



فصل سوم : آمار توصیفی

درس ۱ : توصیف و نمایش داده‌ها

۱) مفاهیم پایه

در یک مسئله آماری، مفاهیم و اصطلاحاتی استفاده می‌شود که دانستن آن‌ها ضروری است. از جمله این مفاهیم عبارتند از:

داده : واقعیت‌هایی درباره یک شیء یا فردند که در محاسبه، برنامه‌ریزی و پیش‌بینی به کار می‌روند.

متغیر : یک ویژگی است که در اعضای جامعه یکسان نیست و معمولاً از یک عضو به عضو دیگر فرق می‌کند.

فراوانی : تعداد دفعاتی که هر داده (x_i) مشاهده می‌شود را فراوانی آن داده گفته و با f_i نشان می‌دهیم. مجموع تمام فراوانی‌ها را با n نشان می‌دهیم. بنابراین :

داده : ۲, ۲, ۴, ۶, ۴, ۲

$$n = f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i$$

جدول فراوانی

x_i	۲	۴	۶
f_i	۳	۲	۱

فراوانی نسبی : نسبت فراوانی یک داده به تعداد کل داده‌ها را فراوانی نسبی آن داده می‌گوییم که برابر با $f'_i = \frac{f_i}{n}$ است. مجموع تمام

فراوانی‌های نسبی برابر ۱ می‌باشد.

x_i	۲	۴	۶
f'_i	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

درصد فراوانی نسبی : اگر فراوانی نسبی یعنی f'_i را در ۱۰۰ ضرب کنیم، درصد فراوانی نسبی حاصل می‌شود. مجموع تمام درصد فراوانی‌های نسبی برابر ۱۰۰ می‌باشد.

مثال ۱) جدول زیر فراوانی داده‌ها را نشان می‌دهد. اگر فراوانی نسبی داده سوم، ۰/۱۵ باشد، x را به دست آورده و مشخص کنید که چند درصد داده‌ها بزرگ‌تر یا مساوی ۹ هستند؟

داده سوم ۵

داده (x_i)	۳	۵	۹	۱۰
فراوانی (f_i)	۵	۸	x	۴

$$f'_3 = 0.15 \Rightarrow \frac{f_3}{n} = 0.15 \Rightarrow \frac{x}{5+8+x+4} = 0.15$$

$$\Rightarrow \frac{x}{x+17} = \frac{15}{100} = \frac{3}{20} \Rightarrow 20x = 3x + 51 \Rightarrow 17x = 51 \Rightarrow x = \frac{51}{17} = 3$$

$$\%35 \rightarrow \frac{3}{7} \times 100 = \frac{3}{7} \times 100 = 35\% \Rightarrow \text{تعداد داده‌ها بزرگ‌تر یا مساوی ۹} = x + 4 = 3 + 4 = 7$$

مثال ۲) قد دانش‌آموزان یک کلاس را اندازه‌گیری کرده‌ایم. با توجه به جدول زیر، چند درصد این دانش‌آموزان قد بیشتر از ۱۵۸ سانتی‌متر دارند؟

قد (سانتی‌متر)	۱۴۵	۱۵۰	۱۵۵	۱۶۰	۱۶۵
فراوانی	۷	۱۱	۱۰	۴	۳

(تعداد افرادی که قد بیشتر از ۱۵۸ دارند) $f_4 + f_5 = 4 + 3 = 7$

$$\Rightarrow \frac{7}{35} \times 100 = 20 \rightarrow 20\%$$

$$n = \sum f_i = 7 + 11 + 10 + 4 + 3 = 35$$





(۷) نمایش داده‌ها

نمایش داده‌ها به کمک نمودار، در ارائه و توصیف آنها اهمیت زیادی دارد. فراوانی و فراوانی نسبی داده‌ها را می‌توانیم به کمک نمودار نشان دهیم. سه نمودار میله‌ای، بافت نگاشت و دایره‌ای را برای نمایش داده‌ها استفاده می‌کنیم.

نمودار میله‌ای :

در این نمودار از میله‌هایی استفاده می‌شود که قاعده این میله‌ها بر داده‌ها منطبق بوده و ارتفاع هر میله به اندازه فراوانی یا فراوانی نسبی آن داده است. این نمودار معمولاً برای داده‌های کمی گسسته و یا کیفی استفاده می‌شود.

نمودار بافت نگاشت :

برای نمایش داده‌های کمی پیوسته یا گسسته که دسته‌بندی شده‌اند، از نمودار بافت نگاشت (مستطیلی) استفاده می‌شود. عرض میله‌ها به اندازه طول دسته و ارتفاع میله به اندازه فراوانی یا فراوانی نسبی آن دسته می‌باشد.

نمودار دایره‌ای :

برای نمایش داده‌ها به کمک این نمودار، مساحت دایره را به نسبت فراوانی هر داده تقسیم می‌کنند. برای انجام این کار، به داده

x_i ، یک قطاع با زاویه مرکزی $\alpha_i = \frac{f_i}{n} \times 360$ اختصاص می‌دهند. این نمودار نیز معمولاً برای داده‌های کمی گسسته و کیفی استفاده می‌شود.



$$\alpha_i = \frac{f_i}{n} \times 360$$

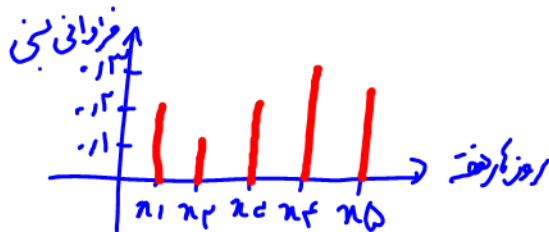
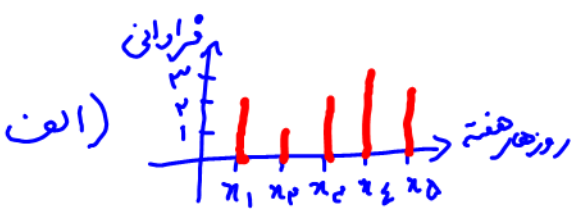
مثال (۳) تعداد غایبین یک کلاس در یک هفته در جدول زیر داده شده است.

الف) نمودار میله‌ای فراوانی و فراوانی نسبی این داده‌ها را رسم کنید.

ب) نمودار دایره‌ای این داده‌ها را رسم کنید.

x_i	x_1	x_2	x_3	x_4	x_5
روز	شنبه	یکشنبه	دوشنبه	سه‌شنبه	چهارشنبه
تعداد غایبین	۲	۱	۲	۳	۲
n	۱۰				
فراوانی نسبی	$\frac{2}{10} = 0.2$	$\frac{1}{10} = 0.1$	$\frac{2}{10} = 0.2$	$\frac{3}{10} = 0.3$	$\frac{2}{10} = 0.2$

$$n = 2 + 1 + 2 + 3 + 2 = 10$$



ب)

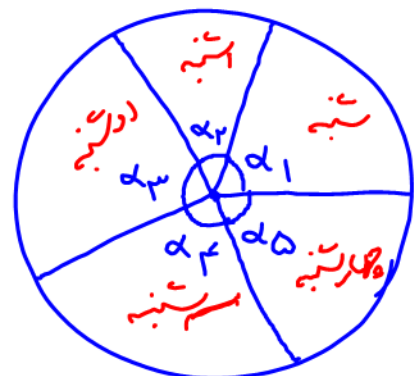
$$\alpha_1 = \frac{f_1}{n} \times 360 = \frac{2}{10} \times 360 = 72^\circ$$

$$\alpha_2 = \frac{f_2}{n} \times 360 = \frac{1}{10} \times 360 = 36^\circ$$

$$\alpha_3 = \alpha_1 = 72^\circ$$

$$\alpha_4 = \frac{f_4}{n} \times 360 = \frac{3}{10} \times 360 = 108^\circ$$

$$\alpha_5 = \alpha_1 = 72^\circ$$

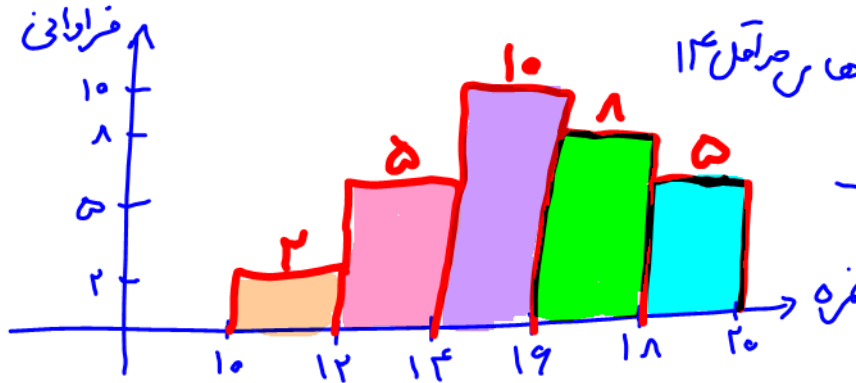




مثال ۴) توزیع نمره دانش‌آموزان یک کلاس در درس فیزیک در جدول زیر دسته‌بندی شده است. برای این نمره‌ها، نمودار بافت نگاشت فراوانی را رسم کنید. چند درصد نمره‌های این کلاس، حداقل ۱۴ است؟

نمره	[۱۰, ۱۲)	[۱۲, ۱۴)	[۱۴, ۱۶)	[۱۶, ۱۸)	[۱۸, ۲۰]
فراوانی	۲	۵	۱۰	۸	۵

$$\rightarrow n = 2 + 5 + 10 + 8 + 5 = 30$$



$$\text{تعداد نمره‌ها بر مبنای ۱۴} = 10 + 8 + 5 = 23$$

$$\rightarrow \frac{23}{30} \times 100 = \frac{230}{3} = 76\%$$

$$\begin{array}{r} 230 \\ 3 \overline{) 76} \\ \underline{69} \\ 70 \\ \underline{69} \\ 10 \end{array}$$

مثال ۵) جدول روبه‌رو فراوانی تعداد شهرهای یک استان را بر حسب میزان بارندگی آنها (بر حسب میلی‌متر) در طول یک سال نشان می‌دهد.

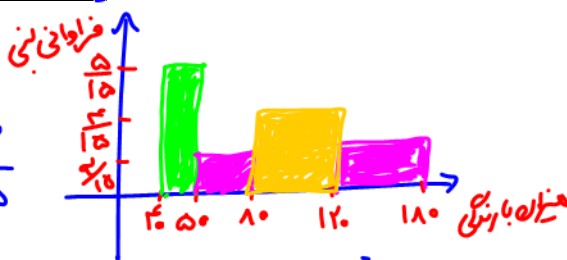
میزان بارندگی	فراوانی
۴۰ - ۵۰	۵ = f_1
۵۱ - ۸۰	۳ = f_2
۸۱ - ۱۲۰	۴ = f_3
۱۲۱ - ۱۸۰	۳ = f_4

الف) نمودار بافت نگاشت فراوانی نسبی را برای این جدول رسم کنید.

ب) چند درصد از شهرهای این استان حداکثر ۸۰ میلی‌متر در سال بیش از بارندگی دارند؟

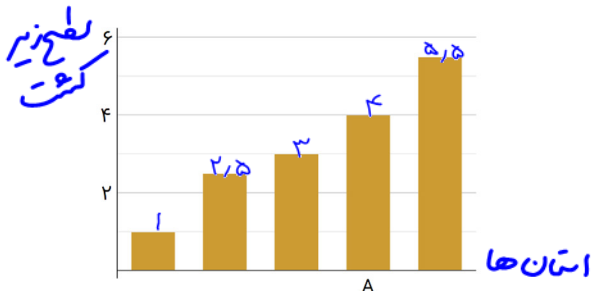
$$n = \sum f_i = 5 + 3 + 4 + 3 = 15$$

$$\begin{aligned} f'_1 &= \frac{5}{15} \\ f'_2 &= f'_3 = \frac{3}{15} \\ f'_4 &= \frac{4}{15} \end{aligned}$$



$$\begin{aligned} &\frac{f_1 + f_2}{n} \times 100 \\ &= \frac{5 + 3}{15} \times 100 = \frac{8}{15} \times 100 = \frac{800}{15} = 53\% \end{aligned}$$

مثال ۶) نمودار میله‌ای زیر، سطح زیر کشت غله را در بخش استان نشان می‌دهد. در نمودار دایره‌ای، زاویه مرکزی متناظر با استان A چند درجه است؟



$$n = \sum f_i = 1 + 2.5 + 3 + 4 + 5.5 = 16$$

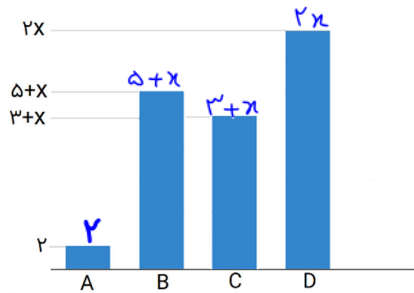
$$\alpha_A = \frac{f_A}{n} \times 360^\circ = \frac{1}{16} \times 360^\circ = \frac{1}{4} \times 360^\circ = 90^\circ$$





مثال ۷ نمودار میله‌ای برای موجودی محصول‌های A، B، C و D در یک فروشگاه به صورت زیر است.

اگر فراوانی نسبی محصول B در این فروشگاه ۳۰٪ باشد، چند محصول B در این فروشگاه وجود دارد؟



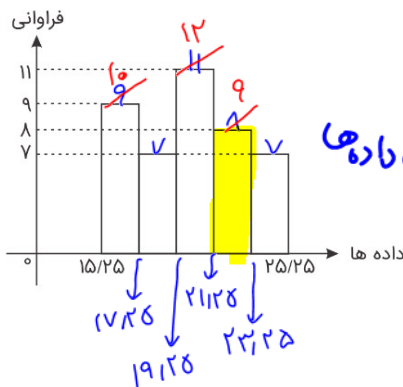
$$n = \sum f_i = 2 + (5+x) + (3+x) + 2x = 10 + 4x$$

$$f'_B = 0.30 \Rightarrow \frac{f_B}{n} = 0.30 \Rightarrow \frac{5+x}{10+4x} = \frac{3}{10} \Rightarrow$$

$$3(10+4x) = 10(5+x) \Rightarrow 30+12x = 50+10x \Rightarrow 2x = 20$$

$$\Rightarrow x = 10 \Rightarrow f_B = 5+x = 5+10 = 15$$

مثال ۸ در نمودار بافت نگاشت زیر، به داده‌های آماری مورد نظر، سه داده ۱۷، ۲۱ و ۲۲ اضافه می‌شود. درصد فراوانی نسبی دسته چهارم چقدر خواهد شد؟



$$\frac{15.25 - 10.25}{5} = \frac{10}{5} = 2$$

$$n = \sum f_i = 10 + 7 + 12 + 9 + 7 = 45$$

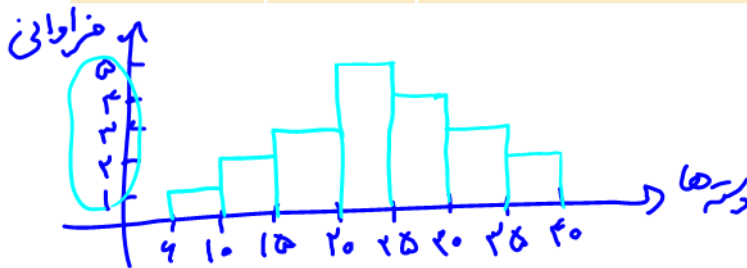
$$f'_4 = \frac{f_4}{n} = \frac{9}{45} = \frac{1}{5} \xrightarrow{\times 100} \frac{1}{5} \times 100 = 20\%$$



۱ داده‌های زیر، مسافتی را که ۲۰ راننده از مکان‌های مختلف برای رسیدن به مقصد A طی می‌کنند نشان می‌دهد. این داده‌ها، در جدول زیر

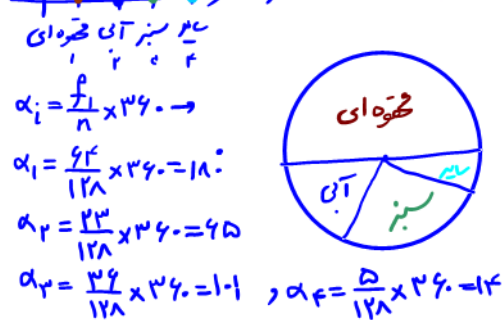
گردآوری شده است. جدول را کامل کرده و نمودار بافت نگاشت مربوطه را رسم کنید.

دسته‌ها	کیلومترهایی که توسط راننده طی شده است	فراوانی f_i	فراوانی نسبی f'_i
۶-۱۰	از ۵/۵ کیلومتر تا ۱۰/۵ کیلومتر	۱	$\frac{1}{20} = 0.05$
۱۱-۱۵	از ۱۰/۵ کیلومتر تا ۱۵/۵ کیلومتر	۲	$\frac{2}{20} = 0.10$
۱۶-۲۰	از ۱۵/۵ کیلومتر تا ۲۰/۵ کیلومتر	۳	$\frac{3}{20} = 0.15$
۲۱-۲۵	از ۲۰/۵ کیلومتر تا ۲۵/۵ کیلومتر	۵	$\frac{5}{20} = 0.25$
۲۶-۳۰	از ۲۵/۵ کیلومتر تا ۳۰/۵ کیلومتر	۴	$\frac{4}{20} = 0.20$
۳۱-۳۵	از ۳۰/۵ کیلومتر تا ۳۵/۵ کیلومتر	۳	$\frac{3}{20} = 0.15$
۳۶-۴۰	از ۳۵/۵ کیلومتر تا ۴۰/۵ کیلومتر	۲	$\frac{2}{20} = 0.10$
	مجموع	$n = 20$	



۲ رنگ چشم ۱۲۸ فرد به شرح زیر است: ۶۴ نفر قهوه‌ای، ۲۳ نفر آبی، ۳۶ نفر سبز و ۵ نفر سایر رنگ‌هاست. چه نمودارهایی می‌توان برای این اعداد رسم کرد. آن نمودار را رسم کنید؟ **رنگ چشم: متغیر کیفی**

نمودار میله‌ای ☐ نمودار دایره‌ای ☐ هر دو ☒



۳ جملات زیر را کامل کنید:

(الف) برای متغیرهای پیوسته از نمودار... استفاده می‌شود.

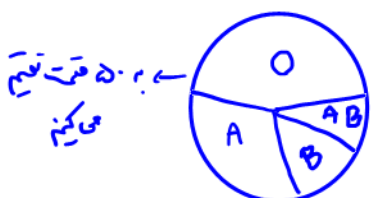
(ب) برای متغیرهای گسسته از نمودارهای... استفاده می‌شود.

(پ) برای متغیرهای کیفی از نمودارهای... استفاده می‌شود.

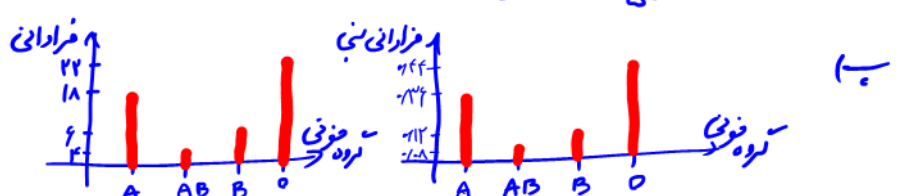
۴ گروه خونی (O) انش آموز پایه یازدهم به صورت زیر گردآوری شده‌اند:

(الف) جدول فراوانی مربوط به گروه خونی این افراد را رسم کنید. (ب) نمودار میله‌ای مربوط به فراوانی و فراوانی نسبی و همچنین نمودار دایره‌ای مربوط به این افراد را رسم کنید؟ (پ) چند درصد افراد، دارای گروه خونی O هستند؟

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



گروه خونی	A	AB	B	O
تعداد (فراوانی)	۱۸	۴	۶	۲۲
فراوانی نسبی	$\frac{18}{50} = 0.36$	$\frac{4}{50} = 0.08$	$\frac{6}{50} = 0.12$	$\frac{22}{50} = 0.44$



$$f'_O \times 100 = 0.44 \times 100 = 44\%$$

۵ اگر فراوانی نسبی مربوط به گروه خونی O، ۴٪ باشد و مجموع فراوانی‌های همه گروه‌های خونی برابر ۲۰ در نظر گرفته شود. فراوانی گروه خونی

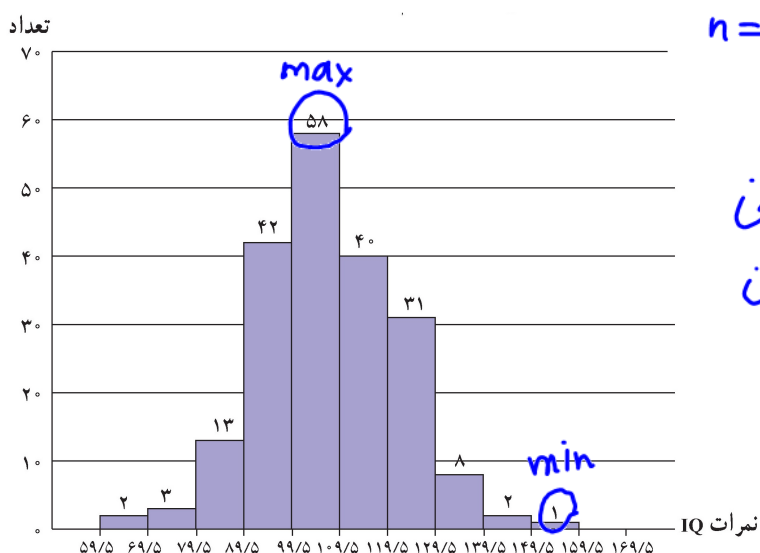
O چه عددی است؟

$$f'_0 = 0.04 \Rightarrow \frac{f_0}{n} = 0.04 \Rightarrow \frac{f_0}{20} = 0.04 \Rightarrow f_0 = 0.04 \times 20 = 0.8$$

$n = 20$

۶ نمودار بافت نگاشت نمرات IQ کودکان یک مهدکودک به صورت زیر رسم شده است. با توجه به این نمودار، به سؤالات زیر پاسخ دهید؟

الف) تعداد کل کودکان که نمره IQ آنها، مورد بررسی قرار گرفته است، چند نفر است؟



$$n = \sum f_i = 2 + 3 + 13 + 42 + 58 + 40 + 31 + 8 + 2 + 1 = 202$$

ب) نمره IQ در کدام رده بیشترین و در کدام رده کمترین فراوانی را دارد؟

در رده (۹۹.۵ - ۱۰۹.۵) بیشترین فراوانی

در رده (۱۳۹.۵ - ۱۴۹.۵) کمترین فراوانی

پ) چند درصد کودکان دارای نمره IQ بین ۱۴۰ تا ۱۶۰ هستند؟

$$\frac{2+1}{202} \times 100 \approx 1.4\%$$

ت) جدول فراوانی آن را رسم کنید؟

رتبه‌ها	۵۹.۵-۶۹.۵	۶۹.۵-۷۹.۵	۷۹.۵-۸۹.۵	...	۱۴۹.۵-۱۵۹.۵
فراوانی	۲	۳	۱۳		۱

۷ جدول فراوانی و نمودارهای مناسب مربوط به تعداد حروف بیت شعر زیر را به دست آورید؟

کیست این پنهان مراد جان و تن کز زبان من همی گوید سخن

حرف	ک	ی	س	ت	ا	ن	...
فراوانی	۲	۴					

حروف فارسی: کمی گستره

نمودارهای مناسب: میله‌ای - دایره‌ای



میانگین - میان - مُد

درس ۲ : معیارهای گرایش به مرکز

A : ۱۴, ۱۵, ۱۶

B : ۱۱, ۱۵, ۱۹

با در اختیار داشتن داده‌های عددی در یک مسئله آماری، سه معیار به نام‌های میانگین، میان و مُد، حد وسط داده‌ها را به گونه‌ای مشخص می‌کنند که به آن‌ها شاخص‌های مرکزی گفته می‌شود. در زیر به بررسی این شاخص‌ها می‌پردازیم.

۱ میانگین

ن بار : $\bar{x}$

الف) میانگین داده‌ها

میانگین n داده آماری $x_1, x_2, \dots, x_n$ را با $\bar{x}$ نشان می‌دهیم که از رابطه زیر به دست می‌آید:

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

چند نکته

۱) اگر $\bar{x}$ میانگین n داده آماری $x_1, x_2, \dots, x_n$ باشد، مجموع اختلاف داده‌ها از $\bar{x}$ برابر با صفر است، یعنی داریم:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x}) = 0$$

۱۵ ← $\bar{x}$ ۱۴, ۱۵, ۱۶

۵ ← $x_i - \bar{x}$: -۱, ۰, ۱

۲) اگر $\bar{x}$ میانگین n داده آماری باشد، آن گاه مجموع کل داده‌ها، حاصل ضرب $\bar{x}$ در n است، یعنی داریم:

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \Rightarrow x_1 + \dots + x_n = n\bar{x}$$

۳) اگر میانگین n داده آماری برابر با A و میانگین m داده آماری دیگر برابر با B باشد، میانگین کل داده‌ها عبارتست از:

$$\bar{x} = \frac{nA + mB}{n + m}$$

مثال ۱) اگر میانگین داده‌های زیر ۱۲ باشد، a را به دست آورید.

$n = 10$: تعداد داده‌ها

۱۳, ۶, ۹, ۱۲, ۱۰, a , ۱۰, ۱۵, ۱۸, ۱۴

$$\bar{x} = 12 \Rightarrow \frac{\sum x_i}{n} = 12 \Rightarrow \frac{13 + 6 + 9 + 12 + 10 + a + 10 + 15 + 18 + 14}{10} = 12 \Rightarrow \frac{a + 107}{10} = 12$$

$$\Rightarrow a + 107 = 120 \Rightarrow a = 120 - 107 = 13$$

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

مثال ۲) اگر میانگین داده‌های $x_1, x_2, \dots, x_{17}$ برابر با ۲۱۹ باشد، میانگین داده‌های زیر چقدر است؟

$$\frac{1}{3}x_1 - 2, \frac{1}{3}x_2 - 4, \frac{1}{3}x_3 - 6, \dots, \frac{1}{3}x_{17} - 34$$

$$\bar{x} = 219 \Rightarrow \frac{x_1 + \dots + x_{17}}{17} = 219 \Rightarrow x_1 + \dots + x_{17} = 17 \times 219 = 3723$$

$$\frac{(\frac{1}{3}x_1 - 2) + (\frac{1}{3}x_2 - 4) + (\frac{1}{3}x_3 - 6) + \dots + (\frac{1}{3}x_{17} - 34)}{17} = \frac{\frac{1}{3}(x_1 + \dots + x_{17}) - (2 + 4 + 6 + \dots + 34)}{17}$$

$$= \frac{\frac{1}{3}(3723) - 2(1 + 2 + 3 + \dots + 17)}{17} = \frac{1241 - 2 \times \frac{17 \times 18}{2}}{17} = \frac{1241 - 306}{17} = \frac{935}{17} = 55$$





مثال ۳) میانگین ۸ داده برابر با ۵ و میانگین ۱۲ داده دیگر برابر ۱۵ است. میانگین این ۲۰ داده را به دست آورید.

روش اول: $\bar{x} = \frac{nA + mB}{n + m} = \frac{1(5) + 12(15)}{1 + 12} = \frac{40 + 180}{20} = \frac{220}{20} = 11$

روش دوم:
$$\begin{cases} \frac{x_1 + \dots + x_8}{8} = 5 \rightarrow x_1 + \dots + x_8 = 8 \times 5 = 40 \\ \frac{y_1 + \dots + y_{12}}{12} = 15 \rightarrow y_1 + \dots + y_{12} = 12 \times 15 = 180 \end{cases} \rightarrow \bar{x} = \frac{(x_1 + \dots + x_8) + (y_1 + \dots + y_{12})}{20} = \frac{40 + 180}{20} = 11$$

مثال ۴) میانگین ۱۰ داده آماری برابر ۵ محاسبه شده است. بعد از محاسبه مشخص شد که دو مقدار ۱۰ و ۱۲ نیز باید به داده‌ها اضافه شود. میانگین جدید چقدر است؟

$$\frac{x_1 + \dots + x_{10}}{10} = 5 \Rightarrow x_1 + \dots + x_{10} = 5 \times 10 = 50$$

$$\bar{x} \text{ (میانگین جدید)} = \frac{(x_1 + \dots + x_{10}) + 10 + 12}{12} = \frac{50 + 22}{12} = \frac{72}{12} = 6$$

مثال ۵) میانگین ۱۰ داده آماری ۳۲/۵ است. اگر دو داده ۳۵ و ۴۰ را از داده‌ها حذف کنیم، میانگین داده‌ها باقی‌مانده چقدر است؟

$$10 \text{ داده آماری: } x_1, x_2, \dots, x_n, 35, 40 \Rightarrow \frac{(x_1 + x_2 + \dots + x_n) + \overbrace{(35 + 40)}^{75}}{10} = \frac{325}{1}$$

$$\Rightarrow (x_1 + \dots + x_n) + 75 = 325 \Rightarrow x_1 + x_2 + \dots + x_n = 325 - 75 = 250$$

$$\bar{x} \text{ (میانگین جدید)} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{250}{n} = \frac{125}{4} = 31,25$$





ب) میانگین موزون (وزن دار)

اگر n داده آماری $x_1, x_2, \dots, x_n$ هر کدام به ترتیب، **فراوانی** یا **وزن** $w_1, w_2, \dots, w_n$ داشته باشند، میانگین آن‌ها که به **میانگین موزون** نیز مشهور است، را با $\bar{x}_w$ نشان می‌دهیم که:

$$\bar{x}_w = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

مثال ۶) نمره‌های یک دانش‌آموز در دروس اختصاصی به صورت زیر است. میانگین آن‌ها را به دست آورید.

درس	ریاضی	فیزیک	شیمی
x_i نمره	۱۶	۱۷	۱۸
w_i ضریب	۴	۳	۲

$$\begin{aligned}\bar{x}_w &= \frac{w_1 x_1 + w_2 x_2 + w_3 x_3}{w_1 + w_2 + w_3} = \frac{(4 \times 16) + (3 \times 17) + (2 \times 18)}{4 + 3 + 2} \\ &= \frac{64 + 51 + 36}{9} \\ &= \frac{151}{9} \\ &\approx 16.78\end{aligned}$$

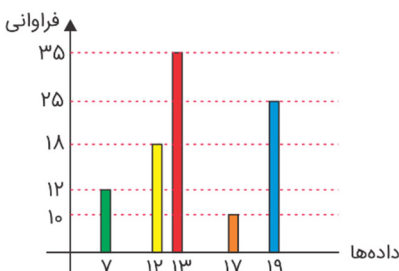
★ مثال ۷) نمره کل آزمون عمومی یک داوطلب مطابق جدول زیر، ۵۸ درصد است. نمره زبان انگلیسی این داوطلب چقدر است؟

درس	ادبیات فارسی	عربی	دینی	زبان انگلیسی
x_i نمره	۶۵	۵۲	۷۰	؟ x
w_i ضریب	۴	۲	۳	۲

$$x = \text{نمره زبان انگلیسی}$$

$$\begin{aligned}\bar{x}_w &= 58 \Rightarrow \frac{(4 \times 65) + (2 \times 52) + (3 \times 70) + (2x)}{4 + 2 + 3 + 2} = 58 \\ &\Rightarrow \frac{260 + 104 + 210 + 2x}{11} = 58 \\ &\Rightarrow \frac{2x + 574}{11} = 58 \Rightarrow 2x + 574 = 638 \Rightarrow 2x = 638 - 574 \Rightarrow 2x = 64 \Rightarrow x = 32\end{aligned}$$

★ مثال ۸) با توجه به نمودار میله‌ای زیر، میانگین کدام است؟



x_i داده	۷	۱۲	۱۳	۱۷	۱۹
w_i فراوانی	۱۲	۱۸	۳۵	۱۰	۲۵

جدول فراوانی :

$$\begin{aligned}\bar{x}_w &= \frac{(7 \times 12) + (12 \times 18) + (13 \times 35) + (17 \times 10) + (19 \times 25)}{12 + 18 + 35 + 10 + 25} \\ &= \frac{84 + 216 + 455 + 170 + 475}{100} = \frac{1400}{100} = 14\end{aligned}$$



اگر داده‌های عددی را به صورت صعودی مرتب کنیم، داده وسط را میانه می‌نامیم و آن را با $\tilde{x}$ نشان می‌دهیم. اگر تعداد داده‌ها زوج باشد، میانگین دو داده میانی، میانه را حاصل می‌کند.

مثال ۹) میانه داده‌های زیر را به دست آورید.

الف) ۲, ۷, ۳, ۵, ۶, ۵, ۱۰ $\xrightarrow{\text{مرتب}} ۲, ۳, ۵, ۵, ۶, ۷, ۱۰ \Rightarrow \text{میانه} = ۵$

ب) ۲, ۳, ۱, ۵, ۶, ۶, ۲, ۴ $\xrightarrow{\text{مرتب}} ۱, ۲, ۲, ۳, ۴, ۵, ۶, ۶ \Rightarrow \text{میانه} = \frac{۳+۴}{۲} = \frac{۷}{۲}$

چند نکته

(۱) میانه برای هر مجموعه‌ای از داده‌ها، همیشه وجود دارد.

(۲) ممکن است میانه، هیچ کدام از داده‌ها نباشد.

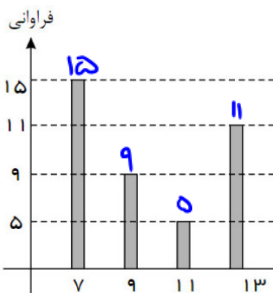
(۳) میانه به مقدار داده‌ها بستگی ندارد، بلکه به تعداد آنها وابسته است.

مثال ۱۰) میانه یک مجموعه از داده‌های مرتب شده، برابر با میانگین چهارمین و پنجمین داده و مجموع کل داده‌ها مساوی ۳۶۰ است. میانگین کل داده‌ها چقدر است؟

$x_1, x_2, x_3, \dots, x_4, x_5, x_6, x_7, x_8 \Rightarrow n=8$
 $\text{میانه} = \frac{x_4 + x_5}{2}$

$\bar{x} = \frac{x_1 + \dots + x_8}{n} = \frac{360}{8} = 45$

مثال ۱۱) با توجه به نمودار میله‌ای فراوانی داده‌های کمی گسسته زیر، میانه را به دست آورید.



جدول فراوانی

$\Rightarrow n = \sum f_i = 15 + 9 + 5 + 11 = 40$

$\Rightarrow$ داده‌های مرتب: $\underbrace{۷, \dots, ۷}_{۱۵ \text{ بار}}, \underbrace{۹, \dots, ۹}_{۹ \text{ بار}}, \underbrace{۱۱, \dots, ۱۱}_{۵ \text{ بار}}, \underbrace{۱۳, \dots, ۱۳}_{۱۱ \text{ بار}}$

$n=40 \Rightarrow \text{میانه} = \frac{x_{20} + x_{21}}{2} = \frac{9+9}{2} = 9$ $(x_1, \dots, x_{20}, x_{21}, \dots, x_{40})$





(۳) مُد (نما)

در میان داده‌های $x_1, x_2, \dots, x_n$ ، داده‌ای که بیشترین فراوانی یا تکرار را دارد، مُد نامیده می‌شود. به عنوان مثال، مُد در داده‌های ۳، ۵، ۲، ۳، ۲، ۲ برابر با ۲ است، زیرا فراوانی آن از دیگر داده‌ها بیشتر است.

چند نکته

- (۱) اگر تمام داده‌ها فراوانی یکسانی داشته باشند، مُد وجود ندارد. به عنوان مثال، داده‌ها ۲، ۲، ۳، ۳، ۴، ۴، ۵، ۵، مُد ندارند.
- (۲) یک جامعه می‌تواند چندین مُد داشته باشد، به عنوان مثال، ۲ و ۵ مدهای داده‌های ۲، ۵، ۵، ۳، ۷، ۱، ۲، ۵ می‌باشند.

مثال ۱۱۲ در داده‌های آماری ۴، ۷، ۹، ۷، ۹، ۴، ۶، ۱، اگر مُد فقط عدد ۹ باشد، میانه داده‌ها کدام است؟

حل) بدون در نظر گرفتن x ، مُد ۴، ۷، ۹ هستند. پس برای اینکه مُد فقط عدد ۹ باشد، باید $x = 9$.

$x = 9 \xrightarrow{\text{مرتب}} 1, 4, 4, 6, 7, 7, 9, 9, 9$
 دوطبقه

$\Rightarrow$ میانه = ۷

مثال ۱۱۳ در مورد داده‌های جدول مقابل، کدام گزینه صحیح است؟

داده	۹	۱۰	۱۳	۱۶	۲۰
فراوانی	۸	۱۴	۹	۱۶	۳

$\uparrow \uparrow$
 $x_{25} \ x_{26}$

- (۱) میانه > میانگین > مُد
 (۲) میانگین > میانه > مُد
 (۳) میانه > مُد > میانگین
 (۴) مُد > میانه > میانگین

$$n = \sum f_i = 8 + 14 + 9 + 16 + 3 = 50 \Rightarrow \text{داده‌ها: } x_1, \dots, x_{25}, x_{26}, \dots, x_{50} \Rightarrow \text{میانه} = \frac{x_{25} + x_{26}}{2} = \frac{13 + 13}{2} = 13$$

✓ $16 = \text{مُد}$ داده مربوطه $16 = \text{بیشترین فراوانی} \Rightarrow \text{داده‌ای با بیشترین فراوانی: مُد}$

$$\bar{x}_w = \frac{9 \times 8 + 10 \times 14 + 13 \times 9 + 16 \times 16 + 20 \times 3}{8 + 14 + 9 + 16 + 3} = \frac{72 + 140 + 117 + 256 + 60}{50} = \frac{645}{50} = 12.9$$

$\Rightarrow$ میانگین > میانه > مُد





تغییرات یکسان روی داده‌ها و اثر آن روی شاخص‌های مرکزی

اگر تمامی داده‌ها را در عددی ضرب و یا با عددی جمع کنیم، همین تغییر روی میانگین، میانه و مُد نیز اعمال می‌شود. به عبارت دیگر اگر $A, \tilde{x}, \bar{x}$ به ترتیب میانگین، میانه و مُد برای داده‌های $x_1, x_2, \dots, x_n$ باشند، آن‌گاه $aA + b, a\tilde{x} + b, a\bar{x} + b$ به ترتیب میانگین، میانه و مُد برای داده‌های زیر هستند:

$$ax_1 + b, ax_2 + b, \dots, ax_n + b$$

(تمام داده‌ها در a ضرب شده و سپس با b جمع شده‌اند.)

مثال ۱۴ اگر میانگین داده‌های $a-1, b+2, c-1$ برابر ۵ باشد، میانگین داده‌های $4c, 2b+6, 2a$ کدام است؟

$$\begin{aligned} x_i &\Rightarrow 2x_i + 2 \\ \downarrow &\quad \downarrow \\ \bar{x} = 5 &\quad \bar{X} = 2\bar{x} + 2 = 2(5) + 2 = 10 + 2 = 12 \end{aligned}$$

مثال ۱۵ میانگین نمرات ریاضی یک کلاس از ۲۰ نمره برابر ۱۶/۵ است. اگر بخواهیم به همه یک نمره اضافه کرده و نمره کلاسی را از ۲۰ به ۳۲ تغییر دهیم، میانگین کلاس چقدر می‌شود؟

$$\bar{x} = 16.5$$

$$x_i \Rightarrow x_i + 1 \Rightarrow \frac{32}{20}(x_i + 1) = \frac{1}{5}(x_i + 1) = \frac{1}{5}x_i + \frac{1}{5}$$

$$\bar{X} = \frac{1}{5}\bar{x} + \frac{1}{5} = \frac{1}{5}(\bar{x} + 1) = \frac{1}{5}(16.5 + 1) = \frac{1}{5} \times 17.5 = \frac{17.5}{5} = 3.5$$

داده دور افتاده

به داده ای که اختلاف بسیار زیادی با بقیه داده‌ها دارد، داده دور افتاده می‌گوییم. مثلاً داده ۳۰۰ در داده‌های زیر، دور افتاده است:

$$4, 2, 8, 300, 10, 5, 6, 8, 3, 5, 4$$

با داشتن داده دور افتاده، میانگین شاخص مناسبی برای این داده‌ها نیست و میانه مناسب‌تر است.

مثال ۱۶ در هر یک از موارد زیر، کدام یک از شاخص‌های آماری میانه، مُد و میانگین را به عنوان یک شاخص مناسب انتخاب می‌کنید؟

(الف) رأی گیری برای انتخابات ریاست جمهوری مُد (فردی رئیس جمهور می‌شود که بیشترین رأی را دارد)

(ب) در اختیار داشتن لیست نمره‌های ریاضی یک کلاس برای بررسی وضعیت این درس میانگین

(پ) متوسط درآمد یک فروشگاه در یک روز با در اختیار داشتن درآمد روزانه آن در یک ماه میانه (زیرا ممکن

است داده دور افتاده وجود داشته باشد)



۱ تعداد حمله‌های یک تیم فوتبال در شش مسابقه گذشته به صورت ۴۸، ۴۵، ۴۴، ۴۵، ۴۲، ۴۳ است.

میانگین تعداد حملات این تیم در شش بازی گذشته را به دست آورید؟

$$\bar{x} = \frac{\sum x_i}{n} = \frac{48+45+44+45+42+43}{6} = \frac{267}{6} = \frac{89}{2} = 44,5$$

۲ بالاترین دما در هریک از روزهای هفته گذشته اندازه‌گیری شده و نتایج زیر به دست آمده است. معدل یا میانگین دما در هفته گذشته چه عددی است؟

۵۵، ۲۷، ۲۹، ۳۲، ۲۸، ۳۱، ۲۹

$$\bar{x} = \frac{55+27+29+32+28+31+29}{7}$$

۳ میانه و مد هر یک از داده‌های زیر را به دست آورید؟

(ب) ۶۰، ۵۰، ۴۰، ۲۴، ۳۰۰

(الف) ۸، ۹، ۹، ۹، ۹

(پ) ۱۵، ۸، ۳، ۱۰

حزب

$$\rightarrow 3, 8, 10, 15 \rightarrow \begin{cases} \text{میان} = \frac{8+10}{2} = 9 \\ \text{مُد} = \text{وجود ندارد} \end{cases}$$

حزب

$$\rightarrow 8, 9, 9, 9, 9 \rightarrow \begin{cases} \text{میان} = 9 \\ \text{مُد} = 9 \end{cases}$$

(ج) ۷، ۴، ۱۳، ۷

(ث) ۲۳، ۱۲، ۱۲، ۲۳

(ت) ۵، ۱۲، ۹، ۶، ۴

حزب

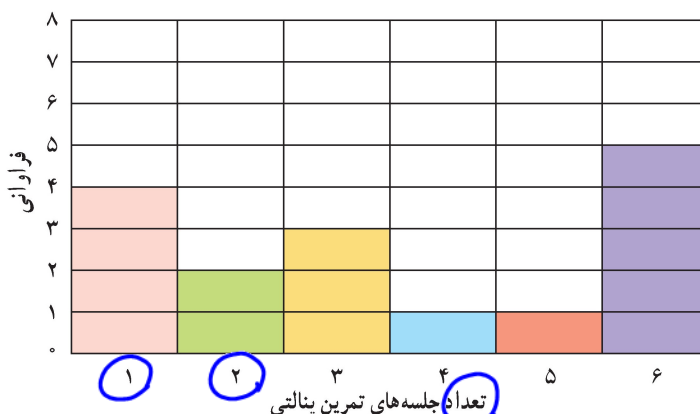
$$\rightarrow 4, 7, 7, 13 \rightarrow \begin{cases} \text{میان} = \frac{7+7}{2} = 7 \\ \text{مُد} = 7 \end{cases}$$

حزب

$$\rightarrow 12, 12, 23, 23 \rightarrow \begin{cases} \text{میان} = \frac{12+23}{2} = \frac{35}{2} = 17,5 \\ \text{مُد} = \text{وجود ندارد} \end{cases}$$

۴ نمودار زیر، نمودار میله‌ای مربوط به تعداد ضربات پنالتی گل شده یک بازیکن در شش جلسه تمرین پنالتی است. با توجه به نمودار، میانگین، میانه و مد

تعداد ضربات گل شده را به دست آورید؟



تعداد گل‌ها در
جلسه‌های مختلف

تعداد گل‌ها در جلسه‌های تمرین پنالتی

$$\rightarrow 1, 2, 3, 4, 5 \rightarrow \begin{cases} \bar{x} = \frac{16}{4} = 4 \\ \text{میان} = \frac{2+3}{2} = 2,5 \\ \text{مُد} = 1 \end{cases}$$

۵ در جدول زیر، نمرات درس ریاضی ۱۰ دانش آموز گردآوری شده و میانگین نمرات داده شده است. علامت‌های سؤال چه اعدادی اند؟

۱۷/۵	۱۹	۱۷	۱۶	۲۰	نمرات درس ریاضی
۱۶	۱۵	۱۸	۱۸	۱۸	

میانگین نمرات = ۱۵/۶۵

مد نمرات = ؟

۱۸ و ۱۶ = مد

$$\bar{x} = 15.65$$

$$n = 10, \bar{x} = 15.65$$

$$\bar{x} = 15.65 \rightarrow \frac{\sum x_i}{n} = 15.65 \Rightarrow \frac{156.5 + x}{10} = 15.65 \Rightarrow$$

$$156.5 + x = 156.5 \Rightarrow \underline{x = 0}$$

۶ داده‌های زیر مدت زمان مطالعه یک دانش آموز را در روزهای هفته نشان می‌دهد.

روزهای هفته	شنبه	یکشنبه	دوشنبه	سه شنبه	چهارشنبه	پنج شنبه	جمعه
مدت زمان مطالعه (ساعت)	۲	۱/۵	۲/۵	۱/۵	۲	۳	۳

این دانش آموز به طور میانگین چند ساعت در روز، در هفته گذشته مطالعه کرده است؟

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2 + 3 + 2 + 1.5 + 2.5 + 1.5 + 2}{7} = \frac{15.5}{7} \approx 2.21$$

۷ یک شرکت بیمه برای تعیین حق بیمه شخص ثالث در سال آینده، نمونه‌ای از خسارت‌های پرداخت شده امسال را جمع‌آوری نموده است.

میانگین خسارت‌های پرداخت شده برابر ۸۵ میلیون ریال به دست آمده است در صورتی که میانه و مد آن برای این خسارت‌های پرداخت شده برابر

۴۲/۲ میلیون ریال و عدد ۹۰ میلیون ریال می‌باشد. به نظر شما مدیر شرکت، کدام معیار گرایش به مرکز را به منظور تعیین حق بیمه در سال آینده

در نظر بگیرد تا اینکه این شرکت ضرر نکند؟

$$\text{میزان خسارت} \begin{cases} \bar{x} = 85 \\ \text{میانه} = 42.2 \\ \text{مد} = 90 \end{cases}$$

میانگین یا مد

۸ دانش آموزی در کنکور سراسری شرکت می کند و نتیجه کارنامه آزمون آن به شرح زیر است :

مواد امتحانی	ریاضیات	فیزیک	شیمی	زبان انگلیسی	ادبیات و زبان فارسی	دین و زندگی
درصد	۵۳	۸؟	۶۷	۳۴	۸۰	۶۷
ضریب درس	۴	۳	۱	۱	۴	۳

x_i

w_i

اگر معدل موزون درصد این دانش آموز ۷۳ باشد، درس فیزیک را چند درصد زده است؟

$$\bar{x}_w = 73 \Rightarrow \frac{\sum x_i w_i}{\sum w_i} = \frac{(4 \times 53) + (3x) + 67 + 34 + (1 \times 80) + (3 \times 67)}{4 + 3 + 1 + 1 + 4 + 3} = 73$$

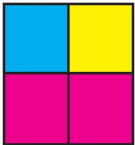
$$\Rightarrow \frac{212 + 3x + 101 + 320 + 201}{16} = 73 \Rightarrow 834 + 3x = 1168 \Rightarrow x = \frac{334}{3}$$

$$\rightarrow x = 111$$

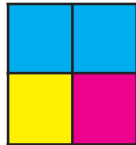
۹ میانگین ۵ داده آماری ۱۷ است، اگر دو عدد ۱۷ و ۱۱ را به داده های قبلی اضافه کنیم، میانگین جدید چه عددی خواهد شد؟

$$\bar{x} = 17 \Rightarrow \frac{\sum_{i=1}^5 x_i}{5} = 17 \Rightarrow \sum_{i=1}^5 x_i = 5 \times 17 = 85$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{85 + (11 + 17)}{5 + 2} = \frac{113}{7} \approx 16.14$$



$$\begin{pmatrix} 470 & 580 \\ 690 & 690 \end{pmatrix}$$



$$\begin{pmatrix} 470 & 470 \\ 580 & 690 \end{pmatrix}$$

۱۰ دو دانش آموز، جدول های چهارخانه ای را به صورت روبه رو رنگ آمیزی کرده اند، بر اساس جدول مربوط به طول موج طیف رنگ ها، جدول عددی این دو شکل به صورت روبه رو نشان داده شده است :

حال جدول عددی مربوط به این دو شکل را ابتدا با هم جمع و سپس هریک از اعضای جدول عددی را به عدد ۲ تقسیم می کنیم. جدول عددی حاصل را به دست آورده و شکل مورد نظر را با توجه به جدول طول موج طیف رنگ ها، به دست آورید. آیا این شکل میانگین دو شکل بالا است؟

برای پاسخ به این سؤال، کاربرد علم آمار در علوم شناختی و مغز را مطالعه کنید. عدد مربوط به طول موج طیف رنگ ها در جدول موجود در حاشیه نشان داده شده است.

طول موج طیف رنگ ها (بر حسب نانومتر)	رنگ ها
۴۵۰ تا ۴۹۵	Blue
۴۹۵ تا ۵۷۰	Green
۵۷۰ تا ۵۹۰	Yellow
۵۹۰ تا ۶۲۰	Orange
۶۲۰ تا ۷۵۰	Pink

درس: آمار توصیفی مبحث: معیارهای پراکندگی



۱۵ = میان = میانگین $\rightarrow A: 14, 15, 16$
 ۱۵ = میان = میانگین $\rightarrow B: 11, 15, 19$
درس ۳: معیارهای پراکندگی

برای تعیین میزان پراکندگی داده‌ها، سه شاخص واریانس، انحراف معیار و ضریب تغییرات وجود دارند که در زیر تعریف می‌شوند.

۱) واریانس

A: $(-1)^2 + (0)^2 + (1)^2 = 2 \rightarrow \frac{2}{3}$
 B: $(-4)^2 + (0)^2 + (4)^2 = 32 \rightarrow \frac{32}{3}$

برای داده‌ها عددی $x_1, x_2, \dots, x_n$ واریانس آنها را σ^2 نشان می‌دهیم و به صورت زیر به دست می‌آوریم:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

تذکره: به کمک فرمول بالا، می‌توان نشان داد که واریانس از فرمول زیر نیز به دست می‌آید:

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - (\bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$$

چند نکته

- ۱) واحد واریانس، مربع واحد داده‌ها است، یعنی اگر داده‌ها بر حسب متر (m) باشند، واریانس بر حسب m^2 خواهد بود.
- ۲) مقدار واریانس، رابطه مستقیم با پراکندگی داده‌ها دارد. واریانس بزرگ نشان دهنده پراکندگی زیاد و واریانس کوچک، پراکندگی کمتر داده‌ها را نشان می‌دهد.
- ۳) اگر واریانس صفر باشد، تمام داده‌ها با یکدیگر برابرند، (پراکندگی وجود ندارد).

مثال ۱) واریانس داده‌های ۲, ۲, ۲, ۲, ۲, ۲, ۶, ۶ را محاسبه کنید.

$$n = 8$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2 + \dots + 2 + 6 + 6}{8} = \frac{12 + 12}{8} = \frac{24}{8} = 3$$

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{8} [1 + 1 + 1 + 1 + 1 + 1 + 9 + 9] = \frac{24}{8} = 3 \rightarrow \sigma^2 = 3$$

مثال ۲) اگر واریانس داده‌های $x_1, x_2, \dots, x_5$ برابر صفر باشد، واریانس داده‌های $x_1 + 1, x_2 + 2, x_3 + 3, x_4 + 4, x_5 + 5$ چقدر است؟

$$x_1, x_2, \dots, x_5 \xrightarrow{\sigma^2 = 0} x_1 = x_2 = \dots = x_5 = a$$

$$x_1, x_2 + 1, x_3 + 2, x_4 + 3, x_5 + 4 : a, a+1, a+2, a+3, a+4$$

$$\bar{x} = a + 2 \xrightarrow{n=5} \sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{5} [4 + 1 + 0 + 1 + 4] = \frac{10}{5} = 2 \Rightarrow \sigma^2 = 2$$





مثال ۳) اگر میانگین داده‌های $x, ۳۹, ۳۶, ۴۱, ۳۸, ۳۵, ۴۰, ۴۲$ برابر ۳۹ باشد، واریانس این داده‌ها را تعیین کنید.

$$\bar{x} = ۳۹ \rightarrow \frac{\sum x_i}{n} = ۳۹ \Rightarrow \frac{۴۲+۴۰+۳۵+۳۸+۴۱+۳۶+۳۹+x}{۸} = ۳۹ \Rightarrow$$

$$\frac{۲۷۱+x}{۸} = ۳۹ \Rightarrow ۲۷۱+x = ۳۱۲ \Rightarrow \underline{x = ۴۱}$$

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{۸} [۹+۱+۱۹+۱+۴+۹+۰+۴] = \frac{۴۴}{۸} = \frac{۱۱}{۲} = ۵.۵$$

مثال ۴) اگر میانگین و واریانس ۲۰ داده آماری به ترتیب ۱۲ و ۱۵ باشند و داده‌های ۱۰ و ۱۴ به آنها اضافه شود، واریانس کل داده‌ها چقدر است؟

$$x_1, x_2, \dots, x_{20} \rightarrow \begin{cases} \bar{x} = ۱۲ \Rightarrow x_1 + \dots + x_{20} = ۲۰ \times ۱۲ = ۲۴۰ \\ \sigma^2 = ۱۵ \Rightarrow \frac{1}{20} \sum (x_i - \bar{x})^2 = ۱۵ \Rightarrow \sum_{i=1}^{20} (x_i - \bar{x})^2 = ۳۰۰ \end{cases}$$

$$x_1, x_2, \dots, x_{20}, ۱۰, ۱۴ \Rightarrow \bar{x} = \frac{(x_1 + \dots + x_{20}) + ۱۰ + ۱۴}{۲۲} = \frac{۲۴۰ + ۲۴}{۲۲} = \frac{۲۶۴}{۲۲} = ۱۲$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{22} (x_i - \bar{x})^2 = \frac{1}{۲۲} \left[\sum_{i=1}^{20} (x_i - \bar{x})^2 + (۱۰ - ۱۲)^2 + (۱۴ - ۱۲)^2 \right]$$

$$= \frac{1}{۲۲} [۳۰۰ + ۴ + ۴] = \frac{۳۰۸}{۲۲} = ۱۴ \Rightarrow \underline{\sigma^2 = ۱۴}$$

۲) انحراف معیار

جزر واریانس، انحراف معیار نامیده می‌شود و آن را با σ نشان می‌دهیم. بنابراین انحراف معیار از روابط زیر می‌تواند به دست آید:

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \Rightarrow \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2}$$

چند نکته

۱) واحد انحراف معیار، همان واحد داده‌ها است، یعنی اگر داده‌ها بر حسب متر (m) باشند، انحراف معیار نیز بر حسب m است.
 ۲) اگر انحراف معیار صفر باشد، تمام داده‌ها با یکدیگر برابرند، (پراکندگی وجود ندارد).

$$\sigma = 0 \Leftrightarrow x_1 = \dots = x_n$$





مثال ۵ ۱۵ داده آماری با واریانس ۱۲ را با ۱۰ داده دیگر با واریانس ۷/۶ ترکیب می‌کنیم. اگر میانگین هر دو گروه یکسان باشند، انحراف معیار کل داده‌ها را به دست آورید.

$$\begin{cases} n_1 = 15 \\ \sigma_1^2 = 12 \rightarrow \frac{1}{15} \sum (x_i - \bar{x})^2 = 12 \Rightarrow \sum_{i=1}^{15} (x_i - \bar{x})^2 = 180 \\ \bar{x} \end{cases}$$

$$\sigma^2 = \frac{1}{n_1 + n_2} (n_1 \sigma_1^2 + n_2 \sigma_2^2)$$

$$\begin{cases} n_2 = 10 \\ \sigma_2^2 = 7/6 \rightarrow \frac{1}{10} \sum (y_i - \bar{y})^2 = 7/6 \Rightarrow \sum_{i=1}^{10} (y_i - \bar{y})^2 = 76 \\ \bar{y} \end{cases}$$

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_{i=1}^{25} (z_i - \bar{z})^2 = \frac{1}{25} \left[\underbrace{\sum (x_i - \bar{x})^2}_{n_1 \sigma_1^2} + \underbrace{\sum (y_i - \bar{y})^2}_{n_2 \sigma_2^2} \right] \\ &= \frac{1}{25} [180 + 76] = \frac{256}{25} \Rightarrow \sigma = \sqrt{\frac{256}{25}} = \frac{16}{5} = 3.2 \end{aligned}$$

مثال ۶ اگر اختلاف ۵ داده آماری از میانگینشان برابر با $a, -1, 0, 2, 2a+5$ باشد، انحراف معیار این داده‌ها کدام است؟

$$\begin{cases} \text{اختلاف داده‌ها از میانگین: } x_i - \bar{x} : 2a+5, 2, 0, -1, a \\ \sigma = ? \end{cases}$$

نکته: $\sum (x_i - \bar{x}) = 0 \Rightarrow (2a+5) + 2 + 0 - 1 + a = 0 \Rightarrow 3a+6=0 \Rightarrow a=-2$

$$\begin{aligned} a = -2 &\Rightarrow x_i - \bar{x} : 1, 2, 0, -1, -2 \Rightarrow n=5 \\ \sigma^2 &= \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{5} [1 + 4 + 0 + 1 + 4] = \frac{10}{5} = 2 \Rightarrow \sigma = \sqrt{2} \end{aligned}$$

(۳) ضریب تغییرات

نسبت انحراف معیار به میانگین را ضریب تغییرات نامیده و با CV نشان می‌دهیم، پس:

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

چند نکته

(۱) ضریب تغییرات، میزان پراکندگی داده‌ها را به ازای یک واحد از میانگین نشان می‌دهد.

(۲) در یک جامعه آماری که $\bar{x} = 0$ یا $\bar{x} < 0$ است، ضریب تغییرات تعریف نمی‌شود.

(۳) ضریب تغییرات، یک شاخص بدون واحد است. بنابراین برای مقایسه میزان پراکندگی دو جامعه آماری از جنس متفاوت (مثلاً زمان و جرم)، از این شاخص برای مقایسه آن‌ها استفاده می‌کنیم.

(۴) از ضریب تغییرات برای تشخیص دقت یک کار و یا میزان یک دست بودن یک محصول استفاده می‌شود. هر چقدر این شاخص کمتر باشد، دقت انجام کار و یا یکنواختی محصول بیشتر است.





مثال ۷) ضریب تغییرات داده‌های زیر را به دست آورید.

$$\underbrace{10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 14, 14, 14, 14, 14, 14}_{n=16}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5(10) + 4(11) + 7(14)}{16} = \frac{50 + 44 + 98}{16} = \frac{192}{16} = 12$$

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{16} [5(4) + 4(1) + 7(4)] = \frac{1}{16} [20 + 4 + 28] = \frac{52}{16} = \frac{13}{4}$$

$$\Rightarrow CV = \frac{\sigma}{\bar{x}} = \frac{\sqrt{\frac{13}{4}}}{12} = \frac{\frac{\sqrt{13}}{2}}{\frac{12}{1}} = \frac{\sqrt{13}}{24}$$

مثال ۸) میانگین و انحراف معیار عمر لاستیک‌های تولیدی کارخانه A به ترتیب برابر با ۸۵۰۰۰ و ۱۰۰۰ کیلومتر و برای کارخانه B، این مقادیر، ۱۲۰۰۰ و ۱۲۰۰ کیلومتر است. محصول کدام کارخانه را انتخاب می‌کنید؟

محصول کارخانه‌ای که ضریب تغییرات آن کم‌تر است و به عبارت دیگر

کارخانه‌ای که ضریب تغییرات کم‌تری دارد.

$$A: CV_1 = \frac{\sigma}{\bar{x}} = \frac{1000}{85000} = \frac{1}{85}$$

$$\frac{\frac{1}{85} \boxed{<} \frac{3}{250}}{250 < 255} \Rightarrow CV_1 < CV_2$$

$$B: CV_2 = \frac{\sigma}{\bar{x}} = \frac{1200}{10000} = \frac{12}{1000} = \frac{3}{250}$$

بنابراین کارخانه A محصولات کم‌تفاوت‌تری دارد.





تغییرات یکسان روی داده‌ها و اثر آن روی شاخص‌های پراکندگی

اگر تمامی داده‌ها را در عددی ضرب و سپس با عددی جمع کنیم، تنها تغییرات ضریبی روی واریانس و انحراف معیار اعمال می‌شود. به عبارت دیگر اگر σ^2 و σ به ترتیب واریانس و انحراف معیار داده‌های $x_1, \dots, x_n$ باشند، آن‌گاه $a^2\sigma^2$ و $a|\sigma|$ به ترتیب واریانس و انحراف معیار برای داده‌های زیر هستند:

$$ax_1 + b, ax_2 + b, \dots, ax_n + b$$

نکته: اگر تمام داده‌ها تنها با عدد b جمع شود و $b > 0$ باشد، واریانس و انحراف معیار تغییر نمی‌کند ولی ضریب تغییرات داده‌های جدید کاهش می‌یابد.

مثال (۹) اگر واریانس قیمت‌ها برای انواع یک محصول در سال گذشته، ۱۰۰۰ تومان و امسال ۳۰ درصد به قیمت‌ها اضافه شود، واریانس قیمت‌های امسال چقدر است؟ ضریب تغییرات چه تغییری می‌کند؟

$$\begin{cases} \sigma_1^2 = 1000 & : x_i \\ \sigma_2^2 = ? & : x_i + 0.3x_i = 1.3x_i \end{cases}$$

$$\sigma_2^2 = a^2 \sigma_1^2 = (1.3)^2 \times 1000 = 1.69 \times 1000 = 1690$$

$$CV_2 = \frac{\sigma_2}{\bar{x}_2} = \frac{1.3\sigma_1}{1.3\bar{x}_1} = \frac{\sigma_1}{\bar{x}_1} = CV_1 \Rightarrow CV_2 = CV_1 \Rightarrow \text{ضریب تغییرات تغییری نمی‌کند}$$

مثال (۱۰) اگر واریانس داده‌های $a+1, a+2, a+3, a+4$ برابر ۳ باشد، واریانس داده‌های $2a+10, 2a+12, 2a+14, 2a+16$ کدام است؟

$$\begin{array}{ccc} x_i & \Rightarrow & 2x_i + 8 \\ \downarrow & & \downarrow \\ \sigma^2 & = & 3 \end{array} \quad \begin{array}{ccc} x_i & \Rightarrow & 2x_i + 8 \\ \downarrow & & \downarrow \\ \sigma'^2 & = & (2)^2 \sigma^2 = 4\sigma^2 = 4 \times 3 = 12 \end{array}$$

مثال (۱۱) اگر ۲۰ داده آماری را دو برابر کرده و از هر کدام ۳ واحد کم کنیم، ضریب تغییرات داده‌های جدید سه برابر می‌شود. میانگین داده‌های اولیه کدام است؟

$$\begin{cases} n = 20 \\ x_i \Rightarrow 2x_i - 3 \\ \frac{CV_2}{CV_1} = 3 \\ \bar{x}_1 = ? \end{cases} \quad \begin{aligned} x_i &\Rightarrow 2x_i - 3 \Rightarrow \begin{cases} \bar{x}_2 = 2\bar{x}_1 - 3 \\ \sigma_2 = 2\sigma_1 \end{cases} \\ \frac{CV_2}{CV_1} = 3 &\Rightarrow \frac{\frac{\sigma_2}{\bar{x}_2}}{\frac{\sigma_1}{\bar{x}_1}} = 3 \Rightarrow \frac{\frac{2\sigma_1}{2\bar{x}_1 - 3}}{\frac{\sigma_1}{\bar{x}_1}} = 3 \Rightarrow \frac{2\bar{x}_1}{2\bar{x}_1 - 3} = 3 \Rightarrow \\ &2\bar{x}_1 - 9 = 2\bar{x}_1 \Rightarrow 4\bar{x}_1 = 9 \Rightarrow \bar{x}_1 = \frac{9}{4} = 2.25 \end{aligned}$$





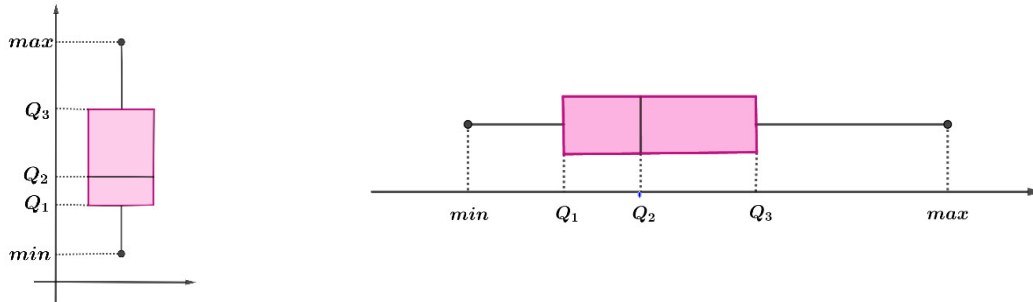
نمودار جعبه‌ای Box plot

نمودار جعبه‌ای، پراکندگی و همچنین **تمرکز داده‌ها** را هم‌زمان نشان می‌دهد. برای رسم این نمودار، داده‌ها را به صورت

صعودی مرتب کرده و ۵ مقدار زیر را مشخص می‌کنیم:

کوچک‌ترین داده (min)، بزرگ‌ترین داده (max)، میانه ($\bar{x}$ یا Q_2)، چارک اول (Q_1) و چارک سوم (Q_3)، که چارک اول، میانه داده‌های قبل از میانه و

چارک سوم، میانه داده‌های بعد از میانه است. بعد از تعیین این مقادیر، نمودار جعبه‌ای به یکی از صورت‌های زیر رسم می‌شود:



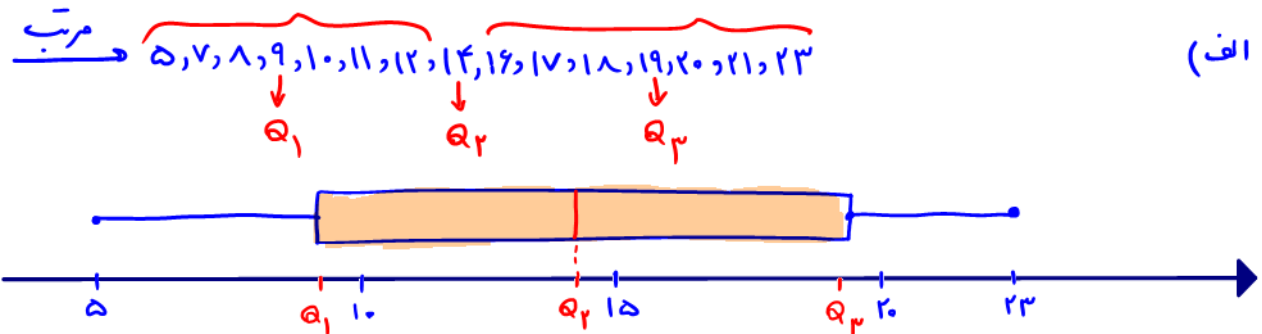
اختلاف چارک اول از چارک سوم را **دامنه میان چارکی** گفته و با IQR نشان می‌دهیم، پس: $IQR = Q_3 - Q_1$ (طول مستطیل)

مثال ۱۲ برای داده‌های آماری : ۸، ۷، ۲۱، ۱۹، ۵، ۱۰، ۱۲، ۹، ۱۱، ۱۷، ۱۴، ۱۶، ۲۳، ۲۰، ۱۸

الف) نمودار جعبه‌ای را رسم کنید.

ب) دامنه میان چارکی این داده‌ها را به دست آورید.

پ) انحراف معیار داده‌های داخل جعبه را تعیین کنید.



ب) $IQR = Q_3 - Q_1 = 19 - 9 = 10$

پ) داده‌های داخل جعبه : ۱۰، ۱۱، ۱۲، ۱۴، ۱۶، ۱۷، ۱۸ $\xrightarrow{n=7} \bar{x} = \frac{\sum x_i}{n} = \frac{91}{7} = 13$

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{7} [1 + 9 + 4 + 0 + 4 + 9 + 16] = \frac{51}{7} \Rightarrow \sigma = \sqrt{\frac{51}{7}}$$

مثال ۱۳ اگر میانگین داده‌های دنباله سمت چپ و راست یک نمودار جعبه‌ای، به ترتیب ۱۰ و ۱۸ و همچنین میانگین ۱۷ داده داخل و روی جعبه برابر

با ۱۵ باشد، میانگین کل داده‌ها چقدر است؟

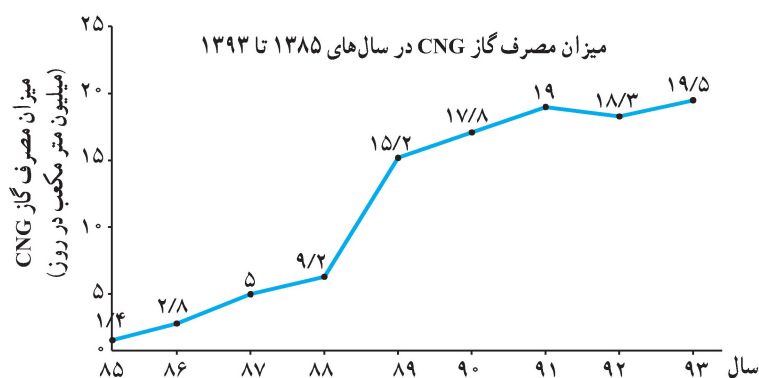


۱ فرض کنید سنّ افرادی که در یک روز سوار اتوبوس شده‌اند، به صورت زیر است :

۳۲،۵۹،۲۶،۵۳،۷۴،۱۷،۴۵،۲۳،۶۴،۵۰،۶۱

انحراف معیار، واریانس و ضریب تغییرات سنّ افراد را به دست آورید.

۲ نمودار زیر میزان مصرف گاز CNG را از سال ۱۳۸۵ تا ۱۳۹۳ نشان می‌دهد. با توجه به این نمودار انحراف معیار، واریانس و ضریب تغییرات میزان مصرف گاز CNG از سال ۱۳۸۵ تا ۱۳۹۳ را به دست آورید.



۳ انحراف معیار، واریانس و ضریب تغییرات را برای هریک از اعداد جدول زیر به دست آورید.

اعداد	انحراف معیار	واریانس	ضریب تغییرات
۱۰۰، ۱۲، ۸، ۱۶، ۱۰، ۴، ۷			
۱۰/۱۱، ۱۱/۳۶، ۱۰/۱۱ ۹/۸۸، ۹/۴۲، ۹/۷۶، ۹/۶۲			
۲، ۳۰۰۰، ۲۵۰۰، ۲۰۰۰			

۴ اعداد دلخواه را در جدول زیر بنویسید و انحراف معیار، واریانس و ضریب تغییرات را برای هریک از اعداد به دست آورید.

اعداد	انحراف معیار	واریانس	ضریب تغییرات

۵ اگر ضریب تغییرات ۱۰ داده ۲ باشد و میانگین آن ۴، واریانس داده‌ها را به دست آورید.

۶ اگر n داده را c برابر کنیم ضریب تغییرات داده‌ها چند برابر می‌شود؟

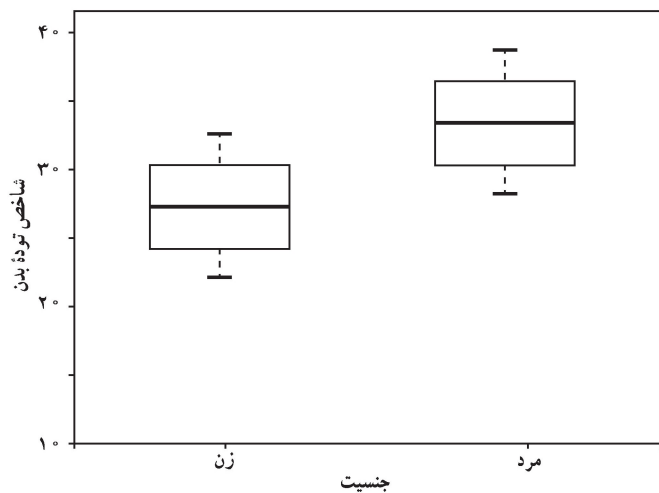
۷ فرض کنید ۲۲ بوته گل قرمز را انتخاب و تعداد گل‌های هر بوته را شمرده‌ایم و نتایج زیر به دست آمده است :

۷، ۴، ۳، ۸، ۶، ۴، ۱، ۷، ۴، ۲، ۱، ۱، ۱، ۱، ۳، ۲، ۲، ۲، ۲، ۵، ۵، ۱، ۲

نمودار جعبه‌ای را برای این داده‌ها رسم کنید.

۸ نمودار جعبه‌ای مربوط به شاخص توده بدن (BMI) به تفکیک جنسیت رسم شده است. این نمودار را تفسیر کنید و به سؤالات زیر پاسخ دهید.

الف) میانۀ شاخص توده بدن در خانم‌ها بیشتر است یا آقایان؟



ب) میزان پراکندگی شاخص توده بدن در خانم‌ها بیشتر است یا آقایان؟

۹ داده‌های زیر مربوط به نرخ بیکاری یک کشور در ده سال گذشته است :

سال	اول	دوم	سوم	چهارم	پنجم	ششم	هفتم	هشتم	نهم	دهم
نرخ بیکاری	۱۱/۵	۱۱/۳	۱۰/۵	۱۰/۴	۱۱/۹	۱۳/۵	۱۲/۳	۱۲/۲	۱۰/۴	۳۰/۱

نمودار جعبه‌ای این داده‌ها را رسم کنید.