~~مثال: میانگین سن یک گروه ۲۰ سال~~

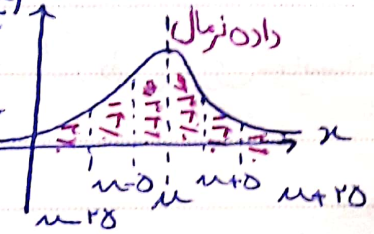
۶- چولگی یا عدم قرینگی:

مقدمه ای بر داده های استاندارد standard data

توزیع استاندارد داده ها normal/standard distribution data

تعریف: داده های که میانگین آنها صفر و واریانس آنها ۱ باشد داده استاندارد نامیده می شود.

زنگوله ای یا شکل ناقوسی



تابعی که این توزیع نرمال را توصیف می کند:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

x دارای توزیع نرمال است با میانگین μ : $X \underset{\text{علامت توزیع داده ها}}{\sim} N(\overset{\text{میانگین}}{\mu}, \overset{\text{انحراف معیار}}{\sigma})$

نکته: در داده های نرمال ⇐ میانه = میانگین

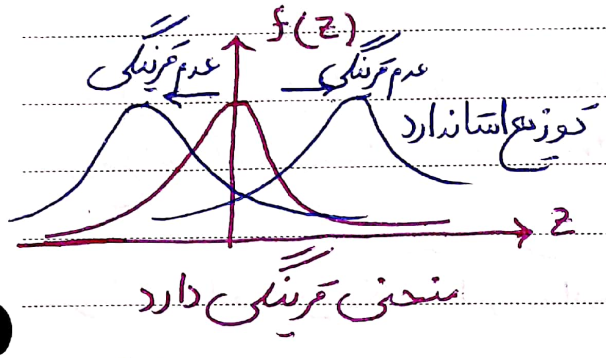
نکته: با تغییر متغیر به صورت ذیل $Z = \frac{x-\mu}{\sigma}$ = تغییر متغیر

$Z \sim N(0,1)$ ، Z دارای توزیع ذیل است با میانگین ۰ و انحراف معیار ۱

مفهوم: اگر توزیع داده‌ها نرمال باشد $(\mu , \delta) N \sim x_n , x_{n-1} , \dots , x_2 , x_1$ نرمال

$$Z = \frac{x - \mu}{\delta}$$

$(0 , 1) N \sim Z_n , Z_{n-1} , \dots , Z_2 , Z_1$ استاندارد.



نکته: به طور کلی برای مقایسه دو نمونه / جامعه می‌بایست داده‌های آنها را استاندارد کنیم

نمره فرد A ⑰ $\begin{cases} \mu_1 = 18 \\ \delta_1 = 1 \end{cases}$: نمرات درس آمار و احتمالات

نمره فرد A ⑭ $\begin{cases} \mu_2 = 15 \\ \delta_2 = 0.8 \end{cases}$: نمرات درس مدار الکتریکی

سوال: عملکرد فرد A در کلاس آمار بهتر بوده است یا؟

جواب کلاس مدار چون Z آن بیشتر است.

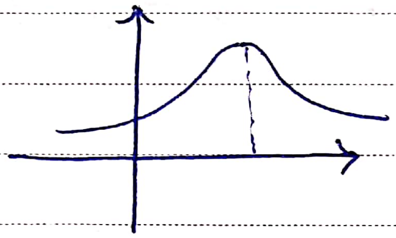
⑥ چولگی یا عدم قرینگی:

در مقایسه دو یا چند جامعه با یکدیگر در صورت تشابه از پارامترهای مرکزی استفاده می‌کردیم

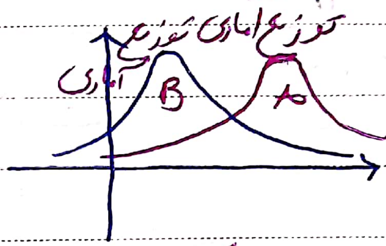
اما در صورت تساوی برخی از پارامترهای مرکزی مانند $\begin{cases} \text{میانگین} \\ \vdots \end{cases}$ اختلاف جوامع

آماری به کمک شاخص های پراکندگی { انحراف معیار ، : } معین می گردید.

در برخی از موارد پارامترهای پراکندگی نیز به علت تساوی جوابگو نیستند.



توزیع در جامعه آماری متقارن



برای مثال : در شکل های رسم شده دو توزیع دارای میانگین و واریانس مساوی هستند

اما توزیع داده یکسان نیست.

توزیع B : دارای تراکم در حول و حوش مبدأ مختصات است

درحالیکه

توزیع A : در نقطه مقابل آن قرار دارد.

نکته ۱ : به این تفاوت در اصطلاح (چولگی) یا (عدم قرینگی) skewness یا (انحراف از قرینگی)

نکته ۲ : چولگی داده ها نسبت به توزیع متقارن سنجیده می شود.

نکته: پارامترهای مرکزی در توزیع نرمال برابرند. $X \sim N(\mu, \delta) \Rightarrow \mu = Md = Mo$

نکته: هرچه توزیع داده‌ها با توزیع متقارن تفاوت بیشتری داشته باشد انحراف از قرینگی آن بیشتر است.

شاخص چولگی:

گشتاور مرکزی مرتبه سوم حول میانگین

ضریب چولگی $SK \triangleq \frac{\mu_3}{\delta^3} = \frac{\sum_{i=1}^{n} \frac{(x_i-\mu)^3}{n}}{\left[\underbrace{\frac{\sum (x_i-\mu)^2}{n}}_{\delta^2}\right]^{3/2}}$

نکته ۱۹:

$$SK \triangleq \frac{\mu_3}{\delta^3} = \begin{cases} < 0 & \text{چوله به چپ} \\ = 0 & \text{نرمال} \\ > 0 & \text{چوله به راست} \end{cases}$$

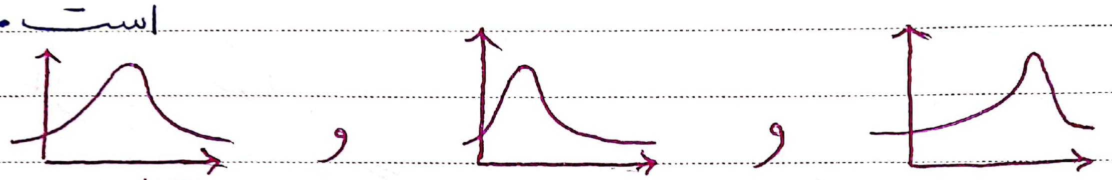
نکته ۲:

$-1 < SK < 1$

اگر SK به ۱- نزدیک باشد چولگی به چپ نزدیک است.

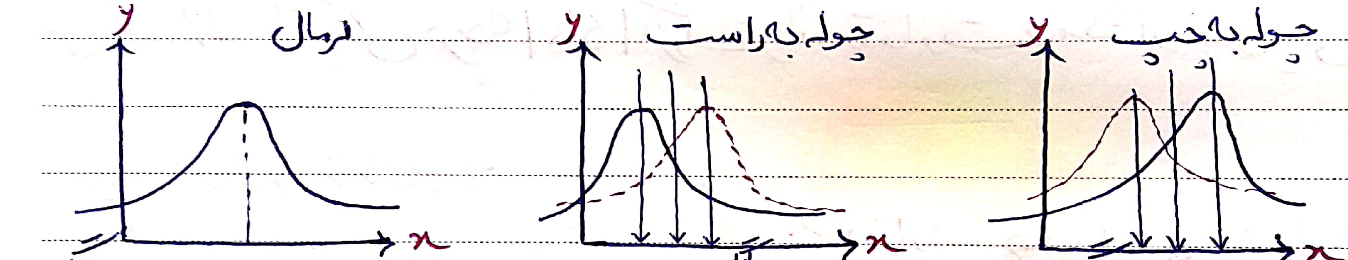
اگر SK به ۱ نزدیک باشد چولگی به راست نزدیک است.

اگر SK به صفر نزدیک باشد ~~[illegible]~~ متقارن است.



چوله به چپ و چوله به راست و متقارن

$SK < 0$ ، $SK > 0$ ، $SK = 0$



مد میانه میانگین ، میانگین مد میانه ، مد = میانه = میانگین

تعبیرهای (SK):

فرمول اول پیرسون:

$$SK_1 \triangleq \frac{\mu - \overset{\text{مُد}}{Mo}}{\delta} = \begin{cases} <0 & \text{چوله به چپ} \\ =0 & \text{متقارن} \\ >0 & \text{چوله به راست} \end{cases}$$

فرمول دوم پیرسون:

$$SK_2 \triangleq \frac{3(\mu - \overset{\text{میانه}}{Md})}{\delta}$$

نکته بسیار مهم: اگر چولگی بسیار خفیف باشد:
(خیلی زیاد نباشد)

$$Sk_1 \cong Sk_2 \Rightarrow \mu - Mo \cong 3(\mu - Md)$$

شرط: چولگی خفیف باشد.

تحلیل ضریب چولگی:

عدد مطلق ضریب چولگی، میزان اختلاف جوامع آماری با توزیع نرمال را نشان می دهد از لحاظ قرینگی هرچه $|Sk|$ بزرگ تر باشد تفاوت جامعه از نظر قرینگی با توزیع نرمال بیش تر خواهد بود.

1- $|Sk| \leqslant 0{,}1 \Rightarrow$ تقریباً چولگی وجود ندارد و جامعه از نظر قرینگی تقریباً نرمال است.

2- $0{,}1 < |Sk| \leqslant 0{,}5 \Rightarrow$ چولگی موجود است اما اندک

3- $|Sk| > 0{,}5 \Rightarrow$ چولگی زیاد است.

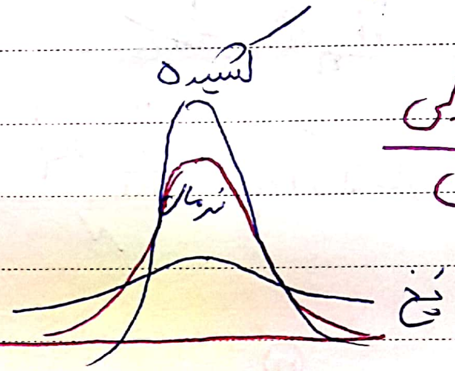
۷- کشیدگی:

کشیدگی معیاری است بدون واحد که ارتفاع منحنی را تجزیه و تحلیل می کند و در ارتباط مستقیم با پراکندگی است.

$$\text{کشیدگی}\ \beta \triangleq \frac{\mu_4}{\delta^4} = \frac{\sum_{i=1}^{n} \frac{(x_i-\mu)^4}{n}}{\left[\frac{\sum (x_i-\mu)^2}{n}\right]^2}$$

$$= \begin{cases} >3 & \text{کشیده} \\ =3 & \text{نرمال} \\ <3 & \text{پخ} \end{cases}$$

$$K = \beta - 3 = \begin{cases} <0 & \text{پخ} \\ =0 & \text{نرمال} \\ >0 & \text{کشیده} \end{cases}$$



درباره تعریف کشیدگی
ضریب کشیدگی

یا $\left[K = \frac{\mu_4}{\delta^4} - 1\right]$

تحلیل ضریب کشیدگی:

قدر مطلق ضریب کشیدگی نشان دهنده میزان اختلاف ارتفاع با توزیع نرمال است.

- اگر $|k| \leqslant 0.1$: توزیع جامعه تقریبا نرمال
- اگر $0.1 < |k| \leqslant 0.5$: توزیع جامعه از نظر کشیدگی تفاوت اندکی با توزیع نرمال دارد.
- اگر $|k| > 0.5$: تفاوت توزیع با توزیع نرمال از نظر کشیدگی بسیار زیاد است.

گشتاورها:

① گشتاور حول مبدأ مختصات

② گشتاور حول میانگین به عنوان یک پارامتر مرکزی

③ گشتاور حول نقطه دلخواه

① ـ گشتاور حول مبدأ مختصات:

$$m_n \triangleq \frac{\sum_{i=1}^{n} x_i^{n}}{N}$$ ، گشتاور n ام حول مبدأ مختصات

مرتبه گشتاور حول مبدأ مختصات

$$m_n \triangleq \frac{\sum_{i=1}^{N} F_i x_i^{n}}{N}$$

میانگین، گشتاور مرتبه اول حول مبدأ مختصات

$$\mu = \frac{\sum_{i=1}^{N} x_i^{1}}{N} = m_1$$ ، گشتاور مرتبه اول

② ـ گشتاور حول میانگین:

گشتاور n ام حول میانگین

$$\begin{cases} \mu_n \triangleq \dfrac{\sum_{i=1}^{N} (x_i-\mu)^{n}}{N} \\ \mu_n \triangleq \dfrac{\sum_{i=1}^{N} F_i (x_i-\mu)^{n}}{N} \end{cases}$$

مرتبه گشتاور حول میانگین