=6$$

۲- واریانس داده های آماری دسته بندی شده در جدول مقابل را بیابید.

مرکز دسته	1	3	5	7	9
فراوانی	2	7	3	5	3

$$\bar{x}_w=\frac{1\times2+3\times7+5\times3+7\times5+9\times3}{2+7+3+5+3}=5$$

$$\sigma^2=\frac{2(1-5)^2+7(3-5)^2+3(5-5)^2+5(7-5)^2+3(9-5)^2}{2+7+3+5+3}=6.4$$

توجه : می دانیم فراوانی هر داده، تعداد تکرار هاِ داده است، لذا در محاسبهٔ واریانس این تعداد را ضرایب هرکدام از پرانتزهای مربوطه قرار داده‌ایم، درضمن تعداد داده ها: مجموع فراوانی هاست.

در حالت کلی اگر فراوانی داده های x_i برابر w_i باشد، آنگاه : $\sigma^2=\dfrac{w_1(x_1-\bar{x})^2+\cdots+w_n(x_n-\bar{x})^2}{w_1+\cdots+w_n}$

حال با کمی تأمل می توان نتیجه گرفت :

اگر F_i فراوانی نسبی دادهٔ x_i باشد آنگاه : $\sigma^2=F_1(x_1-\bar{x})^2+\cdots+F_n(x_n-\bar{x})^2$

۳- جدول روبرو فراوانی نسبی داده های دسته بندی شده است. واریانس آنها را محاسبه نمایید. (سراسری ریاضی خارج ۹۳)

مرکز دسته	8	12	16	20
فراوانی نسبی	0.1	0.25	0.2	α

مجموع فراوانی های نسبی برابر ۱ است $\Rightarrow 0.1+0.25+0.2+\alpha=1 \Rightarrow \alpha=0.45$

$$\bar{x}=0.1\times8+0.25\times12+0.2\times16+0.45\times20=16$$

$$\Rightarrow \sigma^2=0.1(8-16)^2+0.25(12-16)^2+0.2(16-16)^2+0.45(20-16)^2=17.6$$

۴- اطلاعات زیر مربوط به اجرای یک آزمون هوش در دو کلاس A و B است. انحراف معیار کل دو کلاس (انحراف معیار مرکب) را حساب کنید.

کلاس	تعداد دانش آموزان	میانگین	انحراف معیار
A	۲۵	۱۰۳	۹
B	۲۴	۱۰۴	۱۲

کلاس A: $\bar{x} = \frac{\sum x_i}{25} \Rightarrow \sum x_i = 25 \times 103 = 2575$

$$\sigma^2 = \frac{\sum x_i^2}{25} - (\bar{x})^2 \Rightarrow 9^2 = \frac{\sum x_i^2}{25} - (103)^2 \Rightarrow \sum x_i^2 = 267250$$

کلاس B: $\bar{y} = \frac{\sum y_i}{24} \Rightarrow \sum y_i = 24 \times 104 = 2496$

$$\sigma^2 = \frac{\sum y_i^2}{24} - (\bar{y})^2 \Rightarrow 12^2 = \frac{\sum y_i^2}{24} - (104)^2 \Rightarrow \sum y_i^2 = 263040$$

نمرات دانش آموزان کل دو کلاس را با Z_i نمایش می‌دهیم:

$$\bar{Z} = \frac{\sum Z_i}{25+24} = \frac{\sum x_i + \sum y_i}{49} = \frac{2575 + 2496}{49} \simeq 103/5$$

$$\sigma^2 = \frac{\sum Z_i^2}{49} - (\bar{Z})^2 = \frac{\sum x_i^2 + \sum y_i^2}{49} - (\bar{Z})^2 = \frac{267250 + 263040}{49} - (103/5)^2 = 110$$

$$\xRightarrow{\text{جذر}} \sigma_{\text{کل}} = \sqrt{110} \simeq 10/48$$

۵- اگر داده‌های آماری ۱۱و۱۵و۱۷، ۱۶ و۱۴و ۹ و ۱۱ و ۱۲، ۱۵، ۱۸ و ۱۴ را با نمودار جعبه‌ای نشان دهیم، انحراف معیار داده‌های داخل جعبه را حساب کنید. (سراسری تجربی ۸۸)

داده‌های مرتب شده: $9 ، 11 ، \underset{Q_1}{11} ، 12 ، 12 ، 14 ، \underset{Q_2}{14} ، 15 ، 15 ، \underset{Q_3}{16} ، 17 ، 18$

$\Rightarrow$ داده‌های درون جعبه: ۱۲، ۱۴، ۱۴، ۱۵ و ۱۵ $\Rightarrow \bar{x} = \frac{70}{5} = 14$

$$\Rightarrow \sigma^2 = \frac{(12-14)^2 + 2(14-14)^2 + 2(15-14)^2}{5} = \frac{8}{5} = 1/2 \xRightarrow{\text{جذر}} \sigma = \sqrt{1/2} \simeq 1/1$$