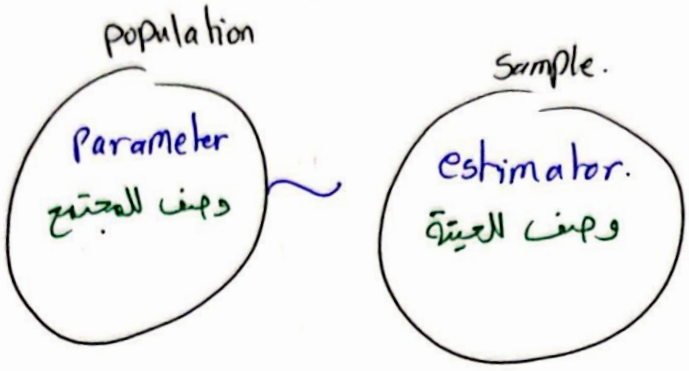
* Chapter (1) full-summary :-

Data sets
- population. → كل المجتمع
- sample → جزء من المجتمع

→ We have two important terms :-



* estimator (statistic) may change by changing the sample taken but parameter remain the same for the population.

statistics.
- Descriptive stat. "Describing data by numbers or graphs"
- Inferential stat. "drawing inferences & conclusions"

Data types
- Qualitative (نوعية) "categorical or non-numirical"
- Quantitative (كمية)
  - Countable → Discrete
  - Measurable → continuous.

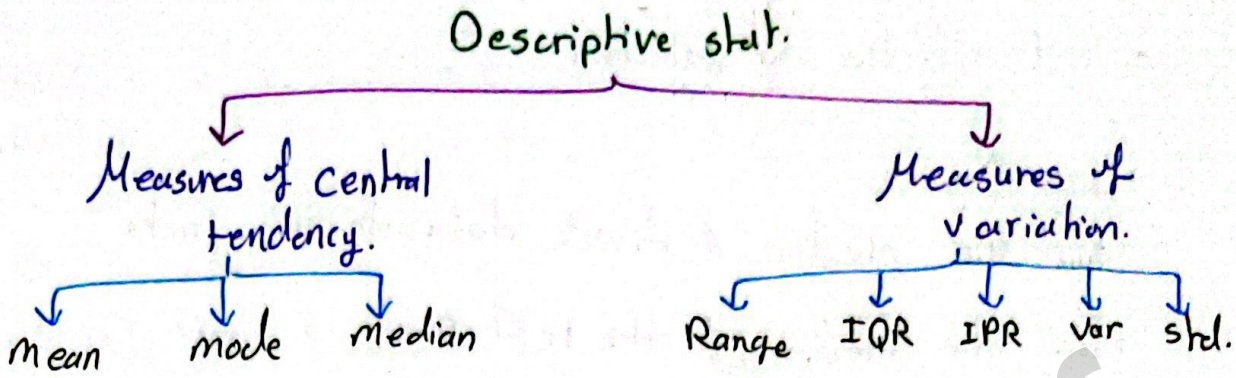
Measurement scales

- **Nominal** — "categorical & No ordering"
- **ordinal** — "categorical & ordered."
- **Interval** — "ordered, meaningful differences & has no true zero"
- **Ratio** — "ordered, meaningful differences & has true zero"

Level	Put data in category	Arrange data values	subtract data values	write values as multiple from each other
Nominal	Yes	NO	No	No
ordinal	Yes	Yes	NO	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes.

← مهم جدا الجدول .

* chapter(2): Descriptive statistics (full-summary).



* Measures of central tendency:-

[1] The Mean:- → Affected by outliers.

1. Raw data 2. stem & leaf.

$$\bar{x} = \frac{\Sigma x_i}{n}.$$

3. frequency dis:-

→ add f.x, then:

$$\bar{x} = \frac{\Sigma f.x}{n}, \quad n = \Sigma f$$

Weighted mean:

$$\bar{x} = \frac{\Sigma x.w}{\Sigma w}.$$

4. grouped frequency distribution:-

→ add x = midpoint = $\frac{L + U}{2}$.

→ add f.x, then:

$$\bar{x} = \frac{\Sigma f.x}{n}, \quad n = \Sigma f$$

Note: $\Sigma x_i = n.\bar{x}$.

[2] The Mode: → the best when data is qualitative

1. Raw data 2. stem & leaf 3. frequency dis:

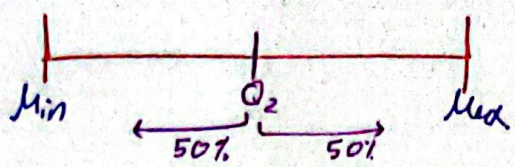
→ take the value that occurs mostly.

4. grouped frequency dis:

Class of maximum freq. → modal class & mod is the midpoint

[1]

3 The Median (Q_2):



The value in the middle of data set & not affected by outliers.

1. Raw data.

→ Arrange data values.

→ $Q_2 = \frac{n}{2}$ { fraction → next int
whole no. → $\frac{k^{th} + (k+1)^{th}}{2}$.

2. stem & leaf.

data is arranged & use same formula & steps.

3. frequency distribution.

→ Add c.f & Intervals. (c.f جدول التكرار)

→ $Q_2 = \frac{n}{2}$ { fraction → next int.
whole no. → $\frac{k^{th} + (k+1)^{th}}{2}$.

4. grouped frequency distribution.

→ Add c.f & URB.

→ $Q_2 = \frac{n}{2}$. (Don't Apply The Rule)

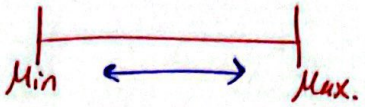
*Note: class length = C.L = URB - LRB.

* The Relative frequency :-

1 $R.f = \frac{f}{\Sigma f}$. 2 $\Sigma R.f = 1$ ← 1 = كل شيء جميع

* Measures of Variation :-

1 The Range :-



1. Raw data + 2. stem & leaf + 3. frequency distribution

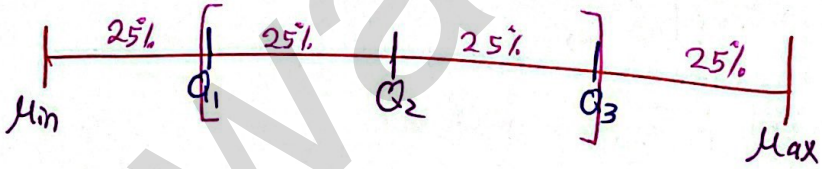
Range = Max - Min

4. grouped frequency distribution :-

Range = Max(URB) - Min(LRB).

* Range is affected by outliers.

2 Inter-quartile-range (IQR) :-



→ takes the 50% of data that falls in the middle.

→ not affected by outliers.

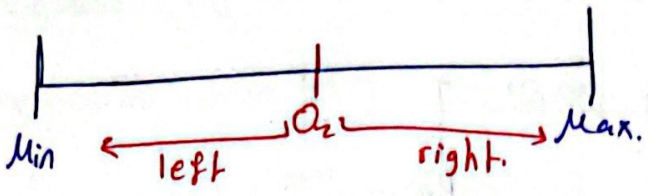
→ $IQR = Q_3 - Q_1$

Q_1 : 1st quartile (lower), 25% of data below it.

Q_3 : 3rd quartile (upper), 75% of data below it

* Mid quartile = $\frac{Q_3 + Q_1}{2}$.

1. Raw data :-



الطريقة الجديدة و المعتمدة.

find the median of left part → call it Q_1

find the median of right part → call it Q_3.

2. Stem & leaf :-

* حوّل الجدول إلى Raw data واتبّع نفس خطوات الـ Raw data *

3. Frequency distribution :-

→ Add c.f & Intervals.

→ $Q_1 = \frac{n}{4}$

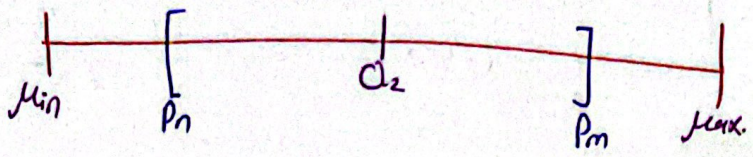
$Q_3 = \frac{3n}{4}$

→ fraction → next int.

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

$IQR = Q_3 - Q_1$.

3 Inter-Percenhile-ranage (IPR):



$IPR = P_m - P_n$ ← هو الفرق بين أي نسبتين

* Percenhile (P_K):

1. Raw data:-

→ Arrange data values.

→ $P_K = \frac{K}{100} \cdot n$ { fraction → next int
whole no. → $\frac{K^{th} + (K+1)^{th}}{2}$.

2. Stem & leaf:-

→ As Raw data, but table give arranged values
لو القيم مرتبة و جاهزة بس لازم نطبق القاعدة.

3. frequency distribution:-

→ Add C.f & Intervals

→ $P_K = \frac{K}{100} \cdot n$ { fraction → next int
whole no. → $\frac{K^{th} + (K+1)^{th}}{2}$.

* Deciles (D_k):

we divide data into 10 equal parts : $D_1, D_2 \dots D_{10}$.

اسهل طريقة هي ناخذ أعاملهم على إنهم percentiles ونطبق قوانينهم

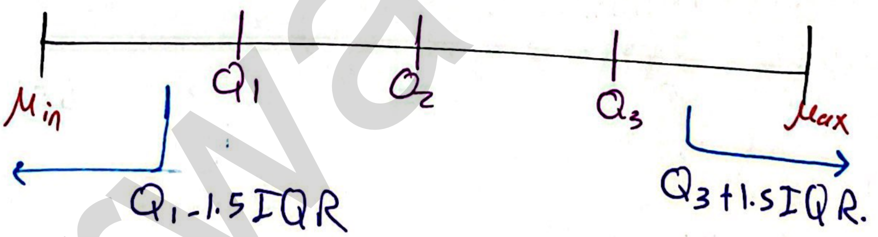
$D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30}$ . . . . $D_9 = P_{90}$.

* Fractiles :- numbers divide data into equal parts.

Fractiles.

- Quartiles
  1. 4 - Parts
  2. Q_1, Q_2, Q_3
- Percentiles
  1. 100 - Parts
  2. $P_1, P_2 \dots P_{99}$
- Deciles.
  1. 10 - parts
  2. $D_1, D_2 \dots D_{10}$.

* Outliers :



less than $Q_1 - 1.5\,IQR$.

More than $Q_3 + 1.5\,IQR$.

4+5 Variance + standard deviation:

→ Deviation = $x_i - \bar{x}_i$

→ for sample: var = s^2 & std = s

for population: var = σ^2 & std = σ.

1. Raw data & stem & leaf:

population variance	sample variance.
1) $\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{N}$	1) $s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$
2) $\sigma^2 = \frac{\Sigma x^2}{N} - (\mu)^2$	2) $s^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$
std = $\sigma = \sqrt{\text{variance}}$	std = $s = \sqrt{\text{variance}}$.
μ: population mean	$\bar{x}$: sample mean.
σ^2: population variance	s^2: sample variance.
N: population size	n: sample size
	* Σx^2: sum of squares.

3. frequency distribution:

→ for population: $\sigma^2 = \frac{\Sigma f \cdot (x_i - \mu)^2}{N}$ ← للمعرفة فقط

→ for sample:

1. $s^2 = \frac{\Sigma f (x_i - \bar{x})^2}{n-1}$ $\boxed{n = \Sigma f}$

2. $s^2 = \frac{\Sigma f \cdot x^2}{n-1} - \frac{(\Sigma f \cdot x)^2}{n(n-1)}$

→ Add f.x / x^2 / $f.x^2$.

std = $\sqrt{\text{variance}}$.

4. grouped frequency distribution :.

→ for sample :-

[1] $S^2 = \frac{\Sigma f.(x_i - \bar{x})^2}{n-1}$

[2] $S^2 = \frac{\Sigma f.x^2}{n-1} - \frac{(\Sigma f.x)^2}{n(n-1)}$ $\quad n = \Sigma f$

Std = $\sqrt{\text{variance}}$.

→ Add x = mid point = $\frac{\text{lower} + \text{upper}}{2}$.

→ add $f.x$ / x^2 / $f.x^2$.

→ Apply the formula.

* Note : When all data values the same, all measures of variation equal zero. * باقي تمنت *

Range = IQR = IPR = Variance = Std = 0.

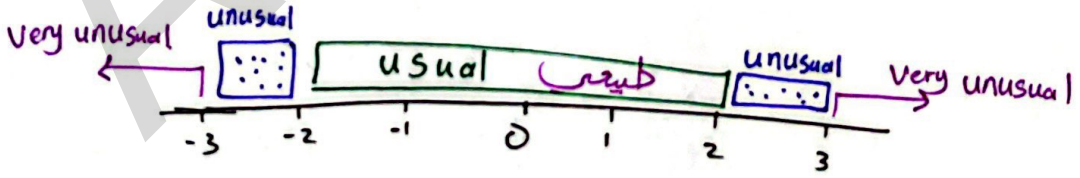
* Comparison two collections :.

[1] Z-score :.

$Z = \frac{x - \bar{x}}{s}$

- Positive. → above the mean.
- Zero → on the mean.
- Negative → below the mean.

ب كل ما زاد يكون أفضل.



← اذا الطالب جاب z-score بين 2- و 2 يكون وضع طبيعي

← اذا جاب بين 2- و 3- أو بين 2 و 3 يكون مش طبيعي

← اذا جاب أقل من 3- أو أكثر من 3 يكون بصورة مش طبيعي.

2 coefficient of variation :

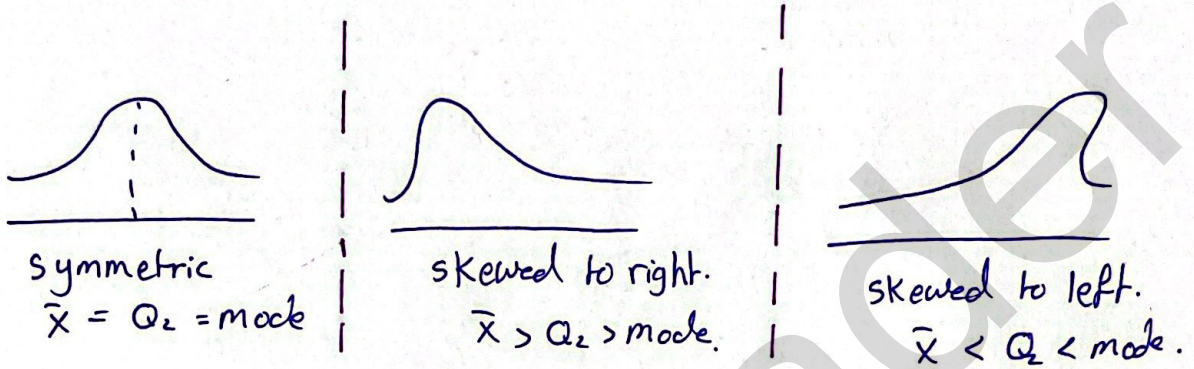
$$C.V = \frac{S}{\bar{X}} * 100\% \rightarrow$$

* كل ما زاد بزيد التشتت .

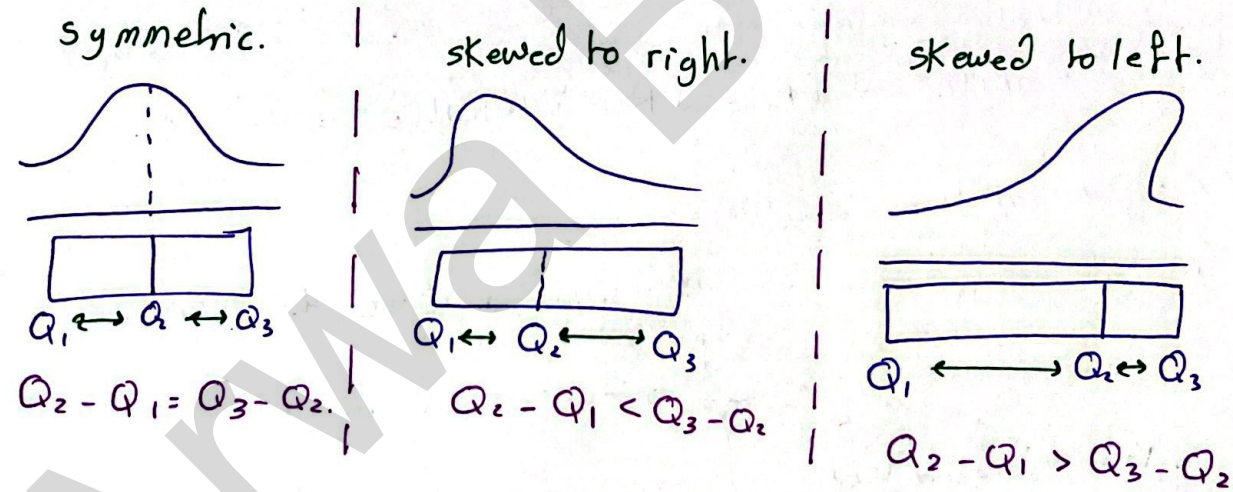
* Skewness :

we have 3 Methods :-

1 By Measures of central tendency :-



2 By Quartiles (Box-plot) :-



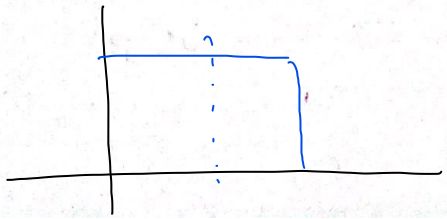
3 By Pearson's Index :-

موضوع جديد

→ compute $P = \frac{3(\bar{X} - Q_2)}{S}$

- $P > 0 \rightarrow$ skewed to right.
- $P < 0 \rightarrow$ skewed to left
- $P = 0 \rightarrow$ symmetric.

* Note :



uniform ≡ symmetric

* Linear Transform (coding) :- الموضوع مطلوب والله

$\underset{\text{new measure}}{y} = a \cdot \underset{\text{old measure}}{x} + b$, a, b : any numbers.

1 For Measures of central tendency :- * يتأثر بالضرب والجمع

1) $\bar{y} = a \cdot \bar{x} + b$. 2) $Q_2(y) = a \cdot Q_2(x) + b$ 3) $mode(y) = a \cdot mode(x) + b$

2 For Measures of variation :- * يتأثر بالضرب فقط وبعده موجب

1) $Range(y) = |a| \cdot Range(x)$ 2) $IQR(y) = |a| \cdot IQR(x)$

3) $IPR(y) = |a|\ IPR(x)$ 4) $S_y = |a| \cdot S_x$.

5) $S_y^2 = a^2 \cdot S_x^2$.

3 For other Measures :-

1) $P_K(y) = a \cdot P_K(x) + b$, $a > 0$.

$P_K(y) = a \cdot P_{(100-K)}(x) + b$, $a < 0$.

* اذا a موجبة باخد نفس النسبة أما اذا سالبة باخد متممة النسبة

2) $Q_1(y) = a \cdot Q_1(x) + b$, $a > 0$.

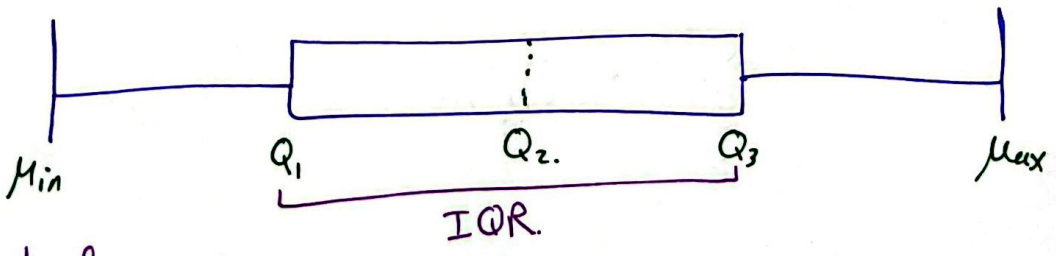
$Q_1(y) = a \cdot Q_3(x) + b$, $a < 0$.

3) $Q_3(y) = a \cdot Q_3(x) + b$, $a > 0$.

$Q_3(y) = a \cdot Q_1(x) + b$, $a < 0$.

* Data representation :-

1 Box & whiskers plot :-



↳ five-number-summary : Min, Q_1, Q_2, Q_3, Max.

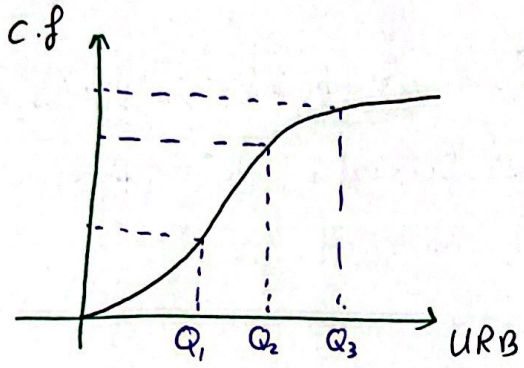
2 Pie-chart :- → represents X & f.

find the angles :-

$$\theta = \frac{f}{\Sigma f} * 360^\circ.$$

3 c.f curve (O-give) :- ٥٢

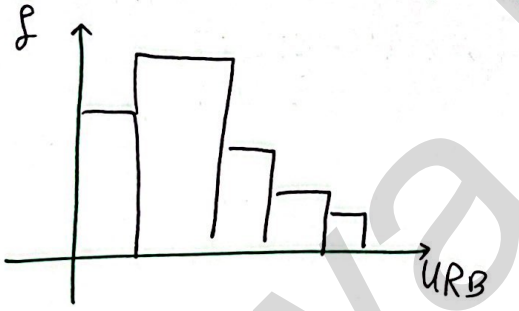
represent c.f & URB.



٢٥ كثير نتنبه كيف نرجع الرحمة لجدول.

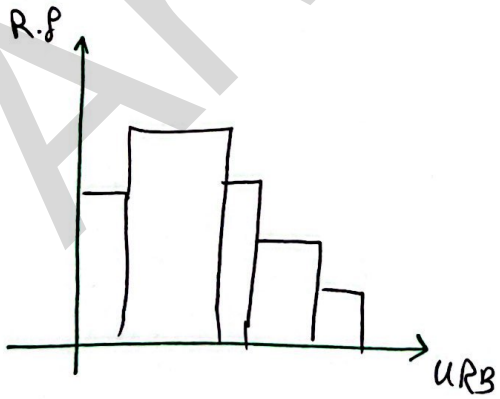
4 Histogram :-

represents URB & frequency. ٥٠



٢٥ نعرف كيف نرجع الرحمة إلى جدول.

Sometimes we represent. URB & relative frequency.



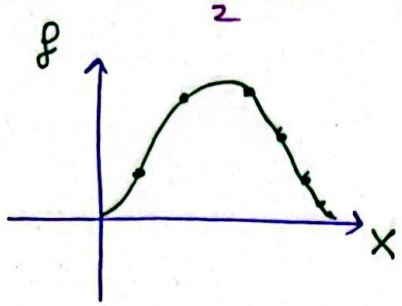
بعض الاحيان يمثل الـ U.R.B مع التكرار النسبي.

انتبه على y-axis شو بمثل.

5 Polygon :-

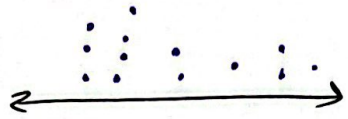
represents X & frequency , X is the midpoint.

$X = \frac{\text{lower} + \text{upper}}{2}$



6 Dot-plot :-

represents X & frequency.

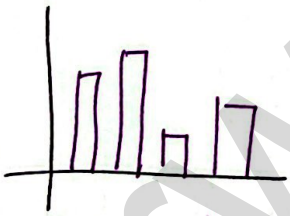


7 Bar-graph & Pareto chart :-

→ represents X & frequency.

→ In Pareto-chart, we put frequency in ascending order

لم يرتب القيم تصاعدي وبعدها بمثل.



Bar-graph

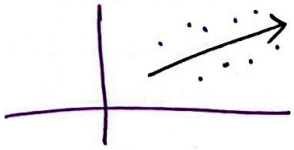
Pareto-chart.

8 Time-series-chart :-

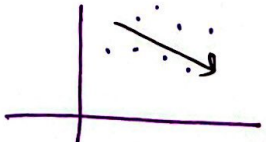
من اسمه يعني بتراقب زيادة أو نقصان خلال مدة زمنية.

9 Scatter-diagram :-

represent relationship between X & y.



Positive correlation



negative correlation.

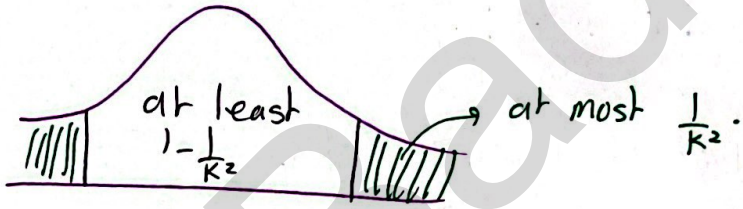
* Notes :-

→ For qualitative data (نوعية), we can use :-
Pie chart , Bar graph & Pareto chart only.

* بالامتحان ما راح ترسم ، فقط حيكون مطلوب توجد شي معين من رسمة لهيك لازم تعرف كيف ترجع الرسمة لجدول.
[شيك اسئلة التيست بانك والـ sheets.]

* Chebyshev's theorem :-

The proportion of observations within K standard deviation from the mean :-



For any set of data :

At least $1-\frac{1}{K^2}$ within $(\bar{x}-k.s, \bar{x}+k.s)$

At most $\frac{1}{K^2}$ outside $(\bar{x}-k.s, \bar{x}+k.s)$

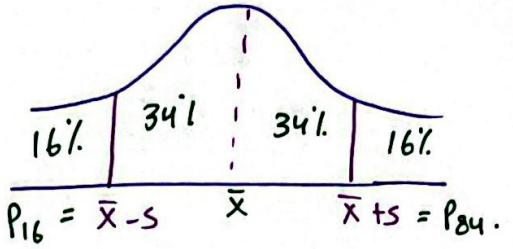
no. of obs. (عدد مشاهدات) ⇆ Percentage (نسبة) ⇆ K ⇆ Interval (فترة)

* number of obs = percentage * n.
- whole no. → ok
- fraction → at least → round up.
- fraction → at most → round down.

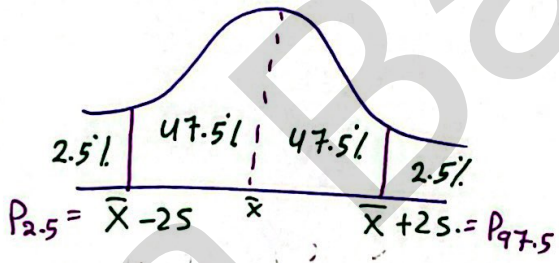
* Empirical Rule :-

For Bell-shaped data only.

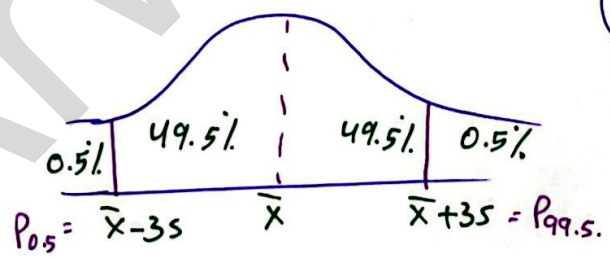
1 For $K=1 \rightarrow (\bar{x}-s, \bar{x}+s)$ contains at least 68%

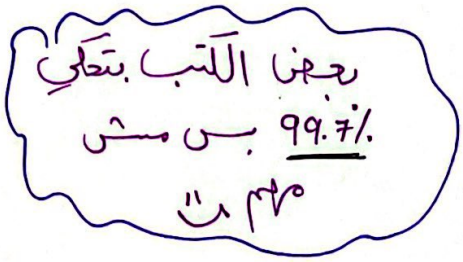


2 For $K=2 \rightarrow (\bar{x}-2s, \bar{x}+2s)$ contains at least 95%



3 For $K=3 \rightarrow (\bar{x}-3s, \bar{x}+3s)$ contains at least 99%





عدد المشاهدات ⟷ نسبة ⟷ قيم ⟷ Interval
no. of obs ⟷ Percentage ⟷ K ⟷ فترة

* number of observations = percentage * n.

* Chapter (3) :- Elements of probability

→ Event : Any subset of the sample space, denoted by A, B, C....

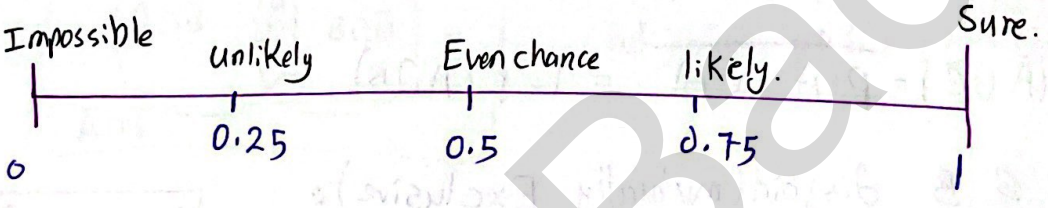
① Simple. → one element only.

② Combined → two or more.

③ Sure → All the sample space.

④ Impossible. → No elements of S.

$$P(A) = \frac{N(A)}{N(S)}$$



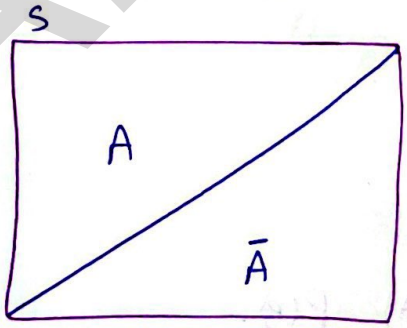
* Note : Any has prob. less than 0.05 called unusual.

→ Facts :

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

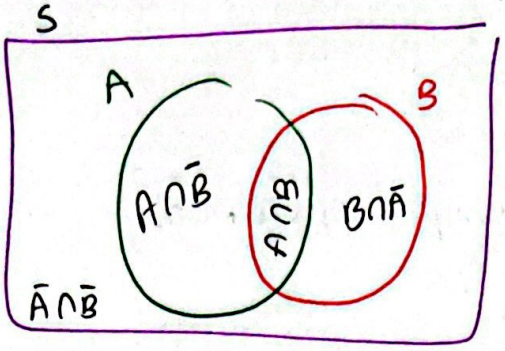
* Rules of probability :-

1)



1 $P(A) = 1 - P(\bar{A})$ OR $P(\bar{A}) = 1 - P(A)$.

2)



تبديلية

$A \cap B = B \cap A$

$A \cup B = B \cup A$.

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

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

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

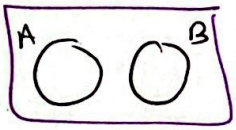
5 Demorgan Law's :-

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

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

6 A & B disjoint (mutually Exclusive) :

$P(A \cap B) = 0$. التقاطع صفر.



7 A & B Independent :

$P(A \cap B) = P(A) * P(B)$ التقاطع حاصل ضربهم

8 Conditional probability :-

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

given if

* If they're independent :-

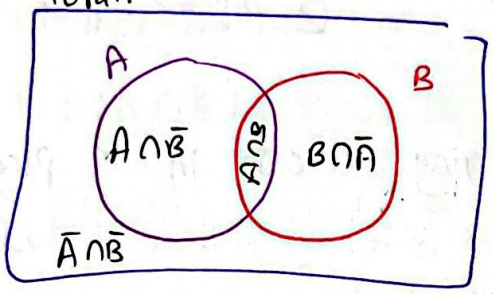
$P(A|B) = P(A)$ & $P(B|A) = P(B)$.

* prob. table :

	A	$\bar{A}$	total
B	$P(A \cap B)$	$P(B \cap \bar{A})$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
tot	$P(A)$	$P(\bar{A})$	1

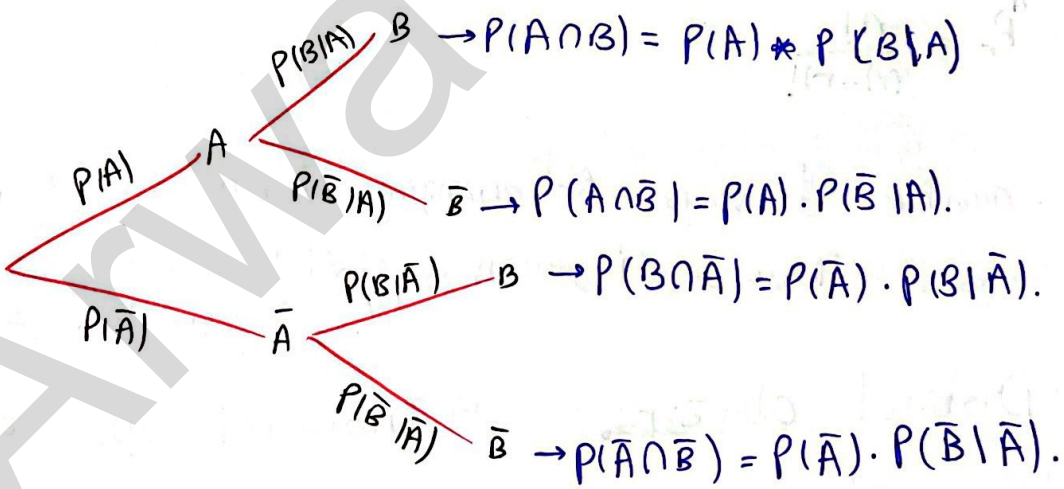
ممكن الجدول يكون معطى أرقام فقط.

* Venn-diagram:



لما ينقسم المجموعة إلى جزئين بينهم عناصر مشتركة.

* Tree diagram :



← لما ينقسم الفضاء العيني مرتين

← لما اختار عنصرين

← لما انفذ تجربة على مرحلتين

* Counting Rules: يا ساتر يارب

[1] Multiplication principle :-

n_1 | n_2

total $n = n_1 * n_2$.

[2] Factorial :-

$n! = n(n-1)(n-2)\dots(2)(1)$.

* Note : ① $0! = 1$ ② $1! = 1$ ③ $2! = 2$.

$n!$: number of ways for arranging n obj in R places.
لترتيب عناصر في أماكن ، عدد العناصر مساوي لعدد الاماكن وماعندي شروط على الخانات.

[3] Permutation :-

$$^nP_r = \frac{n!}{(n-r)!}$$

^nP_r : number of ways for arranging n obj in r places.
لترتيب عناصر في أماكن ، عدد العناصر أكبر من عدد الاماكن وماعندي شروط على الخانات.

A) Distinct obj مختلفة :

$$^nP_r = \frac{n!}{(n-r)!}$$

B) Indistinct obj. مكررة

$$P = \frac{n!}{n_1!\, n_2! \dots n_r!}$$

4 combination :

$$^nC_r = \binom{n}{r} = \frac{n!}{(n-r)! * r!} = \frac{^nP_r}{r!}$$

^nC_r: number of ways for selecting r obj from n obj

لاختيار عناصر بدون ترتيب وبدون إرجاع

* اكتب لنفسك السؤال اذا بطلب احتمالية أو عدد الطرق .

* اذا في شروط على الخانات بطلع الناتج ارقام مضروبة ببعض

*Note: $\binom{n}{r} = \binom{n}{n-r}$.

GOOD LUCK!

T. Arwa M. Bader