فصل هفتم: آمار و احتمال

۱. احتمال شرطی و پیشامدهای مستقل

۲. آمار توصیفی

درس ۱: احتمال شرطی و پیشامدهای مستقل

ما در عصر احتمال به سر می‌بریم ... در عصر قاطعیت تردید ...

$$P(A)=\frac{n(A)}{n(S)}=\frac{\text{تعداد حالت‌های مطلوب}}{\text{تعداد کل حالت‌ها}}$$

۱ احتمال شرطی:

گاهی رخ دادن یک پیشامد بر احتمال رخ دادن پیشامد دیگری اثر می‌کند. مسائلی که در آن محاسبهٔ احتمال یک پیشامد مانند A با توجه به آگاهی از رخ دادن پیشامد دیگری مانند B مطرح است را احتمال شرطی می‌نامیم و آن را با نماد $P(A|B)$ نمایش می‌دهیم که خوانده می‌شود:

"احتمال A به شرط B"

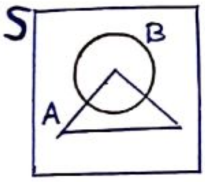
درواقع چون از B آگاهی داریم، B فضای نمونه‌ای محسوب شده و جایگزین S می‌شود، درنتیجه فضای نمونه کاهش یافته است.

مثال. در پرتاب یک تاس اگر بدانیم عدد رو شده زوج است، احتمال آنکه اول باشد چقدر است؟

فضای نمونه‌ای کاهشی $B=\{2,4,6\}$

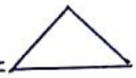
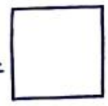
پیشامد اول بودن $A=\{2\}$

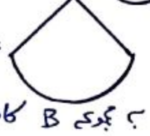
$$\rightarrow P(A|B)=\frac{1}{3}$$



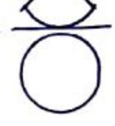
با فرمول احتمال شرطی:

مطابق شکل روبه‌رو

$n(A)=$  ، $n(B)=$  ، $n(S)=$ 

$n(A\cap B)=$ 

حال اگر فضای نمونه‌ای را به گونه‌ای B کاهش دهیم، احتمال وقوع A در B (احتمال A به شرط B) به صورت

$P(A|B)=$ 

محاسبه می‌شود. لذا با تقسیم صورت و مخرج کسر بر $n(S)$ داریم:

$$P(A|B)=\frac{n(A\cap B)}{n(B)} \Rightarrow \boxed{P(A|B)=\frac{P(A\cap B)}{P(B)}}$$

– شرط محاسبه‌پذیری $P(A|B)$ آن است که $P(B)\neq 0$.

مثال. کیسه‌ای شامل ۴ مهرهٔ سیاه، ۲ مهرهٔ سفید و ۳ مهرهٔ قرمز است. مهره‌ای از کیسه بیرون آورده و مشاهده شده که سیاه نیست، احتمال آنکه سفید باشد چقدر است؟

$$P(\text{سفید بودن} \mid \text{سیاه نبودن}) = \frac{P(\text{سیاه نبودن و سفید بودن})}{P(\text{سیاه نبودن})} = \frac{\overset{\text{۲ تا سفید}}{۲}}{\underbrace{۲}_{\text{۲ تا سفید}}+\underbrace{۳}_{\text{۳ تا قرمز}}} = \frac{۲}{۵} = ۴۰\,\%$$

✓ قانون جمع احتمالات:

$$\underbrace{P(A\cup B)}_{\text{حداقل یکی}} = P(A)+P(B)-P(A\cap B)$$

✓ دیدویژه!

۱. بارها شده دانش‌آموز بپرسد، از کجا بفهمیم که مسأله شرطی است یا نه؟ باید بگوییم در مسائل شرطی همواره کلماتی مانند: "می‌دانیم که"، "فرض کنید"، "اگر" می‌آید و در ادامهٔ آن خبری دربارهٔ آزمایش می‌دهند، سپس از ما می‌خواهند با توجه به آن خبر احتمال وقوع چیز دیگری را حساب کنیم.

۲. خوب دقت کنید برای حل مسائل شرطی اکثر مواقع از فرمول فوق استفاده نمی‌کنیم بلکه ابتدا شرط مسأله را روی فضای نمونه‌ای اعمال می‌کنیم و سپس در فضای نمونه‌ای جدید به دنبال خواستهٔ مسأله می‌گردیم.

۳. اگر A و B دو پیشامد ناسازگار باشند یعنی $A\cap B=\varnothing$. لذا $P(A\cap B)=P(\varnothing)=0$ و در نتیجه

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(\varnothing)}{P(B)}=\frac{0}{P(B)}=0$$

A و B ناسازگار $\Longrightarrow P(A|B)=0$

۴. بدیهی است که

$$P(A'|B)=1-P(A|B)$$

۵. رابطه‌های زیر همواره برقرارند

$$P(A|A')=0 \qquad P(A|A)=1 \qquad P(\varnothing|A)=0 \qquad P(A|\varnothing)=\text{تعریف نشده}$$

مثال. اگر $P(A)=\frac{1}{4}$ و $P(B)=\frac{1}{6}$ و $P(A|B)=\frac{1}{3}$ باشد، $P(A\cup B)$ را به دست آورید.

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \Rightarrow \frac{1}{3}=\frac{P(A\cap B)}{\frac{1}{6}} \Rightarrow P(A\cap B)=\frac{1}{18}$$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{4}+\frac{1}{6}-\frac{1}{18}=\frac{13}{36}$$

مثال. دو تاس سفید و سیاه را باهم پرتاب می‌کنیم. اگر مجموع دو تاس بیش‌تر از ۹ باشد، احتمال اینکه دو عدد رو شده برابر باشند را به دست آورید.

مجموع دو تاس بیش‌تر از ۹ باشد $B=\{(۴,۶),(۶,۴),(۵,۶),(۶,۵),(۵,۵),(۶,۶)\}$

دو عدد رو شده برابر باشند $A\cap B=\{(۵,۵),(۶,۶)\}$

$$n(B)=۶ \qquad n(A\cap B)=۲$$

$$P(A|B)=\frac{n(A\cap B)}{n(B)}=\frac{۲}{۶}=\frac{۱}{۳}$$

تمرین. اگر $P(A)=\frac{۲}{۵}$ و $P(B)=\frac{۱}{۳}$ و $P(A\cup B)=\frac{۷}{۱۵}$ باشد آنگاه مقدار $P(A|B)+P(B|A)$ را بیابید.

تمرین. خانواده‌ای دارای ۴ فرزند است. اگر بدانیم این خانواده دقیقاً سه فرزند دختر دارد، احتمال آن که فرزند اول این خانواده پسر باشد را به دست آورید.

۲ پیشامدهای مستقل:

می‌گوییم پیشامد A از پیشامد B مستقل است هرگاه رخ دادن A اثری روی رخ دادن B نگذارد.

به عبارت دیگر رخ دادن B، احتمال رخ دادن A را کم یا زیاد نمی‌کند و در واقع احتمال رخ دادن A با شرط رخ دادن B و بدون این شرط یکسان است، یعنی

$$P(A|B)=P(A)$$

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \xrightarrow[\text{مستقل}]{P(A|B)=P(A)} P(A)=\frac{P(A\cap B)}{P(B)}$$

$$\Rightarrow \boxed{P(A\cap B)=P(A)\times P(B)} \quad *$$

- برای اینکه نشان دهیم آیا دو پیشامد مستقل‌اند یا خیر، باید درستی رابطه * را بررسی کنیم.
- اگر دو پیشامد A و B مستقل از هم نباشند، آن‌ها را وابسته می‌نامیم.

مثال. برای ریاست شرکتی ۴ داوطلب وجود دارد که احتمال انتخاب شدن همهٔ آن‌ها یکسان است. اگر افراد را با x و y و z و t نام‌گذاری کنیم آیا دو پیشامد $A=\{x,t\}$ و $B=\{x,z\}$ مستقل‌اند؟

$$P(A)=\frac{۲}{۴}=\frac{۱}{۲} \qquad P(B)=\frac{۲}{۴}=\frac{۱}{۲} \qquad A\cap B=\{x\}\Rightarrow P(A\cap B)=\frac{۱}{۴}$$

$$P(A\cap B)\overset{?}{=\!=}P(A)\times P(B) \Rightarrow \frac{۱}{۴}=\frac{۱}{۲}\times\frac{۱}{۲} \ \checkmark$$

بله مستقل‌اند زیرا رابطه * برقرار است.

تذکر. اگر پیشامد A مستقل از پیشامد B باشد، آنگاه

۱. B نیز مستقل از A است. ۲. A' نیز مستقل از B است. ۳. B' نیز مستقل از A است. ۴. A' نیز مستقل از B' است.

اثبات رابطه ۲.

می‌دانیم که $A' \cap B = B - A$

$$P(A' \cap B) = P(B - A) = P(B) - \overbrace{P(A \cap B)}^{\text{مستقل}}$$

$$= P(B) - P(A) \times P(B) = P(B)\overbrace{\left(1 - P(A)\right)}^{P(A')} = P(B) \times P(A')$$

$$\Rightarrow P(A' \cap B) = P(A') \times P(B) \blacksquare$$

مثال. احتمال قبولی A در کنکور ۰/۷ و احتمال قبولی B در کنکور ۰/۸ است. چقدر احتمال دارد

الف) دست کم یکی از آن‌ها در کنکور قبول شود؟

ب) فقط A قبول شود.

ج) دست کم یکی از آن‌ها قبول نشود.

د) فقط B قبول شود.

هـ) فقط یکی از آن‌ها قبول شود.

م) هیچکدام قبول نشوند.

قبول شدن یا نشدن یک نفر روی قبول شدن یا نشدن دیگری اثری ندارد پس مستقل‌اند $P(A \cap B) = P(A) \times P(B)$.

حل الف ←

$$P(\text{حداقل یکی}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = ۰/۷ + ۰/۸ - ۰/۵۶ = ۹۴\%$$

$$\downarrow$$

$$P(A \cap B) = P(A) \times P(B) = ۰/۷ \times ۰/۸ = ۰/۵۶$$

حل ب ←

$$P(A\ \text{فقط}) = P(A - B) = P(A \cap B') = P(A) \times P(B')$$

$$= P(A) \times (1 - P(B)) = ۰/۷ \times (۱ - ۰/۸)$$

$$= ۰/۷ \times ۰/۲ = ۰/۱۴$$

حل ج ←

$$P(\text{دست کم یکی قبول نشود}) = 1 - P(\text{هر دو قبول شوند}) = 1 - \left(P(A) \times P(B)\right)$$

$$= ۱ - (۰/۷ \times ۰/۸) = ۴۴\%$$

حل د ←

$$P(B\ \text{فقط}) = P(B - A) = P(B \cap A') = P(B) \times P(A')$$

$$= ۰/۸ \times ۰/۳ = ۰/۲۴$$

حل هـ ←

$$P(\text{فقط یکی}) = P(A\ \text{فقط}) + P(B\ \text{فقط}) = ۰/۱۴ + ۰/۲۴ = ۰/۳۸$$

حل م ←

$$P(\text{هیچکدام}) = P(B \text{ نه } A \text{ نه}) = P(A' \cap B') = P(A') \times P(B')$$

$$= ۰/۳ \times ۰/۲ = ۰/۰۶ \text{ یا } ۶\%$$

مثال. در پرتاب یک سکه و یک تاس باهم، احتمال آن که تاس عدد "مضرب ۳" یا سکه "رو" ظاهر شود را بدست آورید.

عدد روی تاس مضرب ۳ $A=\{3,6\} \Rightarrow P(A)=\frac{2}{6}=\frac{1}{3}$

سکه رو بیاید $B=\{ر\} \Rightarrow P(B)=\frac{1}{2}$

A و B مستقل از یکدیگرند $\Rightarrow P(A\cap B)=P(A)\cdot P(B)=\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$

$$P(A\underset{\text{یا}}{\cup}B)=P(A)+P(B)-P(A\cap B)=\frac{1}{3}+\frac{1}{2}-\frac{1}{6}=\frac{2}{3}$$

مثال. اگر A و B دو پیشامد مستقل باشند بطوریکه $P(A)=\frac{2}{5}$ و $P(A\cup B)=\frac{1}{2}$ آنگاه $P(B')$ را بدست آورید.

A و B مستقل‌اند $\Rightarrow P(A\cap B)=P(A)\times P(B)$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B) \Rightarrow \frac{1}{2}=\frac{2}{5}+P(B)-(P(A)\times P(B))$$

$$P(B)(1-P(A))=\frac{1}{2}-\frac{2}{5}=\frac{5-4}{10}=\frac{1}{10}$$

$$\Rightarrow P(B)\left(1-\frac{2}{5}\right)=\frac{1}{10} \Rightarrow P(B)=\frac{\frac{1}{10}}{\frac{3}{5}} \Rightarrow P(B)=\frac{1}{6} \Rightarrow P(B')=1-P(B)=\frac{5}{6}$$

تمرین. دو تیر به سوی هدفی شلیک می‌کنیم. احتمال اصابت تیر اول به هدف ۰/۸ و این احتمال برای تیر دوم ۰/۹ است. احتمال اینکه

- حداقل یکی از آن‌ها به هدف بخورد چقدر است؟
- فقط یک تیر به هدف بخورد، را بدست آورید.

تمرین. دو نفر با نام‌های A و B با یک فرد مبتلا به بیماری ویروسی ملاقات کرده اند. احتمال ابتلا به این بیماری برای این دو نفر به ترتیب ۰/۵ و ۰/۶ می‌باشد. مطلوب است محاسبه احتمال اینکه هیچکدام از این دو نفر به این بیماری مبتلا نشوند؟

تمرین. اگر A و B دو پیشامد مستقل، $P(A)=\frac{1}{4}$ و $P(A\cup B')=\frac{3}{4}$ باشند. مقدار $P(B)$ را بدست آورید.

درس ۲. آمار توصیفی

آمار هنر پرورش داده‌هاست.

به مجموعه روش‌هایی که برای سازمان دادن، خلاصه‌کردن و توصیف مشاهدات مورد استفاده قرار می‌گیرد، آمار توصیفی می‌گویند.

- این روش‌های آماری بیشتر به مشخص‌کردن داده‌ها، تنظیم و ارائهٔ آن‌ها بصورت جدول و نمودار و تعیین روابط بین آن‌ها می‌پردازد.

۱ معیارهای گرایش به مرکز (شاخص‌های مرکزی):

داده‌های آماری اغلب تمایل دارند حول یک عدد بنام شاخص مرکزی جمع شوند. این شاخص‌ها عبارتند از:

I. میانگین mean　　II. میانه median　　III. مُد mod

I. میانگین:

متوسط یا معدل داده‌های آماری را میانگین می‌نامیم. به عبارت دیگر، میانگین مرکز ثقل داده‌هاست که آن را با نماد $\bar{x}$ نشان می‌دهیم.

اگر x_1، x_2، ...، x_n، n دادهٔ آماری باشند میانگین از رابطهٔ زیر بدست می‌آید:

$$\text{میانگین} = \frac{\text{مجموع داده‌ها}}{\text{تعداد داده‌ها}} \Rightarrow \boxed{\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}}$$

مثال. اگر میانگین داده‌های آماری ۵، ۳، x، ۲، ۱ و ۱ برابر ۴ باشد، مقدار x را بیابید.

$n = 6$ تعداد داده‌ها

$$\bar{x} = \frac{5+3+x+2+1+1}{6} \Rightarrow 4 = \frac{x+12}{6}$$

$$x + 12 = 6 \times 4 \Rightarrow x = 24 - 12 \Rightarrow x = 12$$

تذکر.

۱. میانگین مرسوم‌ترین معیار مرکزی است که تمام داده‌های آماری را تحت تأثیر خود قرار می‌دهد. به عبارتی هر تغییری در داده‌ها، در میانگین اثر می‌گذارد (خیلی حساس است!). اگر داده‌های آماری با مقدار ثابتی جمع و یا در عدد ثابتی ضرب شوند، آنگاه میانگین نیز چنین خواهد بود.

۲. میانگین داده‌های مساوی برابر با یکی از آن‌هاست. یعنی اگر تمام داده‌ها با هم برابر باشند، میانگین نیز با آن‌ها برابر است. در غیر اینصورت میانگین همواره عددی بین کوچک‌ترین و بزرگ‌ترین داده است.

$$x_1 = x_2 = \dots = x_n = a \Rightarrow \bar{x} = \frac{\overbrace{a + a + \dots + a}^{n \text{ تا}}}{n} = \frac{na}{n} = a$$

۳. تنها ایراد میانگین این است که به شدت تحت تأثیر نقاط پایانی (دادهٔ دورافتاده یا پرت) قرار می‌گیرد. در این صورت میانگین کارآمد نبوده بلکه به جای آن از میانه استفاده می‌کنیم.

- داده دورافتاده داده‌ای است که از بقیه داده‌ها فاصله زیادی دارد یعنی یا بسیار کوچک‌تر و یا بسیار بزرگ‌تر از بقیه داده‌هاست.

مثال. اگر میانگین ۱۰ دادهٔ آماری $x_1 ، x_2 ، \ldots ، x_{10}$ برابر ۱۲ باشد، میانگین داده‌های آماری زیر را به دست آورید.

$$2x_1-1 \;،\; 2x_2-1 \;،\; \ldots \;،\; 2x_{10}-1$$

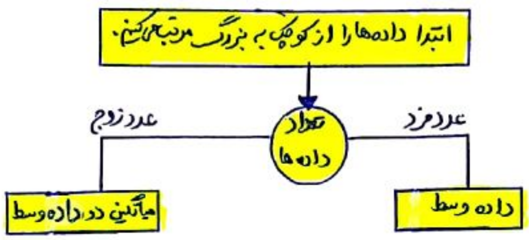
داده‌های جدید ۲ برابر منهای ۱ شده‌اند پس میانگین نیز چنین خواهد شد:

$$\bar{x}_{\text{جدید}} = 2(12)-1 = 23$$

II. میانه:

میانه شاخصی برای به دست آوردن وسط داده‌ها در بین داده‌های مرتب شده است و معمولاً در مواردی استفاده می‌شود که دادهٔ دورافتاده داشته باشیم.

برای به دست آوردن میانهٔ داده‌ها، ابتدا آن‌ها را از کوچک به بزرگ مرتب می‌کنیم (صعودی ↗). اگر تعداد داده‌ها فرد باشد، دادهٔ وسط میانه است، در غیر اینصورت اگر تعداد داده‌ها زوج باشد، عدد وسط نداریم! ، میانگین دو دادهٔ وسطی را به عنوان میانه در نظر می‌گیریم. میانه را با Q_2 نمایش می‌دهیم.



مثال. در داده‌های آماری زیر، میانه را به دست آورید.

الف) ۱۰ ۱۱ ۱۴ ۷۲ ۲۰ ۵ ۱۴

ب) ۹ ۶ ۱۶ ۱۲ ۱۸ ۶

حل الف: → ۵ ۱۰ ۱۱ $\boxed{14}$ ۱۴ ۲۰ ۷۲

میانه $Q_2 = 14$

حل ب: → ۶ ۶ ۹ ۱۲ ۱۶ ۱۸

میانه $Q_2 = \frac{9+12}{2} = \frac{21}{2} = 10/5$

مثال. علی در درس ریاضی نمرات زیر را کسب کرده است:

۲۰ و ۷ و ۱۹ ، ۱۷ و ۱۷

الف) میانگین و میانهٔ نمرات را حساب کنید.

ب) کدام یک از شاخص‌های فوق بیانگر وضع این دانش‌آموز در درس ریاضی است؟

ج) اگر معلم درس ریاضی برای جبران نمرهٔ ۷ امکان امتحان مجدد را به او بدهد، برای اینکه میانگین او در این درس بیش‌تر از ۱۸ شود او در این امتحان چه نمره‌ای باید کسب کند؟

حل الف ←
$$n=5 \qquad \bar{x} = \frac{17+17+19+7+20}{5} = \frac{80}{5} = 16 \Rightarrow \bar{x}=16$$

$$\to \quad 7 \quad \boxed{17} \quad 17 \quad 19 \quad 20 \qquad Q_2 = 17$$

حل ب ← چون دادهٔ دورافتاده داریم (در اینجا نمرهٔ ۷ از بقیه کوچک‌تر است) بهتر است از میانه استفاده کنیم.

حل ج ← می‌خواهیم: $\bar{x} > 18$

به جای نمرهٔ ۷ ، x قرار می‌دهیم و نامعادلهٔ زیر را حل می‌کنیم:

$$\frac{x+17+17+19+20}{5} > 18 \Rightarrow \frac{x+73}{5} > 18$$

$$\Rightarrow x+73 > 5 \times 18$$

$$x+73 > 90 \Rightarrow x > 17$$

باید نمرهٔ بالاتر از ۱۷ کسب کند.

تذکر.

۱. میانه داده‌ای است که نصف ($\frac{1}{2}$ یا ٪۵۰) داده‌ها از آن کم‌تر و نصف داده‌ها از آن بیش‌تر هستند.

۲. زمانی که داده‌ها باهم برابر باشند ، میانه با یکی از آن‌ها برابر است.

۳. درصورت وجود دادهٔ پرت ، بهترین شاخص مرکزی ، میانه خواهد بود.

۴. میانه نسبت به اندازه داده‌ها حساسیت نشان نمی‌دهد به عبارت دیگر میانه فقط به این نکته توجه دارد که تعداد داده‌های بعد از آن برابر تعداد داده‌های قبل از آن باشد و به بزرگی و یا کوچکی داده‌ها کاری ندارد.

III . مُد :

مُد داده‌ای است که بیش‌ترین تکرار (فراوانی) را دارد ، مثلاً وقتی رنگی مُد می‌شود ، بیش‌تر افراد جامعه از آن استفاده می‌کنند و آن رنگ بیش‌تر از سایر رنگ‌ها در جامعه دیده می‌شود. مُد هیچ اطلاعاتی درمورد سایر داده‌ها به ما نمی‌دهد و فقط به داده‌ای نگاه می‌کند که بیش‌ترین تکرار را دارد. نظرسنجی‌ها ، رأی‌گیری‌ها و انتخابات ، مهم‌ترین مواردی هستند که در آن‌ها از مُد استفاده می‌شود.

تمرین. میانگین ۱۰ عدد ۳ ، ۸ ، ۵ ، ۹ ، ۳ ، ۸ ، ۷ ، x ، ۳ ، ۴ برابر ۵/۵ است . مقدار x و میانه را به دست آورید.

$x = ?$ و Q_2

۲ معیارهای پراکندگی (شاخص های پراکندگی):

به معیارهایی که نزدیکی یا دوری داده ها را از میانگین برای ما اندازه گیری می کنند، شاخص پراکندگی می گوییم.

شاخص های پراکندگی در این قسمت عبارتند از:

I. دامنه تغییرات Range   II. واریانس Variance   III. انحراف معیار (خطای استاندارد) deviation

IV. ضریب تغییرات coffficient of variation   V. دامنه میان چارکی Inter quartile Range

I. دامنه تغییرات:

اختلاف بین بزرگ ترین (Max) و کوچک ترین (min) داده آماری را دامنه تغییرات می نامند و با R نمایش می دهند.

$$\boxed{R = Max - min}$$

مثال. در داده های آماری ۱۳ و ۵ و ۲۰ و ۷۲ و ۱۴ و ۱۱ و ۱۰ دامنه تغییرات را به دست آورید.

$Max = 72 \qquad min = 5 \qquad R = Max - min = 72 - 5 = 67$

نکات

۱. اگر داده ها برابر باشند، دامنه تغییرات صفر است و برعکس.

$$x_1 = x_2 = \cdots = x_n \iff R = 0$$

۲. چون داده های بین min و Max اثری روی دامنه تغییرات ندارد پس معیار خوبی برای پراکندگی داده ها نیست.

II. واریانس:

واریانس در لغت به معنای تفاوت و تغییر است. میانگین مجذور انحرافات از میانگین داده ها را واریانس می گویند و آن را با نماد δ^2 (سیگما ۲) نشان می دهند.

با به کارگیری یک فرمول از داده های توان واریانس را محاسبه کرد. برای این کار ابتدا میانگین داده ها را به دست می آوریم ($\bar{x}$)، میانگین را از تمام داده ها کم می کنیم و به توان ۲ می رسانیم سپس آن ها را با هم جمع کرده و بر تعداد داده ها (یعنی n) تقسیم می کنیم.

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

مثال. واریانس داده‌های آماری ۱ ، ۳ ، ۵ ، ۷ را به دست آورید.

$n = ۴$ تعداد داده‌ها $\qquad$ میانگین $\bar{x} = \frac{۱+۳+۵+۷}{۴} = \frac{۱۶}{۴} = ۴$

واریانس $\delta^۲ = \frac{(۱-۴)^۲+(۳-۴)^۲+(۵-۴)^۲+(۷-۴)^۲}{۴} = \frac{۹+۱+۱+۹}{۴} = \frac{۲۰}{۴} = ۵$

$\Rightarrow \delta^۲ = ۵$

نکات.

۱. واحد واریانس برابر با توان دوم واحد دادهٔ مورد نظر است. مثلاً اگر قد دانش‌آموزان برحسب cm موضوع آماری باشد، واحد واریانس $(cm)^۲$ سانتی‌متر مربع خواهد بود.

۲. هرچه واریانس داده‌ها کوچک‌تر باشد، پراکندگی داده‌ها کم‌تر است، یعنی داده‌ها به هم و در نتیجه به میانگین خودشان نزدیک‌ترند.

۳. اگر داده‌های آماری برابر باشند، واریانس صفر است و برعکس.

$$x_۱ = x_۲ = \cdots = x_n \Leftrightarrow \delta^۲ = ۰$$

۴. اگر داده‌ها با مقدار ثابتی جمع شوند واریانس آن‌ها تغییر نخواهد کرد اما اگر داده‌ها را در عدد ثابتی ضرب کنیم، واریانس آن‌ها در مجذور این عدد ضرب خواهد شد. با بیان ریاضی، اگر واریانس داده‌های $x_۱$ و $x_۲$ و ... و x_n برابر $\delta^۲$ باشد آنگاه واریانس داده‌های $x_۱+b$ و $x_۲+b$ و ... و x_n+b برابر با $\delta^۲$ و واریانس داده‌های $ax_۱$ ، $ax_۲$ ، ... ، ax_n برابر $a^۲ . \delta^۲$ خواهد بود و در حالت کلی داریم:

$$\delta^۲_{ax+b} = a^۲ . \delta^۲$$

مثال. واریانس تعدادی دادهٔ آماری برابر ۵ است. ابتدا به هر دادهٔ آماری ۴ واحد اضافه می‌کنیم و سپس آن‌ها را ۲ برابر می‌کنیم. واریانس داده‌های جدید را به دست آورید.

فرض کنید $x_۱$ و $x_۲$ و ... و x_n داده‌های اولیه با واریانس $\delta^۲ = ۵$ باشند، داده‌های جدید به صورت زیر می‌باشند:

$۲(x_۱+۴)$ ، $۲(x_۲+۴)$ ، ... ، $۲(x_n+۴)$ می‌باشند. می‌دانیم جمع عدد ثابت در واریانس تأثیری ندارد و داریم:

$$\delta^۲_{۲(x+۴)} \xlongequal{\text{حذف عدد ۴}} \delta^۲_{۲x} = (۲)^۲ \times \underset{۵}{\delta^۲} = ۴ \times ۵ = ۲۰$$

مثال. اگر $x > ۰$ و واریانس داده‌های x و x و $۳x$ و $۳x$ برابر ۲۵ باشد، آنگاه دامنهٔ تغییرات را بیابید. $R = ?$

$n = ۴ \qquad \bar{x} = \frac{x+x+۳x+۳x}{۴} = ۲x$

$$\delta^۲ = \frac{(x-۲x)^۲+(x-۲x)^۲+(۳x-۲x)^۲+(۳x-۲x)^۲}{۴} = x^۲$$

$\delta^۲ = ۲۵ \qquad x^۲ = ۲۵ \xrightarrow[x>۰]{\sqrt{\ }} x = ۵ \Rightarrow$ داده‌ها: ۵ ، ۵ ، ۱۵ ، ۱۵

$$R = Max - min = ۱۵ - ۵ = ۱۰$$

III. انحراف معیار:

در عمل واریانس عدد بزرگی است و به جای آن از انحراف معیار استفاده می‌کنیم. به این ترتیب که از واریانس داده‌های آماری جذر می‌گیریم و آن را با نماد δ (سیگما) نمایش می‌دهیم.

$$\boxed{\text{انحراف معیار } \delta = \sqrt{\text{واریانس}}}$$

این شاخص میزان انحراف داده‌ها از میانگین را نشان می‌دهد و یک معیار معمول برای سنجش میزان پراکندگی داده‌هاست.

انحراف معیار تمامی ویژگی‌های واریانس را دارد (صفحه ۱۰۲ جزوه). با این تفاوت که واحد انحراف معیار همان واحد اندازه‌گیری داده‌هاست و اگر داده‌ها در عدد a ضرب شوند، انحراف معیار در $|a|$ ضرب خواهد شد. یعنی

$$\delta_{ax} = |a| \cdot \delta$$

مثال. انحراف معیار داده‌های آماری ۱ و ۲ و ۵ و ۴ را به دست آورید.

$$n = ۴ \qquad \bar{x} = \frac{۱+۲+۵+۴}{۴} = \frac{۱۲}{۴} = ۳$$

$$\text{واریانس } \delta^2 = \frac{(۱-۳)^2+(۲-۳)^2+(۵-۳)^2+(۴-۳)^2}{۴} = \frac{۴+۱+۴+۱}{۴} = \frac{۱۰}{۴} = ۲/۵$$

$$\text{انحراف معیار } \delta = \sqrt{۲/۵} \simeq ۱/۵۸$$

مثال. انحراف معیار داده‌های x_1 و x_2 و ... و $x_{۱۰}$ برابر ۳ است. انحراف معیار داده‌های زیر را به دست آورید:

$$-۲x_1+۲ \;\text{ و }\; -۲x_2+۲ \;\text{ و }\; \cdots \;\text{ و }\; -۲x_{۱۰}+۲$$

داده‌ها در ۲- ضرب و با ۲+ جمع شده‌اند. جمع عدد ثابت در انحراف معیار تأثیری ندارد.

$$\delta_{-۲x+۲} = |-۲| \times \delta = ۲ \times ۳ = ۶$$

IV. ضریب تغییرات:

اگر بخواهیم پراکندگی دو دسته از داده‌های آماری را که هم‌واحد نیستند با هم مقایسه کنیم از شاخصی بنام ضریب تغییرات استفاده می‌کنیم. (مثل تناسب بین قد m و وزن Kg). نسبت انحراف معیار به میانگین داده‌ها را ضریب تغییرات می‌نامیم و با نماد C.V نمایش می‌دهیم:

$$\text{ضریب تغییرات} = \frac{\text{انحراف معیار}}{\text{میانگین}} \Rightarrow \boxed{C.V = \frac{\delta}{\bar{x}}}$$

ضریب تغییرات بدون واحد اندازه‌گیری بیان می‌شود، یعنی به واحد اندازه‌گیری بستگی ندارد.

مثال. ضریب تغییرات داده‌های ۶ ، ۵ ، ۴ ، ۴ ، ۴ ، ۳ ، ۲ را به دست آورید.

$$n=7 \qquad \text{میانگین}\ \bar{x}=\frac{6+5+4+4+4+3+2}{7}=\frac{28}{7}=4$$

$$\text{واریانس}\ \delta^2=\frac{(6-4)^2+(5-4)^2+(4-4)^2+(4-4)^2+(4-4)^2+(3-4)^2+(2-4)^2}{7}=\frac{10}{7}\simeq 1/42$$

$$\text{انحراف معیار}\ \delta=\sqrt{1/42}\simeq 1/19$$

$$\text{ضریب تغییرات}\ C.V=\frac{\delta}{\bar{x}}=\frac{1/19}{4}\simeq /29$$

نکات.

۱. ضریب تغییرات میزان پراکندگی به ازای یک واحد از میانگین را نشان می‌دهد. هرچه ضریب تغییرات کم‌تر باشد، پراکندگی داده‌ها کم‌تر است.

۲. ضریب تغییرات معمولاً به صورت درصد بیان می‌شود. $C.V=\frac{\delta}{\bar{x}}\times 100$

۳. اگر داده‌های آماری برابر باشند، ضریب تغییرات صفر است و برعکس.

$$x_1=x_2=\dots=x_n \Leftrightarrow C.V=0$$

۴. در محصولات تولیدی هرچه $C.V$ کم‌تر باشد به معنای کیفیت بالاتر است.

مثال. ۵ سال دیگر ضریب تغییرات سن دانش‌آموزان کلاس شما چه تغییری می‌کند؟

کم می‌شود. زیرا بنابر رابطهٔ $C.V=\frac{\delta}{\bar{x}}$ اگر به داده‌ها عددی اضافه شود انحراف معیار تغییری نخواهد کرد اما این عدد به میانگین اضافه می‌شود. بنابراین صورت کسر ثابت و به مخرج ۵ واحد اضافه می‌شود و کل کسر کوچک شده لذا ضریب تغییرات کم خواهد شد.

مثال. اگر ضریب تغییرات داده‌های $x-3$ و $2y-7$ و $4z-12$ و ۳ برابر صفر باشد، آنگاه میانگین x و y و z را به دست آورید.

چون ضریب تغییرات صفر است پس داده‌ها همگی با هم برابرند:

$$x-3=2y-7=4z-12=3 \Rightarrow x=6\ ,\ y=5\ ,\ z=4$$

$$\bar{x}=\frac{x+y+z}{3}=\frac{6+5+4}{3}=\frac{15}{3}=5$$

$$\Rightarrow \bar{x}=5$$

۷. دامنهٔ میان چارکی:

اگر انحراف معیار زیاد باشد بدین معنی است که میزان پراکندگی داده‌ها زیاد بوده و دادهٔ دورافتاده داریم. لذا در این حالت انحراف معیار شاخص خوبی برای محاسبات آماری نخواهد بود بلکه به جای آن از معیار پراکندگی دیگری که به کمک میانه تعریف می‌شود، استفاده می‌کنیم.

این معیار دامنهٔ میان چارکی نام دارد که با IQR نمایش می‌دهیم و برای محاسبهٔ آن به صورت زیر عمل می‌کنیم:

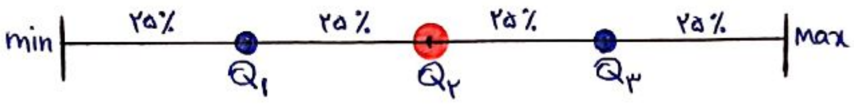
۱. ابتدا با مرتب کردن داده‌ها، میانه را پیدا کرده و آن را Q_2 یا چارک دوم می‌نامیم.

۲. داده‌های قبل از میانه را جدا کرده و میانه آن‌ها را پیدا می‌کنیم و آن را Q_1 یا چارک اول می‌نامیم.

۳. 〃 بعد 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃 Q_3 〃 چارک سوم 〃.

اختلاف بین چارک اول و چارک سوم را دامنه میان چارکی می‌نامیم.

دامنه میان چارکی = چارک سوم − چارک اول ⇐ $\boxed{IQR = Q_3 - Q_1}$



چارک‌ها (چارک اول، چارک دوم، چارک سوم) ۳ عدد هستند که داده‌های مرتب شده را به ۴ قسمت مساوی تقسیم می‌کنند.

مثال. در داده‌های آماری زیر، چارک‌های اول و سوم را مشخص کنید.

۱۳ و ۱۶ و ۱۵ و ۱۳ و ۱۴ و ۱۱ و ۱۳ و ۱۲ و ۱۰ و ۱۲ و ۱۰ و ۹

ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم:

$\rightarrow \quad 9 \quad 10 \quad 10 \quad 11 \quad 12 \quad 12 \quad 13 \quad 13 \quad 13 \quad 14 \quad 15 \quad 16$

میانه Q_2

چارک اول $Q_1 = \frac{10+11}{2} = 10/5$

چارک سوم $Q_3 = \frac{13+14}{2} = 13/5$

تمرین. داده‌های آماری ۱۵، ۵، ۴، ۳، ۱۱، ۱، ۸، ۹، ۱۱ مفروضند. دامنه تغییرات و چارک اول و چارک دوم و چارک سوم را مشخص کنید.

تمرین. کارخانه‌ای دو نوع تسمه تایم تولید می‌کند. میانگین طول عمر برای نوع A و B به ترتیب ۷۵۰۰۰ کیلومتر و ۹۰۰۰۰ کیلومتر و واریانس برای نوع A و B به ترتیب ۴۰۰ کیلومتر و ۶۲۵ کیلومتر مربع است. کدام نوع بهتر است؟

تمرین. اگر واریانس داده‌های زیر صفر باشد، میانگین a و b را به دست آورید.

$a+2$ و ۴ و $b+3$

تمرین. ضریب تغییرات داده‌های ۱۴، ۱۲، ۱۰، ۸، ۶ را به دست آورید.

ریاضی الفبای زندگی است.

موفقیت آمیز با رشد طلیب حیدری