جزوه آمار و احتمال

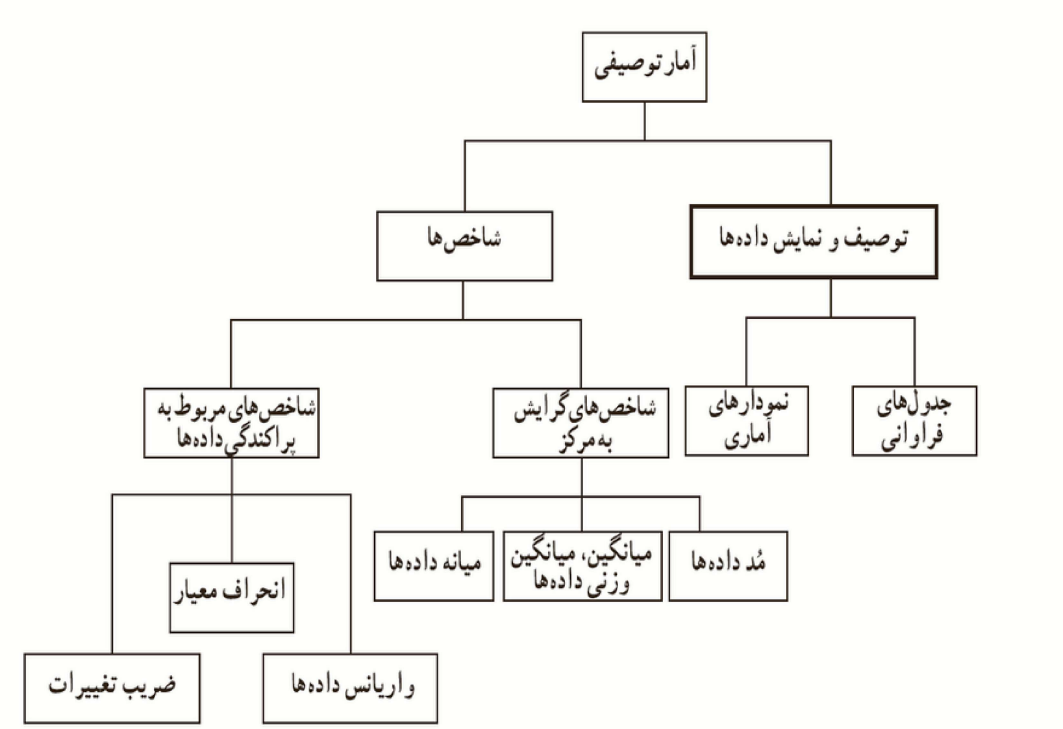
پایه یازدهم ریاضی

فصل سوم و چهارم: آمار توصیفی و استنباطی

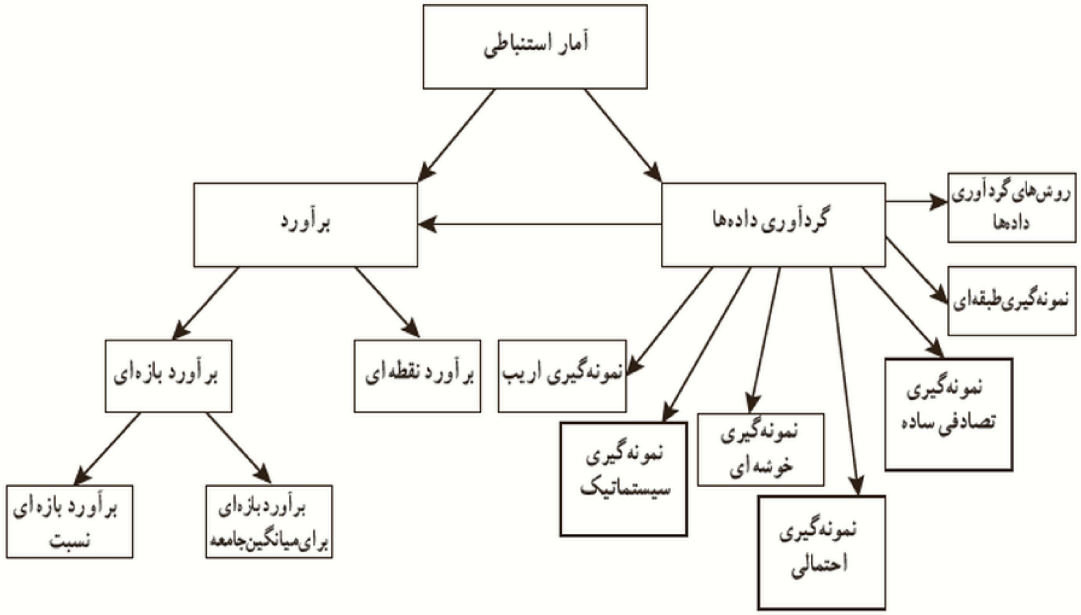


مولف: سید ابوالفضل فاضلیان

نمای کلی فصل سوم:



نمای کلی فصل چهارم:



فصل سوم: آمار توصیفی

یادآوری:

آمار: مجموعه ای از اعداد، ارقام و اطلاعات است.

علم آمار: مجموعه روش هایی است شامل جمع آوری اعداد و ارقام، سازماندهی و نمایش، تحلیل و تفسیر داده ها و در آخر نتیجه گیری، قضاوت و پیش بینی، در مورد آزمایش های تصادفی است.

جامعه آماری: مجموعه تمام افراد یا اشیایی که درباره یک یا چند ویژگی آن ها تحقیق صورت می گیرد.

نمونه: بخشی از جامعه را که برای مطالعه انتخاب شود، نمونه گوئیم.

به تعداد اعضای جامعه، اندازه جامعه یا حجم جامعه و به تعداد اعضای نمونه، اندازه نمونه یا حجم نمونه می گوئیم.

متغیر: هر ویژگی از جامعه آماری که در اعضای جامعه یک سان نیست و از عضوی به عضو دیگر تغییر می کند را متغیر می گوئیم و به خود یا عبارتی که به متغیر نسبت داده می شود مقدار متغیر می گوئیم.

انواع متغیر: دو نوع متغیر داریم: الف) کمی ب) کیفی

الف) متغیر کمی: متغیرهایی که قابل اندازه گیری بوده و مقادیر عددی به خود می گیرند. مانند قد، وزن، تعداد و ...

متغیرهای کمی به دو نوع پیوسته و گسسته تقسیم می‌شوند:

۱) متغیر کمی پیوسته: اگر دو متغیر a و b را بگیرید هر مقدار بین آن‌ها را هم می‌توانند اختیار کنند. مانند قد، سن، وزن و...

۲) متغیر کمی گسسته: متغیر کمی‌ای که پیوسته نباشد مانند تعداد فرزندان، تعداد مکالمات تلفنی در شبانه‌روز و...

ب) متغیر کیفی: متغیرهایی که قابل اندازه‌گیری نیستند و صرفاً برای دسته‌بندی افراد یا اشیاء در گروه‌ها به کار می‌روند و لزوماً مقدار عددی ندارند مانند: جنسیت، گروه خونی و رنگ خودرو و...

متغیرهای کیفی به دو نوع ترتیبی و اسمی تقسیم می‌شوند:

۱) متغیر کیفی ترتیبی: متغیرهایی که نوعی ترتیب طبیعی در آن‌ها وجود دارد مانند سطح تحصیلات، کیفیت یک میوه (درجه ۱، درجه ۲ و...) و مراحل زندگی (کودکی، جوانی و...)

۲) متغیر کیفی اسمی: متغیرهایی که ترتیبی نداشته مانند گروه خونی، جنسیت افراد و...

در این فصل می‌خواهیم دو بخش از علم آمار یعنی سازماندهی و نمایش، تحلیل و تفسیر داده‌ها که به آمار توصیفی معروف است را بررسی کنیم.

توصیف و نمایش داده‌ها:

بعد از گردآوری داده‌ها، به تنظیم، رده‌بندی و خلاصه کردن آن‌ها می‌پردازیم. برای این منظور می‌توان از روش‌های زیر استفاده نمود:

الف) تنظیم و طبقه‌بندی داده‌ها در یک جدول به نام جدول فراوانی

ب) رسم کردن نمودارها براساس مقادیر جدول فراوانی و نوع متغیر

فراوانی یک داده: تعداد دفعاتی که هر داده مشاهده شده است را فراوانی آن داده می‌نامیم.

فراوانی داده شماره i را با f_i نمایش می‌دهیم. مجموع فراوانی داده‌ها برابر تعداد داده‌های نمونه یعنی: $n = f_1 + f_2 + \dots + f_k$.

فراوانی نسبی یک داده: با تقسیم فراوانی هر داده بر تعداد کل داده‌ها، فراوانی نسبی آن داده بدست می‌آید.

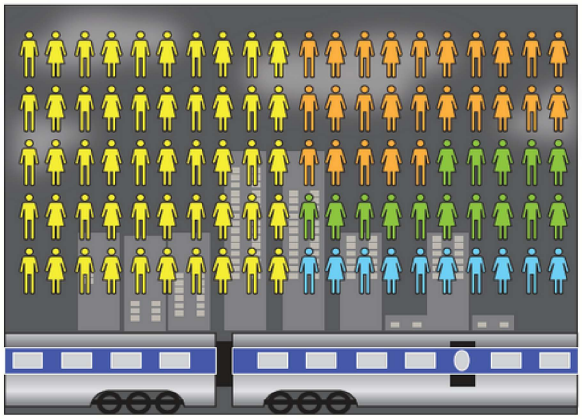
فراوانی نسبی داده شماره i را با f_i نمایش می‌دهیم. پس $f_i = \frac{f_i}{n}$.

واضح است که جمع فراوانی‌های نسبی یک می‌شود. اگر فراوانی نسبی را در ۱۰۰ ضرب کنیم درصد فراوانی نسبی آن داده بدست می‌آید. جمع فراوانی‌های نسبی برابر ۱۰۰ است.

مثال: در مورد اینکه مسافران یک قطار در طول سفر چگونه از وقت خود استفاده می‌کنند، تحقیقی صورت گرفته است و نتایج زیر به دست آمده است.

در شکل ت، تعداد مسافران یک قطار به عنوان متغیر گسسته را ملاحظه می‌کنید.

- افرادی که با رنگ زرد مشخص شده‌اند، مسافرانی‌اند که در قطار استراحت می‌کنند.
- افرادی که با رنگ نارنجی مشخص شده‌اند، مسافرانی‌اند که در قطار با تلفن همراه خود بازی می‌کنند.
- افرادی که با رنگ سبز مشخص شده‌اند، مسافرانی‌اند که در قطار مطالعه می‌کنند.
- افرادی که با رنگ آبی مشخص شده‌اند، مسافرانی‌اند که در قطار غذا می‌خورند.



شکل ت

جدول فراوانی مربوط به فراوانی تعداد مسافران را کامل کنید.

مسافران قطار	شماره مسافران	فراوانی مسافران	فراوانی نسبی مسافران
مسافرانی که استراحت می‌کنند	۱	۵۰	۰/۵۰
مسافرانی که با تلفن همراه خود بازی می‌کنند	۲	۲۵	۰/۲۵
مسافرانی که مطالعه می‌کنند	۳	۱۵	۰/۱۵
مسافرانی که غذا می‌خورند	۴	۱۰	۰/۱۰
تعداد کلّ مسافران		۱۰۰	۱

مثال: جدول فراوانی نسبی ۲۰۰ داده به صورت زیر است. فراوانی داده شماره ۴ را محاسبه کنید.

شماره داده	۱	۲	۳	۴	۵
فراوانی نسبی	۰/۱	۰/۱۸	۰/۲۴	۹۸	۰/۱۸

$$0.11 + 0.18 + 0.24 + x + 0.18 = 1 \Rightarrow x = 1 - 0.71 = 0.14$$

$$0.13 = \frac{f_4}{200} \Rightarrow f_4 = 60$$

دسته بندی داده ها :

فرض کنید کوچکترین داده a و بزرگترین داده b باشد. دامنه تغییرات داده ها $R = b - a$ تعریف می‌شود.

اگر بخواهیم این دامنه را به k دسته تقسیم کنیم طول هر دسته $c = \frac{R}{k}$ خواهد بود. حال بازه ها را از a شروع می‌کنیم و k واحد k واحد جلو می‌رویم. اینرا بازه ها را بسته و انتهای آنها را باز در نظر می‌گیریم به جز دسته آخر که هر دو بسته می‌شود. اگر بازه‌ای به صورت $[a_1, a_2)$ باشد به a_1 کران پائین دسته و به a_2 کران بالای دسته می‌گوییم.

مثال : اعداد زیر را به ۴ دسته طبقه بندی کنید سپس جدول فراوانی آن را رسم کنید.

$$2, 4, 4, 2, 5, 2, 5, 7, 6, 6, 7, 7, 4, 17, 16, 11$$

$R = 17 - 2 = 15$ و $c = \frac{15}{4} = 5$

دسته	$[2, 7)$	$[7, 12)$	$[12, 17]$
فراوانی	10	4	2

انواع نمودارها: هیستوگرام

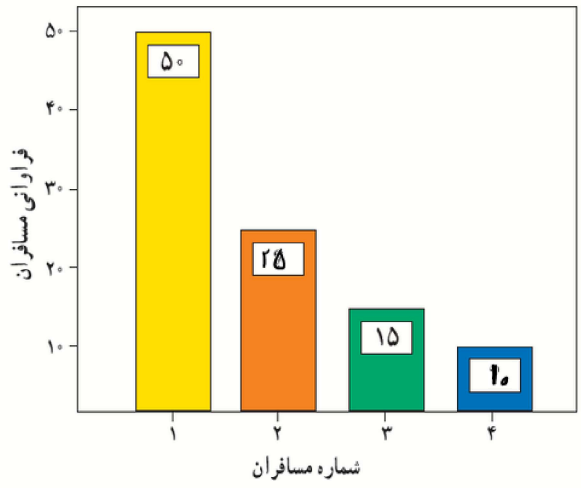
۱) نمودار میله‌ای ۲) نمودار دایره‌ای ۳) نمودار بافت نقاط (متقطعی) ۴) نمودار چندمتغیره

نکته: نمودار میله‌ای و دایره‌ای برای متغیرهای کیفی و کمی گسسته مناسب‌اند. نمودار بافت نقاط و چندمتغیره فقط مخصوص متغیرهای کمی پیوسته‌اند.

۱) نمودار میله‌ای: روی محور طول‌ها، شماره داده یا جزء داده می‌نویسیم و روی محور عرض‌ها، فراوانی یا فراوانی نسبی.

ارتفاع میله‌ها متناسب با فراوانی یا فراوانی نسبی دسته‌هاست.

مثال: برای مثال مسافران تعداد نمودار میله‌ای فراوانی را رسم کنید.



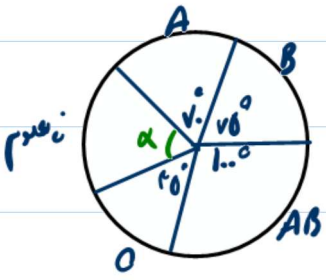
۲) نمودار دایره‌ای: دو روش برای رسم نمودار دایره‌ای داریم:

روش اول: دایره را به ۱۰۰ قسمت مساوی تقسیم می‌کنیم. هر قسمت معادل ۱ درصد است. با توجه به درصد فراوانی‌های بین داده‌ها را رسم می‌کنیم.

روش دوم: هر قسمت یک زاویه مرکزی دارد. اگر α_i زاویه مربوط به داده iام باشد داریم:

$$\alpha_i = f_i \times 360^\circ$$

مثال: نمودار دایره‌ای زیر مربوط به گروه‌های خونی دانش‌آموزان یک کلاس است. گروه خونی ۴۲ نفر از آن‌ها مشخص شده است. چند نفر از آن‌ها دارای نوع خون AB هستند؟



$$\alpha = 360^\circ - (70 + 75 + 100 + 45) = 70^\circ$$

$$70 = \frac{42}{n} \times 360 \Rightarrow 70n = 42 \times 360 \Rightarrow n = \frac{42 \times 360}{70}$$

$$\Rightarrow n = 144$$

$$\alpha_B = 75 = \frac{f_B}{144} \times 360 \Rightarrow f_B = \frac{144 \times 75}{360} = 30$$

۳) نمودار بافت نگاشت (متعلیی)

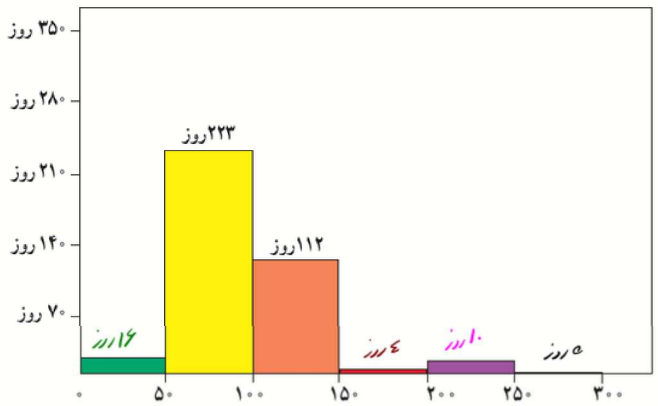
دسته‌ها یا بازه‌ها را روی محور xها و فراوانی یا فراوانی نسبی را روی محور yها مشخص می‌کنیم. نمودار از مقداری مستطیل به هم پیوسته تشکیل می‌شود.

مثال :

میزان شاخص کیفیت هوا در شهر تهران برای تمام روزهای سال ۱۳۹۳ در جدول زیر گزارش شده است. این جدول را کامل کنید :

وضعیت هوا	شاخص کیفیت هوا	فراوانی	فراوانی نسبی
پاک	$0 \le AQI \le 50$	۱۶	۰/۰۴
سالم	$50 < AQI \le 100$	۲۲۳	۰/۶۱
ناسالم برای گروه‌های حساس	$100 < AQI \le 150$	۱۱۲	۰/۳۱
ناسالم	$150 < AQI \le 200$	۴	۰/۰۱
بسیار ناسالم	$200 < AQI \le 250$	۱۰	۰/۰۳
خطرناک	$250 < AQI \le 300$	۰	۰
تعداد کلّ روزهای یک سال		۳۶۵	۱

نمودار مربوط به فراوانی تعداد روزها براساس وضعیت آلودگی هوا را کامل کنید؟



نکته : مجموع مساحت مستطیل ها در نمودار بافت نگاشت فراوانی nc و در فراوانی نسبی برابر c است.

۴) نمودار چندبر فراوانی :

اگر در نمودار مستطیلی وسط عرض بازه‌ی مستطیل‌ها را به هم وصل کنیم و روی محور xها به اندازه نصف طول بازه‌ها جلو و عقب رفته و به آن نقاط وصل کنیم نمودار چندبر فراوانی به دست می‌آید. در واقع نقاط شکستگی با مختصات (x_i, f_i) هستند که x_i وسط دسته‌هاست.

شاخص های مرکزی و پراکندگی:

(الف) معیارهای گرایش به مرکز (شاخص های مرکزی)

معیارهایی که محل تمرکز داده ها را نشان می دهد. سه معیار گرایش به مرکز داریم: میانگین - میانه - مد

۱) میانگین: فرض کنید داده ها $x_1, x_2, \dots, x_n$ باشند. جمع داده ها تقسیم بر تعداد آن را میانگین آن را تعریف می شود که با $\bar{x}$ نشان داده می شود. پس:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

نماد سیگما (Σ): برای خلاصه نویسی جمع از نماد Σ استفاده می کنیم به این صورت که:

$$\sum_{i=a}^{b} (\text{عبارت برحسب } i)$$

$i=a$ ← از این

مثلاً

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

پس فرمول میانگین را می توان به صورت $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ هم نوشت.

مثال: میانگین داده های $x_1, x_2, \dots, x_{10}$ برابر ۲۴ است. میانگین داده های $x_1, x_2, \dots, x_{10}, 14, 26$ را بدست آورید.

$$24 = \frac{\sum_{i=1}^{10} x_i + 14 + 26}{12} \Rightarrow \sum_{i=1}^{10} x_i + 40 = 288 \Rightarrow \sum_{i=1}^{10} x_i = 248$$

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i + 19 + 10}{12} = \frac{242 + 29}{12} = \frac{271}{12} = 22{,}58$$

نکته: اگر میانگین n داده $\bar{x}$ و میانگین m داده $\bar{y}$ باشد میانگین همه آن‌ها با هم $\frac{n\bar{x} + m\bar{y}}{n+m}$ می‌شود.

نکته: فرض کنید a داده در عدد b ضرب و میانگین داده‌های $x_1, x_2, \ldots, x_n$ برابر $\bar{x}$ باشد و داده‌های جدید y_i به صورت $y_i = ax_i + b$ تعریف شوند. در این صورت میانگین $y_1, \ldots, y_n$ برابر $a\bar{x} + b$ می‌شود.

نکته: اگر داده‌ای بیشتر از میانگین به داده‌ها اضافه شود میانگین زیاد و اگر کمتر از میانگین اضافه کنیم، میانگین کم می‌شود و اگر داده‌ای مساوی میانگین حذف یا اضافه کنیم میانگین تغییر نمی‌کند.

انحراف از میانگین: $(x_i - \bar{x})$ را انحراف داده iام از میانگین می‌نامیم. می‌توان ثابت کرد

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

می‌شود.

تذکر: اگر داده‌ها تشکیل دنباله حسابی بدهند میانگین آن‌ها $\frac{\text{آخری} + \text{اولی}}{2}$ می‌شود.

مثال: میانگین n داده آماری برابر ۱۲ و میانگین ۱۲ داده دیگر برابر ۱۰ است. دوسوده را با هم ارزیابی کنیم میانگین همگی داده‌ها ۱۱،۵ می‌شود. n را بیابید.

$$\frac{n\bar{x}+m\bar{y}}{n+m}=11{,}5 \Rightarrow \frac{12n+12\times 10}{n+12}=11{,}5 \Rightarrow 11{,}5n+138=12n+120$$

$$\Rightarrow 0{,}5n=18 \Rightarrow n=\frac{18}{0{,}5}=36$$

مثال: اگر میانگین داده‌های $2x_1+5$، $2x_2+5$، $2x_3+5$ برابر ۲۰ باشد میانگین $1+\frac{x_1}{3}$، $1+\frac{x_2}{3}$ و $1+\frac{x_3}{3}$ را بیابید.

$$2x_i+5 \xrightarrow{-5} 2x_i \xrightarrow{\times\frac{1}{2}} x_i \xrightarrow{+1} \frac{x_i}{3}+1$$

$$\bar{X}=20 \xrightarrow{-5} \bar{X}=15 \xrightarrow{\times\frac{1}{2}} \bar{X}=\frac{15}{6}=\frac{5}{2} \xrightarrow{+1} \bar{X}=\frac{5}{3}+1=\frac{7}{3}=4{,}5$$

نکته: برای محاسبه سریع میانگین اعداد بزرگ، از همه اعداد یکبار و یکجا کم کنید به میانگین داده‌ها نزدیک باشد، کم می‌کنیم. میانگین اعداد جدید را می‌یابیم و در آخر عددی را که کم کردیم به میانگین اضافه می‌کنیم.

مثال: میانگین اعداد ۹۰، ۹۲، ۹۷، ۱۰۰، ۱۰۵، ۱۱۰ و ۱۲۰ را بیابید.

۹۰، ۹۲، ۹۷، ۱۰۰، ۱۰۵، ۱۱۰، ۱۲۰

۱۰۵ را از همه کم می‌کنیم:

-۱۵، -۱۳، -۸، -۵، ۰، ۵، ۱۵

$$\bar{X}_{\text{جدید}}=\frac{-15+\cdots+15}{7}=-3 \rightarrow \bar{X}_{\text{قدیم}}=-3+105=102$$

میانگین موزون (وزن‌دار)

اگر به هر داده x_i وزنی (این وزن‌ها می‌توانند ضریب ارزش یا فراوانی و ... باشند) مانند w_i نسبت دهیم، میانگین وزن‌دار را با $\bar{x}_w$ نمایش می‌دهیم و برابر $\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$ است.

تذکر: اگر وزن‌ها، فراوانی‌های نسبی باشند چون مجموع فراوانی‌های نسبی برابر یک است پس داریم

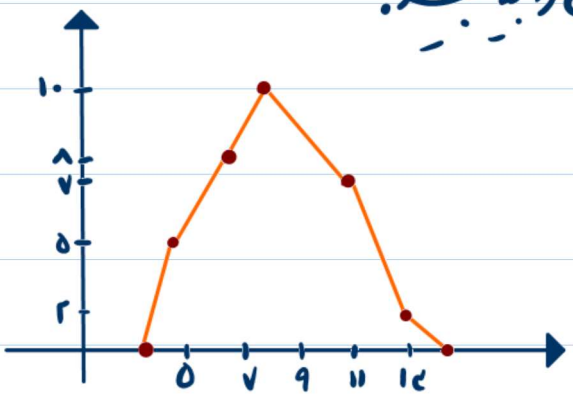
$$\bar{x}_w = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{1}$$

مثال: کارنامه‌ی یک دانش‌آموز به صورت زیر است. معدل او را حساب کنید.

درس	زبان	عربی	دینی	ادبیات	شیمی	هندسه	ریاضی
نمره	۱۲	۱۶	۱۹	۱۸	۱۰٫۵	۱۴	۱۹
ضریب درس	۱	۲	۴	۲	۱	۳	۴

$$\bar{x}_w = \frac{1\times 12 + 2\times 16 + 4\times 19 + 2\times 18 + 1\times 10{,}5 + 3\times 14 + 4\times 19}{1+2+4+2+1+3+4} = \frac{284{,}5}{17} = 16{,}73$$

مثال: میانگین داده‌ها با نمودار چند ضابطه‌ای مقابل را بیابید.



$$\bar{x}_w = \frac{\sum f_i x_i}{\sum f_i} = \frac{(0\times 5) + (8\times 7) + (10\times 9) + (7\times 11) + (2\times 14)}{5+8+10+7+2} = \frac{274}{32} = 8{,}56$$

مثال: تعداد عمل‌های جراحی یک پزشک در یک روز مطابق جدول زیر است. این پزشک به طور متوسط در یک روز چند عمل جراحی انجام می‌دهد؟

تعداد عمل‌ها در روز	۱	۲	۴	۵
فراوانی نسبی	٪۱۴	٪۱۲	٪n	٪۴۵

$\bar{x}_w = \frac{1\times ٪۱۴ + ۲\times ٪۱۲ + ۴\times ٪n + ۵\times ٪۴۵}{۱} = ۲,۶$

$٪۱۴ + ٪۱۲ + ٪n + ٪۴۵ = ۱ \Rightarrow n = ٪۱$

۲) **میانه:** عددی که بعد از مرتب کردن داده‌ها از کوچک به بزرگ در وسط قرار می‌گیرد.

اگر n داده داشته باشیم دو حالت رخ می‌دهد:

الف) n فرد باشد: فقط یک عدد در وسط قرار می‌گیرد که همان میانه خواهد بود. شماره این داده $\frac{n+1}{2}$ است.

ب) n زوج باشد: دو داده در وسط قرار می‌گیرند که میانگین این دو داده میانه است. شماره این دو داده $\frac{n}{2}$ و $\frac{n}{2}+1$ است.

مثال: میانه داده‌های زیر را بیابید.

۱) ۱,۲,۴,۸,۱۵,۱۰,۱۱,۶,۱۲,۱۰,۴,۱۷,۲,۱۹

$n=۱۴$: داده $\frac{۱۴+۱}{۲}$ یعنی ۸ میانه است.

۱,۱,۲,۲,۴,۴,۶,۸,۱۰,۱۰,۱۲,۱۵,۱۷,۱۹

میانه

۲) ۱۲,۵,۱۵,۴,۶,۹,۱۶,۲۰

$n=۸$: میانگین دو داده وسطی داده چهارم و پنجم میانه است.

$\frac{۸}{۲}$ $\frac{۸}{۲}+۱$

۳,۵,۶,۹,۱۲,۱۵,۱۶,۲۰

میانه $= \frac{۹+۱۲}{۲} = ۱۰,۵$

مثال: به ازای کدام مقدار a میانه داده‌های $3, 4, 15, 2, 5, 11, 21$ برابر 8.5 است؟

۱) ۵ ✓ ۲) ۶ ۳) ۷ ۴) ۸

چون تعداد داده‌ها ۶ تاست پس میانگین داده‌های سوم و چهارم میانه خواهد بود.

$3, 4, 5, 11, 15, 21$ $\quad \frac{a+11}{2} = 8.5 \rightarrow a+11 = 17 \Rightarrow a = 6$

مثال: میانگین یازده داده برابر ۶ و میانه این داده‌ها برابر ۷ است. حداکثر از این داده‌ها عبارت‌اند از:

$9, 8, 8, 8, 4, 4, 3, 2, 2, 1$

تفاضل دو داده‌ی دهم کدام است؟

۱) ۵ ۲) ۶ ۳) ۷ ۴) ۸ ✓

چون $\bar{x} = 6$ و $n = 11$ پس: $\frac{\sum x_i}{11} = 6$ در نتیجه $\sum x_i = 66$. حداکثر دو داده‌ی دهم را x و y در نظر بگیریم:

$x+y+1+2+2+4+4+8+8+8+9 = 66 \Rightarrow x+y = 22$

از طرفی چون تعداد داده‌ها $n = 11$ است پس داده‌ی ششم میانه است که برابر ۷ است. پس حداکثر یکی از x یا y باشد و چون $x+y = 22$ پس دیگری ۱۵ یا کمتر و تفاضل دو به ۸ است.

$\frac{11+1}{2} = 6$

چارک اول و سوم: میانه داده‌های قبل از میانه را چارک اول Q_1 و میانه داده‌های بعد از میانه را چارک سوم Q_3 می‌نامیم. به عبارت دیگر به میانه $\left[\frac{n}{4}\right]$ داده‌ی اشاره‌ی چارک اول و به میانه $\left[\frac{3n}{4}\right]$ داده‌ی اشاره‌ی چارک سوم می‌گوییم.

مثال: چارک اول و سوم را برای داده‌های زیر بیابید.

۱) $1, 2, \dots, 9$: $1, 2, 4, 5, 6, 7, 8, 9$ $\quad n=9$

$Q_1 = \frac{2+4}{2} = 3$ $\quad Q_3 = \frac{7+8}{2} = 7.5$

Q_2 یا میانه: 5

۲) $5, 6, 4, 3, 7, 8, 10, 9$ $\quad n=8$

$3, 3, 5, 6, 7, 8, 9, 10$

$Q_1 = 4$ $\quad Q_3 = 8.5$

$Q_2 = \frac{6+7}{2} = 6.5$

مثال: مجموع چارک اول و سوم داده‌های جدول مقابل کدام است؟

داده	فراوانی
۷	۷
۱۰	۵
۱۱	۶
۱۵	۵
۱۷	۶

۱) ۲۲ ۲) ۲۴ ۳) ۲۴٫۵ ✓ ۴) ۲۶٫۵

$n = \sum f_i = 7+5+6+5+6 = 29$

Q_3: داده‌ی $\frac{29+1}{2} = 15$ام

پس چارک اول میانه چهارده داده‌ی اول یعنی میانگین داده‌های هفتم و هشتم است $Q_1 = \frac{7+10}{2} = 8.5$

چارک سوم میانه چهارده داده‌ی دوم یعنی میانگین داده‌های بیست و دوم و بیست و سوم است $Q_3 = \frac{15+15}{2} = 15$.

$Q_1 + Q_3 = 8.5 + 15 = 23.5$

۳) مد: به داده‌ای که بیشترین فراوانی را داشته باشد، مد یا نما می‌گوییم.

تذکر: اگر همه داده‌ها یک فراوانی داشته باشند، این داده‌ها مد ندارند و اگر دو یا چند (البته نه همه) داده دارای بیشترین فراوانی باشند این داده‌ها بیش از یک مد دارند.

مثال: در مورد داده‌های زیر کدام گزینه درست است؟

۱, ۱, ۲, ۲, ۲, ۳, ۳, ۳, ۳, ۴, ۴, ۴, ۴, ۴, ۵, ۶, ۷

۱) میانگین > مد = میانه
۲) میانگین < مد = میانه
۳) میانگین > مد > میانه
۴) مد > میانگین > میانه ✓

$n=17$ پس داده 9 ام $= \frac{17+1}{2}$ یعنی ۴ میانه است. ۴ دارای بیشترین فراوانی است پس مد است.

$\bar{x} = \frac{1+1+2+\cdots+7}{17} = \frac{58}{17}$ $\quad 4 > \frac{58}{17} > 3$ (۴ به ۳ نزدیک‌تر است)

تذکر: در تفسیر و تحلیل معانی آماری، ارتباط نزدیک شاخص‌های مرکزی با مرکز کار است. باید هر سه معیار میانگین، میانه و مد حساب شود و برای یک هدف مورد نظر، معیار مناسب انتخاب و از آن برای انجام تفسیر، قضاوت و پیش‌بینی استفاده شود.

تعریف داده دور افتاده: مشاهده‌ای که تفاوت بسیار زیادی با سایر مشاهدات مجموعه داده‌ها داشته باشد.

نکته: داده دور افتاده میانگین داده‌ها را تحت تأثیر قرار می‌دهد و بی تأثیری بر میانه و مد ندارد.

پس اگر داده دور افتاده داشته باشیم، میانگین معیار مناسبی برای سنجش تمرکز داده‌ها نیست.

ب) معیارهای پراکندگی:

معیارهایی که دوری یا نزدیکی داده‌ها از میانگین را بررسی می‌کنند. سه معیار هم برای سنجش میزان پراکندگی

از میانگین عبارتند از: انحراف معیار - واریانس - ضریب تغییرات

۱) انحراف معیار و واریانس:

اگر مجموع مربعات انحراف از میانگین را بر تعداد داده‌ها تقسیم کنیم واریانس بدست می‌آید که آن را با σ^2

نمایش می‌دهیم یعنی:

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

جذر واریانس را انحراف معیار می‌نامیم و با σ نشان می‌دهیم یعنی:

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

نکته: اگر واریانس و انحراف معیار داده‌ها کوچک باشد به این معناست که پراکندگی حول میانگین کم و در نتیجه

داده‌ها به هم نزدیک‌ترند واریانس و انحراف معیار داده‌ها عددی کوچک‌تر باشد یعنی پراکندگی داده‌ها حول میانگین شان زیاد و در نتیجه داده‌ها از هم دورند.

مثال: واریانس و انحراف معیار داده‌های زیر را بدست آورید.

۱۰, ۱۵, ۱۸, ۱۴, ۱۴, ۲۰

$$\bar{x} = \frac{10+15+18+14+14+20}{6} = \frac{90}{6} = 15$$

$$\sigma^2 = \frac{\sum_{i=1}^{6}(x_i-15)^2}{6} = \frac{(10-15)^2+(15-15)^2+(18-15)^2+(14-15)^2+(14-15)^2+(20-15)^2}{6}$$

$$= \frac{25+0+9+1+1+25}{6} = \frac{61}{6} = \frac{32}{4}$$

$$\sigma = \sqrt{\frac{32}{4}} = \frac{\sqrt{32}\times\sqrt{4}}{\sqrt{4}\times\sqrt{4}} = \frac{4\sqrt{6}}{4}$$

واریانس و انحراف داده‌های وزن دار:

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

مثال: انحراف معیار داده‌های جدول زیر را بیابید.

x_i	1	2	4	5
w_i	1	2	2	1

$$\bar{x} = \frac{1\times1+2\times2+4\times2+5\times1}{1+2+2+1} = \frac{18}{6} = 4$$

$$\sigma^2 = \frac{\sum_{i=1}^{4} w_i (x_i - 4)^2}{\sum_{i=1}^{4} w_i} = \frac{1(1-4)^2 + 2(2-4)^2 + 2(4-4)^2 + 1(8-4)^2}{1+2+2+1} = \frac{12}{6} = 2$$

$$\sigma = \sqrt{2}$$

چند نکته :

۱) اگر n داده تشکیل دنباله حسابی با قدرنسبت d بدهند واریانس آن‌ها $\sigma^2 = \frac{n^2-1}{12} d^2$ می‌شود.

۲) اگر به همه داده‌ها عدد ثابتی اضافه یا کم کنیم واریانس و انحراف معیار تغییر نمی‌کند. از این نکته وقتی داده‌ها بزرگ باشند می‌توان استفاده کرد.

۳) اگر واریانس یا انحراف معیار صفر باشد آنگاه همه داده‌ها مساوی با میانگین (و در نتیجه میانه و هم) هستند و برعکس.

۴) اگر داده‌ها را در k ضرب کنیم انحراف معیار $|k|$ برابر و واریانس k^2 برابر می‌شود.

۵) فرمول واریانس را می‌توان به صورت : $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2$ هم نوشت.

۶) واحد اندازه‌گیری واریانس، مربع واحد داده‌های اولیه است و واحد اندازه‌گیری انحراف معیارها واحد اندازه‌گیری داده‌های اولیه است.

مثال: مجموع مربعات ده داده برابر ۲۲۵۰ و میانگین آنها ۱۰/۲ است. انحراف معیار کدام است؟

۱) ۴ ۲) ۴ ۳) ۵ ۴) ۶

$$\sigma^2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - \bar{x}^2 = \frac{2250}{10} - (10.2)^2 = 225 - 200 = 25 \Rightarrow \sigma = 5$$

مثال: اگر انحراف معیار داده‌های ۲، ۴، ۸، ۱۰، ۱۲ برابر a باشد، واریانس داده‌های ۴، ۵، ۷، ۸، ۹ کدام است؟

۱) $\frac{a^2}{4}$ ۲) $\frac{a}{2}$ ۳) $\frac{a^2}{4}$ ۴) $\frac{a^2}{2}$

x_i : ۲، ۴، ۸، ۱۰، ۱۲ $\quad y_i$: ۴، ۵، ۷، ۸، ۹ $\Rightarrow y_i = \frac{1}{2}x_i + 4$

$$\sigma_x = a \Rightarrow \sigma_x^2 = a^2, \quad \sigma_y^2 = (\tfrac{1}{2})^2 \sigma_x^2 = \frac{1}{4}a^2 = \frac{a^2}{4}$$

مثال: واریانس داده‌های $x_1, x_2, \ldots, x_{10}$ برابر صفر است. انحراف معیار داده‌های $x_1+2, x_2-2, x_3, \ldots, x_{10}$ کدام است؟

۱) ۲ ۲) ۲۲ ۳) ۸ ۴) $\sqrt{8}$

$\sigma_x^2 = 0 \Rightarrow x_1 = x_2 = \cdots = x_{10} = 10$ $\Rightarrow$ داده‌های جدید: ۱۴، ۸، ۱۰، ۱۲، ۱۶ $\rightarrow$ ۱۶، ۱۴، ۱۲، ۱۰، ۸

دنباله‌های با قدرت ۲ :

$$\sigma^2 = \frac{n^2-1}{12} \times d^2 = \frac{5^2-1}{12} \times 2^2 = \frac{24}{12} \times 4 = 2 \times 4 = 8$$

$$\sigma = \sqrt{8}$$

۲) ضریب تغییرات (CV)

ی مفهوم دیگری برای سنجش پراکندگی داده‌ها تعریف کنیم که در آن هم پراکندگی داده‌ها و هم میانگین نقش داشته باشه. مثلاً در یک تیم والیبال هر چقدر میانگین قد بازیکنان بلندتر و پراکندگی قدشون کمتر باشه، از نظر طول قد مطلوب‌تره.

این معیار ضریب تغییرات نامیده می‌شود که میزان پراکندگی داده‌ها را برای یک واحد از میانگین نشان می‌دهد.

تعریف: تقسیم انحراف معیار بر میانگین را ضریب تغییرات می‌گوییم: $CV = \frac{\sigma}{\bar{x}}$

چند نکته:

۱) ضریب تغییرات به واحد اندازه‌گیری بستگی ندارد. (مثلاً اگر داده‌ها بر حسب cm اندازه‌گیری شده باشند و ما آن‌ها را به m تبدیل و بعد CV را حساب کنیم.) بنابراین برای مقایسه پراکندگی داده‌های دو واحد متفاوت ندارند بهترین شاخص پراکندگی CV است که بدون واحد است. (مثلاً مقایسه پراکندگی قد با وزن دانش‌آموزان)

۲) هر چقدر ضریب تغییرات داده‌ها کمتر باشد، میزان پراکندگی داده‌ها کمتر است.

۳) اگر داده‌ها را k برابر کنیم، CV تغییر نمی‌کند.

۴) اگر به همه داده‌ها عدد k ($k \neq 0$) را اضافه کنیم، ضریب تغییرات به صورت $CV = \frac{\sigma}{\bar{x}+k}$ تغییر می‌کند و مقدار CV کم می‌شود.

مثال: اگر میانگین و انحراف معیار داده‌های $x_1, x_2, \dots, x_{10}$ به ترتیب ۴ و ۱ باشد، ضریب تغییرات داده‌های $3+x_1, 3+x_2, \dots, 3+x_{10}$ کدام است؟ ۱) ۱۲۵٪ ۲) ۱۲۵٪ ۳) ۱۲۵٪ ۴) ۱

$$(CV)_y = \frac{\sigma_y}{\bar{y}} = \frac{1}{4+3} = \frac{1}{8} = 0.125$$

مثال: اگر انحراف معیار داده‌های $x_1, \dots, x_n$ برابر ۲ و ضریب تغییرات آنها برابر ۵ باشد، ضریب تغییرات داده‌های $5x_1+2, \dots, 5x_n+2$ کدام است؟

۱) $\frac{5}{5+5}$ ۲) $\frac{5+5}{5}$ ۳) $\frac{5}{5+25}$ ۴) $\frac{5+25}{5}$

$$(CV)_x = \frac{\sigma_x}{\bar{x}} \Rightarrow \bar{x} = \frac{\sigma_x}{(CV)_x} = \frac{2}{5} \qquad \bar{y} = 5\bar{x} + 2 = \frac{10+25}{5}$$

$$\sigma_y = 2 \times 5 = 10 \Rightarrow (CV)_y = \frac{10}{\frac{10+25}{5}} = \frac{10}{10+25} \xrightarrow{\div 5} (CV)_y = \frac{5}{5+5}$$

مثال: اگر میانگین و ضریب تغییرات اندازه اضلاع مربع‌های ۱ و ۲ به ترتیب ۱۵ و ۲۰٪ باشد، میانگین مساحت این مربع‌ها کدام است؟

۱) ۲۴۲ ۲) ۲۲۹ ۳) ۲۲۷ ۴) ۲۲۴

$$CV = \frac{\sigma}{\bar{x}} \Rightarrow 0.2 = \frac{\sigma}{15} \Rightarrow \sigma = 3 \Rightarrow \sigma^2 = 9$$

اگر ضلع i ام باشد مساحت x_i^2 می‌شود. میانگین مساحت‌ها همان میانگین مربعات داده‌ها، از فرمول زیر واریانس داریم:

$$\sigma^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 \Rightarrow 9 = \frac{\sum x_i^2}{n} - (15)^2 \Rightarrow \frac{\sum x_i^2}{n} = 234$$

دامنه میان چارکی: به حاصل $Q_3 - Q_1$ دامنه میان چارکی گوییم و آن را با IQR نشان می‌دهیم.

مثال: در نمودار جعبه‌ای ۴۲ داده آماری، مینیمم داده‌های سمت راست جعبه برابر ۱۱، داخل جعبه ۷ و سمت چپ ۵ است. مینیمم کل ۴۲ داده کدام است؟ ۱) ۵ ✓ ۲) ۷٫۵ ۳) ۹٫۵ ۴) ۱۱

از ۴۲ داده ۸ تا سمت چپ ۸ تا راست و ۱۶ تا داخل جعبه است.

$$\bar{X}_L = \frac{\sum x_i}{8} \Rightarrow 5 = \frac{\sum x_i}{8} \Rightarrow \sum x_i = 40$$

$$\bar{X}_R = \frac{\sum x_i}{8} \Rightarrow 11 = \frac{\sum x_i}{8} \Rightarrow \sum x_i = 88$$

$$\bar{X}_m = \frac{\sum x_i}{16} \Rightarrow 7 = \frac{\sum x_i}{16} \Rightarrow \sum x_i = 112$$

$$\bar{X} = \frac{40 + 88 + 112}{42} = \frac{240}{42} = 7/5$$

نمودار جعبه‌ای:

برای رسم نمودار جعبه‌ای ابتدا بزرگ‌ترین و کوچک‌ترین داده، میانه و چارک اول و سوم را به دست می‌آوریم و روی یک محور مشخص می‌کنیم. از Q_1 تا Q_3 یک جعبه رسم می‌کنیم.

نکته: در نمودار جعبه‌ای تقریباً یک چهارم داده‌ها سمت چپ، یک چهارم سمت راست و نیمی از داده‌ها داخل جعبه قرار دارند.

مثال: برای داده‌های زیر نمودار جعبه‌ای رسم کنید. سپس داده‌های داخل جعبه را بیابید.

۱۴, ۱۲, ۱۲, ۲۱, ۱۷, ۱۲, ۱۱, ۱۰, ۹, ۱۷, ۱۶, ۲۰, ۷, ۱۸

۷, ۹, ۱۰, ۱۱, ۱۲, ۱۲, ۱۴, ۱۶, ۱۷, ۱۷, ۱۸, ۲۰, ۲۱ $\quad n=14$

$Q_1=10.5$ $\quad Q_2$ $\quad Q_3=18.5$



داده‌های داخل جعبه : ۱۱, ۱۲, ۱۲, ۱۴, ۱۶, ۱۷, ۱۷

برای یافتن ساده‌ترین داده‌ها ۱۴ را از همه داده‌ها کم می‌کنیم:

۴, ۴, ۳, ۰, −۱, −۱, −۲

$$\bar{x}=\frac{-2+\cdots+4}{7}=\frac{7}{7}=1 \Rightarrow \sigma^2=\frac{4+1+1+9+16+16}{7}-1^2=\frac{47}{7}-1=6.71-1=5.71$$

مثال: کنکوری ۱۴۰۱

در یک دسته از اعداد، چهار عدد فرد یک رقمی در اختیار داریم. دوتای آنها را با کوچک‌ترین عدد زوج بعد از خود و دوتای دیگر را با بزرگ‌ترین عدد زوج قبل از خود جایگزین می‌کنیم به‌طوری‌که اعداد در دسته دوم (دسته جدید) تک رقمی و غیرتکراری باشند. نسبت ضریب تغییرات دسته اول به دسته دوم کدام است؟

۱) $0.2\sqrt{5}$ ۲) $0.4\sqrt{10}$ ۳) $\sqrt{2}$ ✓ ۴) ۱٫۲۵

x_i : ۱, ۳, ۷, ۹ $\quad : \bar{x}=\frac{1+3+7+9}{4}=5$

y_i : ۲, ۴, ۶, ۸ $\quad \bar{y}=\frac{2+4+6+8}{4}=5$

$$\sigma_x^2=\frac{(-4)^2+(-2)^2+2^2+4^2}{4}=\frac{16+4+4+16}{4}=\frac{40}{4}=10 \rightarrow \sigma_x=\sqrt{10}$$

$\sigma_y^2 = \frac{(-2)^2+(-1)^2+1^2+2^2}{4} = \frac{9+1+1+9}{4} = 5 \to \sigma_y = \sqrt{5} \Rightarrow \frac{c\sigma_x}{c\sigma_y} = \frac{\frac{12}{5}}{\frac{15}{5}} = \frac{12}{15}$

مثال: کنکور تیر ۱۴۰۱

میانگین شش داده آماری، عددی طبیعی است و توان دوم انحراف از میانگین این داده‌ها، به‌صورت $1, b^2, 0, 9, a^2, 9$ است. اگر واریانس این داده‌ها برابر ۴ باشد، مقدار ab کدام است؟ $(a, b \in \mathbb{Z})$

۱) ۴- ۲) ۴ ۳) ۲ ۴) ۲- ✓

$4 = \frac{9+a^2+0+9+b^2+1}{6} \Rightarrow 24 = 19 + a^2 + b^2 \Rightarrow a^2+b^2 = 5 \begin{cases} a=\pm1, b=\pm2 \\ a=\pm2, b=\pm1 \end{cases}$

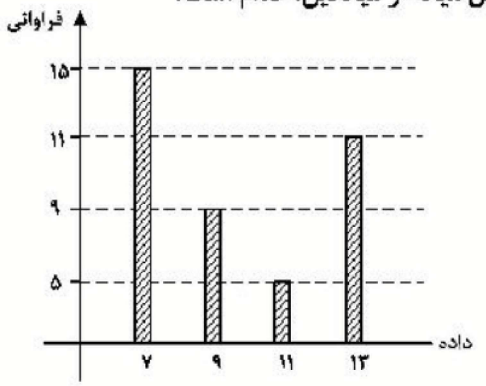
مجموع انحراف از میانگین‌ها باید ۰ باشد

$|x_i - \bar{x}|$: $\pm4, \pm1, 0, \pm4, \pm2, \pm1$

$a=-1, b=2 \Rightarrow ab=-2$ — شدنی‌ها: $4, -1, 0, -4, 2, -1$

مثال: کنکور ۹۹ خارج

با توجه به نمودار میله‌ای فراوانی داده‌های کمّی گسسته، تفاضل میانه از میانگین، کدام است؟



۱) ۰/۳
۲) ۰/۴
۳) ۰/۵
۴) ۰/۶ ✓

$\sum f_i = 8 \Rightarrow 10+11+9+10 = 40$ — میانه = میانگین دو داده وسط = $\frac{9+9}{2} = 9$

$\bar{x} = \frac{7\times10 + 9\times9 + 11\times8 + 13\times11}{40} = \frac{100+81+88+143}{40} = \frac{412}{40} = 9.2$

میانگین − میانه $= 9.2 - 9 = 0.2$