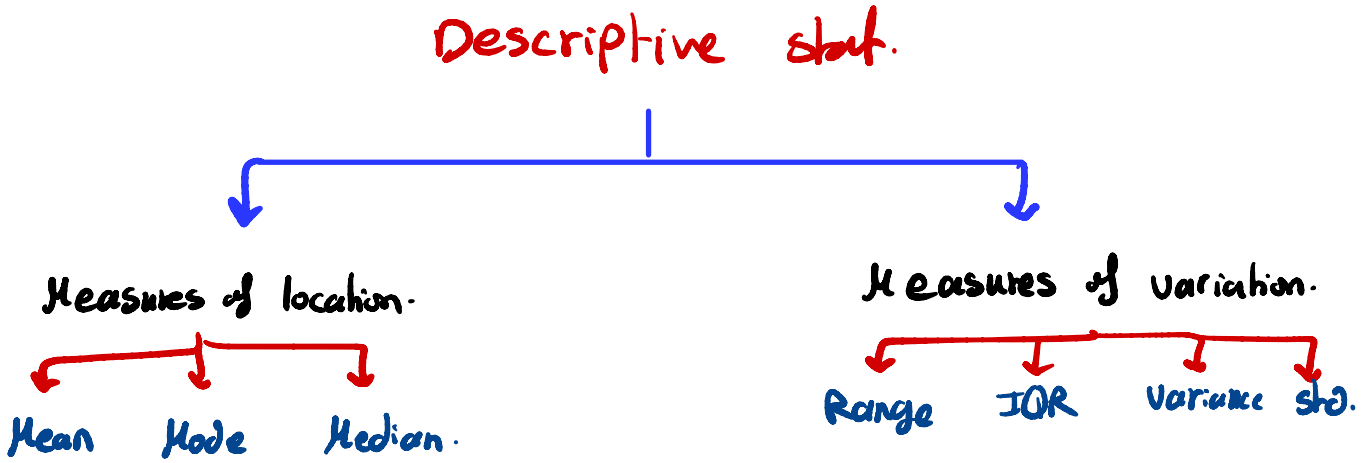
* Chapter (2) : Descriptive Statistics :-



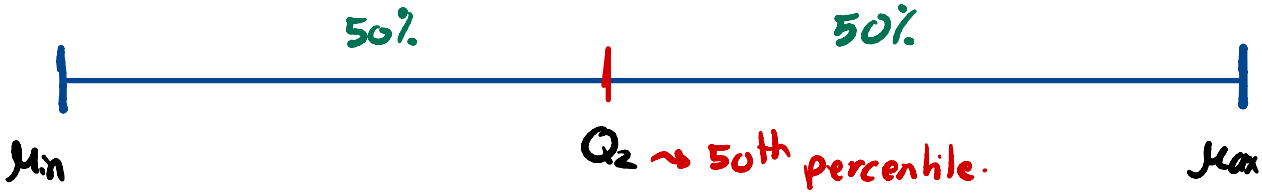
⇒ Measures of location :-

1) The Mean :-

$$\bar{X} = \frac{\Sigma x_i}{n}$$

↝ affected by outliers.

2) The Median (Q_2) :-



↝ The value that falls in the middle & it's not affected by outliers.

$$Q_2 \rightarrow \frac{n}{2} \begin{cases} \text{fraction} \rightarrow \text{take next int.} \\ \text{whole no} \rightarrow \dfrac{k^{th} + (k+1)^{th}}{2} \end{cases}$$

3) The mode :-

↝ The most frequent value & used for qualitative (not numiric) data.

one mode → unimodal

two modes → bimodal

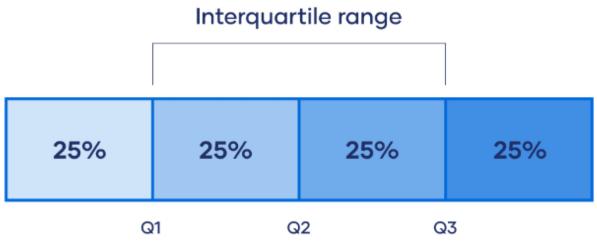
three modes → trimodal.

⇒ Measures of Variation:-

1) Range = Max - Min.

→ affected by outliers & not representitive for the data.

2) IQR :-



Q_1 :- The first quartile (lower), 25% of the data below it.

Q_3 :- the third quartile (upper), 75% of the data below it.

↳ find its location by :-

$Q_1 \rightarrow \frac{n}{4}$

$Q_3 \rightarrow \frac{3n}{4}$

- fraction → take next int.
- whole no. → $\frac{K^{th} + (K+1)^{th}}{2}$

* percentiles :-

$P_K \rightarrow \frac{K}{100} \cdot n$

- fraction → next int.
- whole no. → $\frac{K^{th} + (K+1)^{th}}{2}$

3) + 4) variance & std :-

Deviation = $x_i - \bar{x}$. → $\Sigma (x_i - \bar{x}) = 0$ always.

$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$ OR $s^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$ (better)

$SD = S = \sqrt{var(x)}$.

s^2 & s affected by outliers

* The coefficient of variation :-

$C.V = \frac{S}{\bar{x}} \cdot 100\%$ → Unit free.

↳ used to compare the variation between two samples.

* Outliers :-

- Any value less than $Q_1 - 1.5IQR$ OR More $Q_3 + 1.5IQR$.

  Is an outlier.

- Any value less than $Q_1 - 3IQR$ OR More than $Q_3 + 3IQR$

  Is an extreme outlier.

* The Relative frequency :-

↝ $R.f = \frac{f}{\Sigma f}$. Note: $\Sigma R.f = 1$. always.

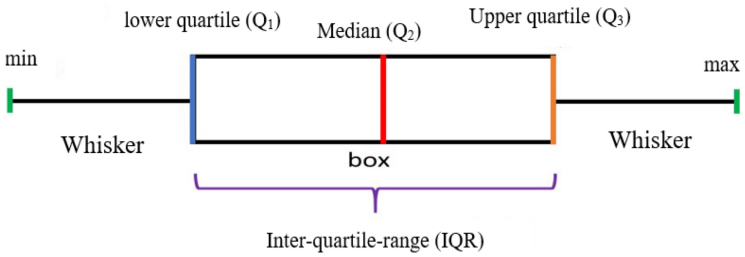
* Graphic methods :

[1] Bar graph :

↝ used for discrete data. & used to compare two datasets.

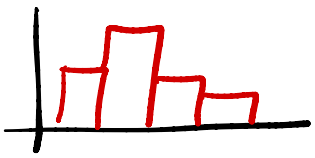
[2] Box & whiskers plot :-

↝ represents 5- number summary :- Min, Q_1, Q_2, Q_3, Max.



* ↝ outliers.

[3] Histogram :-



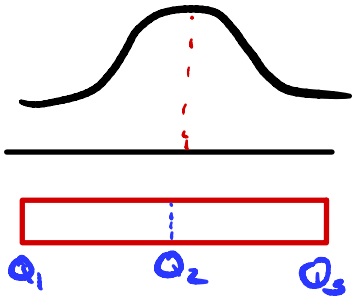
↝ used for continous data.

↝ plot URB (or X) Vs frequency (OR relative frequency).

↝ used to know the shape of data distribution.

* Skewness :-

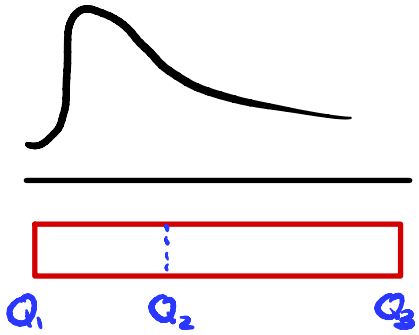
① Symmetric :-



$\bar{x} = Q_2 = \text{mode}.$

$Q_3 - Q_2 = Q_2 - Q_1.$

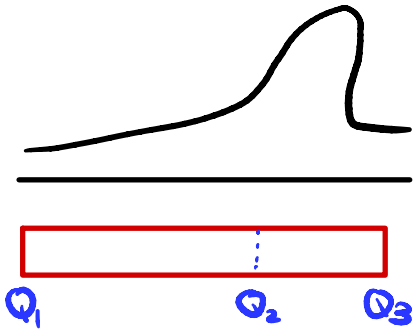
② Skewed to right (positive) :-



$\bar{x} > Q_2 > \text{mode}.$

$Q_3 - Q_2 > Q_2 - Q_1$

③ Skewed to left (negative) :-



$\bar{x} < Q_2 < \text{mode}.$

$Q_3 - Q_2 < Q_2 - Q_1$

* Coding (linear transformation) :-

$y = ax + b$

① $\bar{y} = a.\bar{x} + b.$

② $s_y = |a|.\, s_x.$

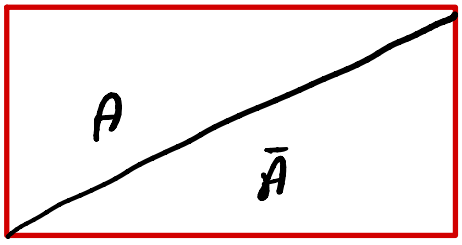
③ $s_y^2 = a^2.\, s_x^2$

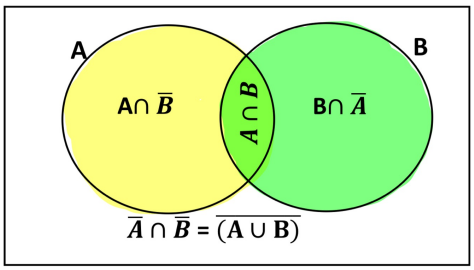
} → not affected by addition.

* chapter (3): Elements of probability.

① $P(S) = 1$ ② $P(\emptyset) = 0$ ③ $0 \le P(A) \le 1$

⇒ Rules of prob.:





① $P(A) + P(\bar{A}) = 1$.

② $P(A \cap \bar{B}) = P(A) - P(A \cap B)$.

③ $P(\bar{A} \cap B) = P(B) - P(A \cap B)$.

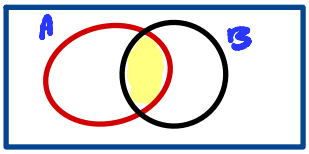
④ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

⑤ Demorgan Law's:

1. $P(\bar{A} \cup \bar{B}) = \overline{P(A \cap B)} = 1 - P(A \cap B)$

2. $P(\bar{A} \cap \bar{B}) = \overline{P(A \cup B)} = 1 - P(A \cup B)$.

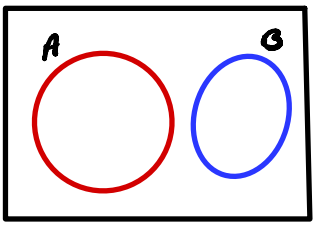
⑥ If A and B are independent:



- $P(A \cap B) = P(A) \cdot P(B)$.
- $P(A \cap \bar{B}) = P(A) \cdot P(\bar{B})$.
- $P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B)$.
- $P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B})$.

⑦ If A and B mutually Exclusive :

$P(A \cap B) = 0 \rightsquigarrow P(A \cup B) = P(A) + P(B)$.



⑧ conditional prob. :-

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

If A and B independent : $P(A|B) = P(A)$ & $P(B|A) = P(B)$.

Note :- RR of B|A :-

$$\frac{P(B|A)}{P(B|\bar{A})} \begin{cases} 1 \rightsquigarrow \text{Independent.} \\ >1 \rightsquigarrow \text{dependent.} \end{cases}$$

* Screening test :-

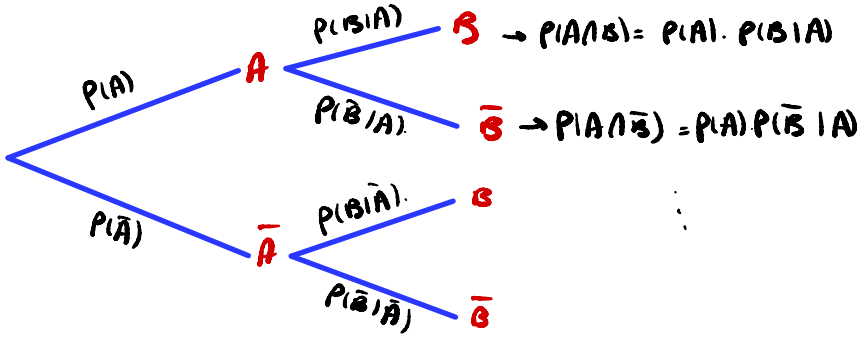
Test (symptoms) \ Actual (Disease)	Positive (D)	Negative ($\bar{D}$)
Positive (T)	TP	FP
Negative ($\bar{T}$)	FN	TN.

① $P(FN) = P(\bar{T}|D)$ ② $P(FP) = P(T|\bar{D})$

③ $P(PV+) = P(D|T)$ ④ $P(PV-) = P(\bar{D}|\bar{T})$

⑤ sensitivity $= P(T|D)$ ⑥ specificity $= P(\bar{T}|\bar{D})$.

* Tree diagram & Bayes theorem :-



* chapter (4) :- Discrete prob. distributions

Any Random variable can be : ① Discrete "countable"
② cts "measurable".

→ prob. mass function (P.M.F) :-

① $P(X=x_i) \geq 0$ ② $\sum P(X=x_i) = 1$.

→ The Expected value :- weighted mean.

$E(x) = \mu = \sum x \cdot p(x)$.

→ The variance & s.d :-

$var(x) = \sigma^2 = E(x-\mu)^2 = E(x^2) - (E(x))^2$

s.d $= \sigma = \sqrt{var(x)}$

Note : $E(x^2) = \sum x^2 \cdot p(x)$.

* The Binomial distribution :-

X is said to follow Binomial if it satisfy the following :-

① we have n independent & identical trials.

② we have two outcomes only.

③ The probability of success p, is always the same for all trials

If $X \sim Bin(n, p)$ $x = 0, 1, 2, \ldots n$.

n : no. of trials p : prob. of success.

$q = 1-p$: prob. of failure.

If $X \sim Bin(n, p)$:-

① $P(X = K) = \binom{n}{K} p^K \cdot q^{n-K}$.

↳ calc. $\binom{n}{K} = {}^nC_K$

② $P(X \geq K)$, $P(X \leq K)$ ⇝ use tables.

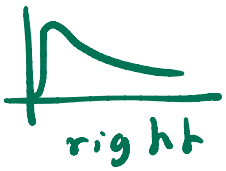
③ $E(X) = n \times p$.

④ $var(X) = n \times p \times q$.

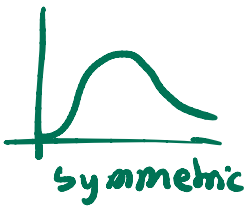
⑤ $std(X) = \sqrt{var(X)} = \sqrt{n \times p \times q}$.

* Skewness for p :-

① $p < 0.5$



② $p = 0.5$



③ $p > 0.5$

---

* chapter(5) :- cts prob. distribution.

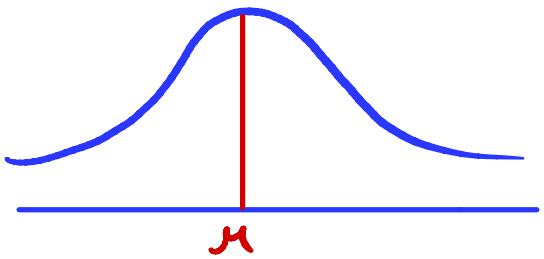
$X \sim N(\mu, \sigma^2)$

① symmetric = bell-shaped.

② Mean = Mode = Median.

③ prob. = area ⇝ tables

④ total area = 1.



* The standard normal :- $Z \sim N(0,1)$.

1) $P(Z \le K)$ → col A.

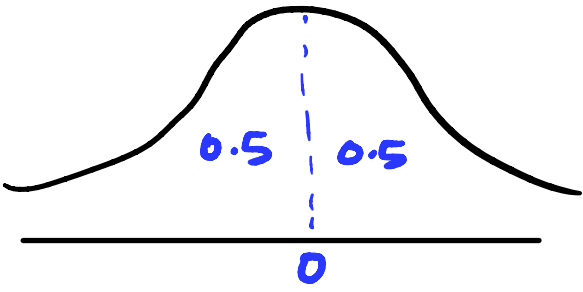
2) $P(Z > K)$ → col B.

3) $P(-a < Z < a)$ → col D.

4) $P(a < Z < b) = P(Z < b) - P(Z < a)$.

5) $P(Z < -K) = P(Z > +K)$.

6) $P(Z = K) = 0$.



⚠ في الحل العكسي، لما يطلب مني الرقم اللي بحصر تحته مساحة معينة:

1) إذا كانت المساحة أكبر من 0.5 بروح لجدول الموجب وببحث داخله.

2) إذا كانت المساحة اقل من 0.5 بروح على جدول السالب وببحث داخله.

3) إذا كان المساحة 0.5 فالرقم هو صفر.

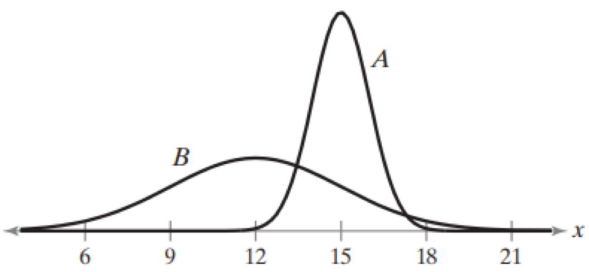
* Transforming to standard normal :-

$X \sim N(\mu, \sigma^2)$

$Z = \frac{X - \mu}{\sigma} = \frac{x - mean}{sd}$.

* Notes :-

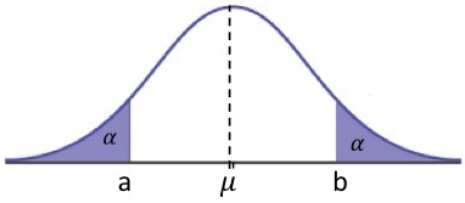
①



→ As the variance increase, the normal curve becomes wider & flatter.

Very important notes:

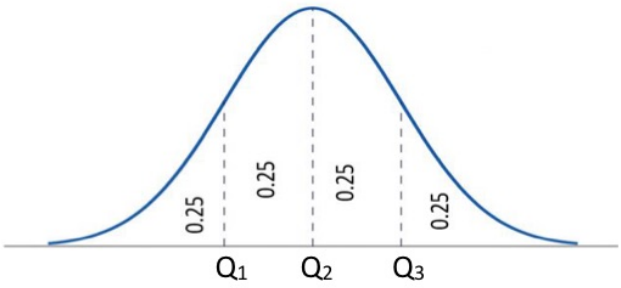
[1]



- If P(X<a) = P(X>b) then μ is the mid-point between a and b.

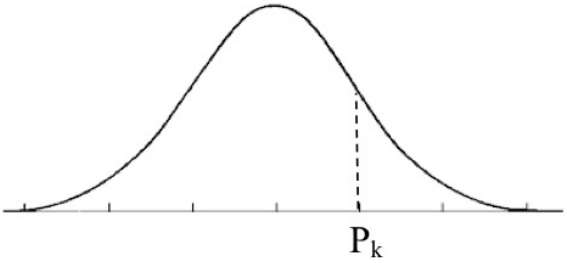
  $\therefore \ \mu = \frac{a+b}{2}$

[2]



- **$P(X \leq Q_1) = 0.25$** "المساحة المحصورة تحت الربع الاول"
- **$P(X \leq Q_2) = 0.5$** "المساحة المحصورة تحت الربع الثاني"
- **$P(X \leq Q_3) = 0.75$** "المساحة المحصورة تحت الربع الثالث"

[3]



- **$P(X \leq P_k) = k\%$** "المساحة المحصورة تحت نسبة معينة"

[4]