* Chapter (1) full-summary :

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

→ We have two important terms :

population — Parameter — وصف المجتمع

sample. — estimator. — وصف العينة

* estimator (statistic) may change by changing the sample taken but paramete 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 كثير الجدول .

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

Descriptive stat.

- Measures of central tendency.
  - mean
  - mode
  - median
- Measures of variation.
  - Range
  - IQR
  - IPR
  - Var
  - std.

* Measures of central tendency:-

$\boxed{1}$ The Mean:- → Affected by outliers.

1. Raw data 2. stem & leaf

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

3. frequency dis:-

→ add $f.x$, then:

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

weighted mean:

$$\bar{x} = \frac{\sum x.\omega}{\sum \omega}.$$

4. grouped frequency distribution:-

→ add x = midpoint = $\frac{L+U}{2}$.

→ add $f.x$, then:

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

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

$\boxed{2}$ The Mode: → the best when data is qualitative

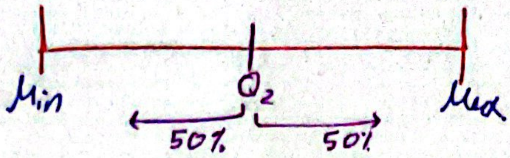
1. Raw data 2. stem & leaf 3. frequency dis:

→ take the value that occurs mostly.

4. grouped frequency dis:

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

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. (بالاعتماد على c.f)

→ $Q_2 = \frac{n}{2}$ { 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)

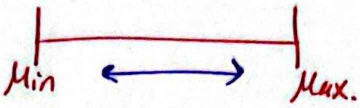
* Note: class length = C.L = URB - LRB.

* The Relative frequency :-

[1] $R.f = \frac{f}{\sum f}$. [2] $\sum R.f = 1$ → 1 = مجموع قيم ثابت

* Measures of Variation :-

[1] The Range :-



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

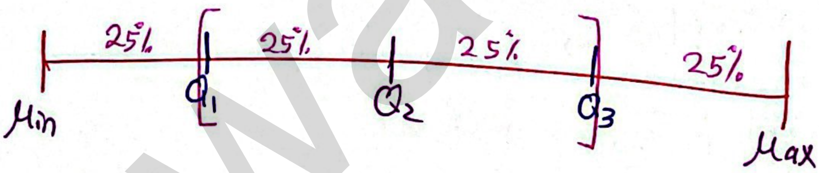
Range = Max − Min

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.

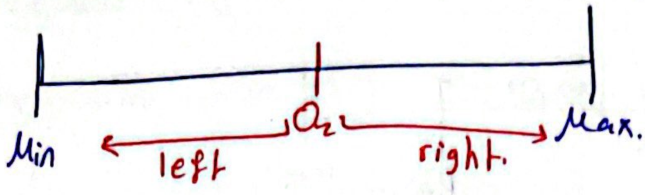
→ $IQR = Q_3 - Q_1$

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 والمعتمدة.

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 وطبّق نفس خطوات الـ Raw data *

3. Frequency distribution :

→ Add c.f & Intervals.

→ $Q_1 = \frac{n}{4}$

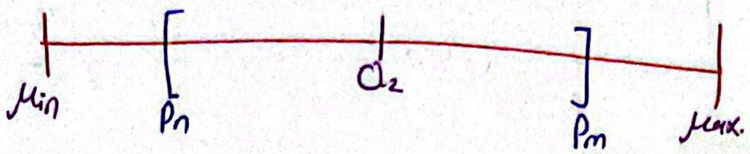
$Q_3 = \frac{3n}{4}$

→ fraction → next int.

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

$IQR = Q_3 - Q_1$.

3. Inter-percentile-range (IPR):



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

✱ Percentile (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
لـ 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 انو أتعاملهم على إنهم Percentiles ، وبطبق قوانينهم

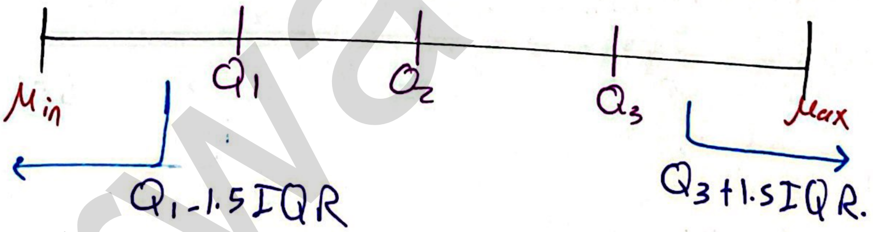
$D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30} \dots$ $D_9 = P_{90}$.

* 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 $Q_1 - 1.5IQR$.

More than $Q_3 + 1.5IQR$.

4 + 5 Variance + standard deviation:

→ Deviation = $x_i - \bar{x}_i$

→ for sample: Var $= s^2$ & std $= s$

for population: var $= \sigma^2$ & std $= \sigma$.

1. Raw data & stem & leaf:

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

3. frequency distribution

→ for population : $\sigma^2 = \dfrac{\Sigma f.(x_i - \mu)^2}{N}$ → للمجزة فقط

→ for sample,

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

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

→ Add $f.x$ / x^2 / $f.x^2$.

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

4. grouped frequency distribution :

→ for sample :-

$\boxed{1}$ $$S^2 = \frac{\sum f \cdot (x_i - \bar{x})^2}{n-1}$$

$\boxed{2}$ $$S^2 = \frac{\sum f \cdot x^2}{n-1} - \frac{(\sum f \cdot x)^2}{n(n-1)}$$

$\boxed{n = \sum f}$

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

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

→ add $f \cdot x$ / x^2 / $f \cdot x^2$.

→ Apply the formula.

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

Range = IQR = IPR = Variance = std = 0.

* Comparison two collections :-

$\boxed{1}$ Z-score :

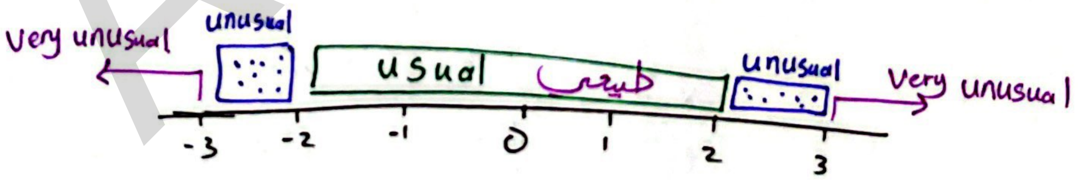
$$Z = \frac{x - \bar{x}}{S}$$

← كل ما زاد بكون أفضل.

→ Positive. → above the mean.

→ Zero → on the mean.

→ Negative → below the mean.



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

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

← اذا جاب أقل من -3 أو أكثر من 3 بكون بالمرة مش طبيعي.

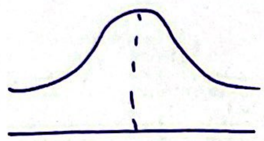
2 coefficient of variation :

$$C.V = \frac{s}{\bar{x}} * 100\% \rightarrow$$. كل ما زاد يزيد التشتت *

* Skewness :-

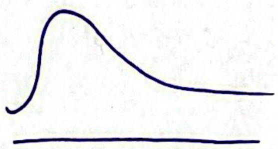
we have 3 Methods :-

1 By Measures of central tendency :-



symmetric

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



skewed to right.

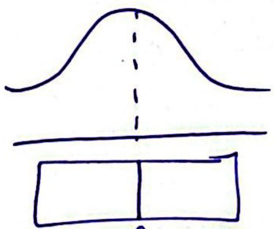
$\bar{x} > Q_2 > \text{mode}$.

skewed to left.

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

2 By Quartiles (Box-plot) :-

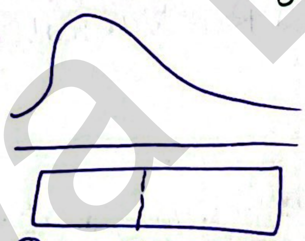
symmetric.



$Q_1 \leftrightarrow Q_2 \leftrightarrow Q_3$

$Q_2 - Q_1 = Q_3 - Q_2$.

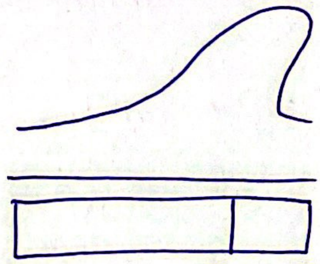
skewed to right.



$Q_1 \leftrightarrow Q_2 \longleftrightarrow Q_3$

$Q_2 - Q_1 < Q_3 - Q_2$

skewed to left.



$Q_1 \longleftrightarrow Q_2 \leftrightarrow Q_3$

$Q_2 - Q_1 > Q_3 - Q_2$

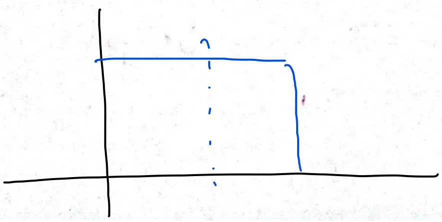
3 By Pearson's Index :-

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

الموضوع مطلوب والله.

* Linear Transform (coding):

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

[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 ، بعدد 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:

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

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

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

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

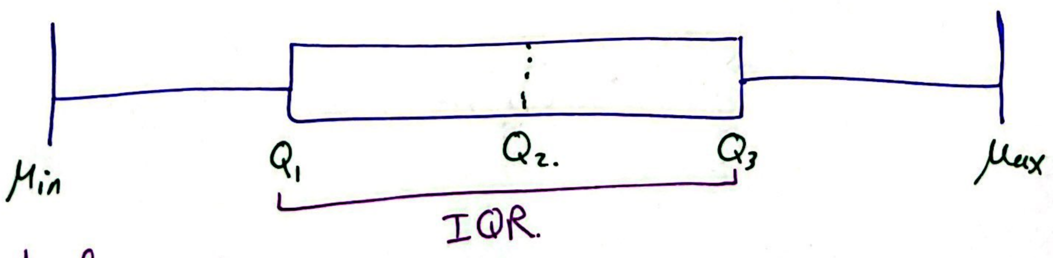
$Q_1(y) = a \cdot Q_3(x) + b \quad , \; a < 0.$

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

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

* Data representation:

[1] Box & whiskers plot:



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

2 Pie-chart: → represents X & f.

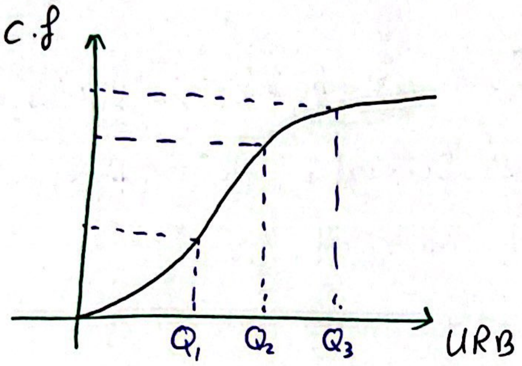
find the angles:

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

3 C.f curve (O-give):

مهم

represent C.f & URB.

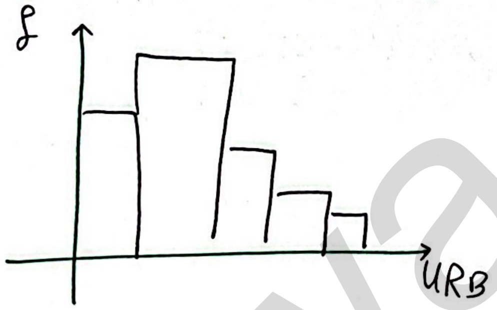


مهم كثير ننتبه كيف
نرجع الرسمة لجدول.

4 Histogram:

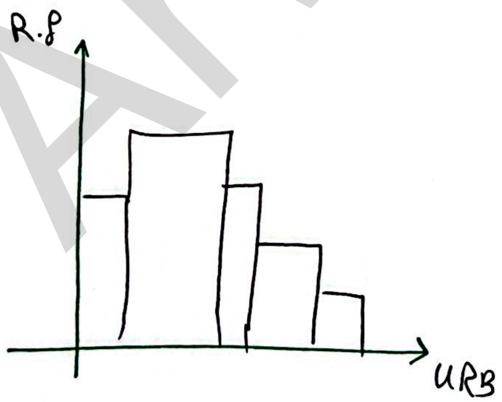
represents URB & frequency.

مهم.



مهم نعرف كيف
نرجع الرسمة إلى
جدول.

Sometimes we represent URB & relative frequency.



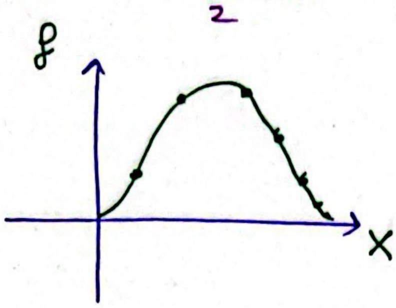
بعض الاحيان يمثل الـU.R.B
مع التكرار النسبي.

انتبه على y-axis شو بمثّل.

5 Polygon :-

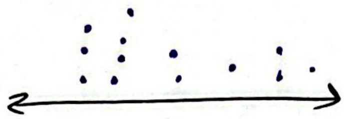
represents x & frequency , x is the midpoint.

$$x = \frac{\text{Lower} + \text{upper}}{2}$$



6 Dot-plot :-

represents x & frequency.



7 Bar-graph & Pareto chart :-

→ represents x & frequency.

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

لم يرتّب القيم تصاعدي وبعدها بمثّل.

Bar-graph

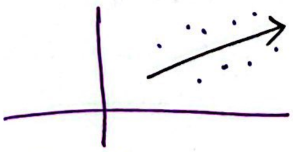
Pareto-chart.

8 Time-series-chart :-

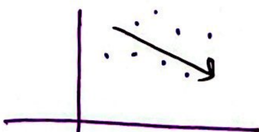
من اسمه يعني بنراقب زيادة أو نقصان خلال مدة زمنية.

9 Scatter-diagram :-

represent relationship between x & y.



Positive correlation



negative correlation.

* Notes :-

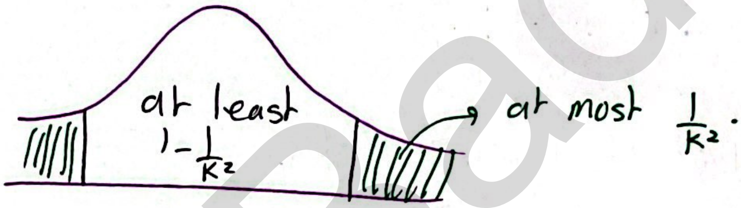
→ For qualitative data (نوعية), we can use :-
Pie chart, Bar graph & Pareto chart only.

* بالامتحان ما رح 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 $(\bar{x}-k\cdot s, \bar{x}+k\cdot s)$

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

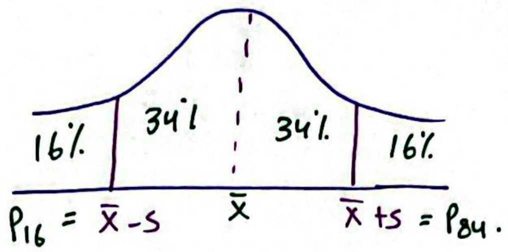
no. of obs. ⇆ Percentage ⇆ K ⇆ Interval
عدد مشاهدات — نسبة — فترة

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

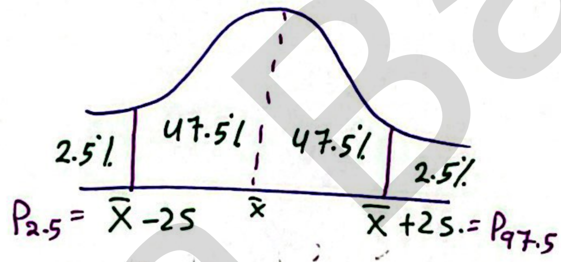
* Empirical Rule :-

For Bell-shaped data only.

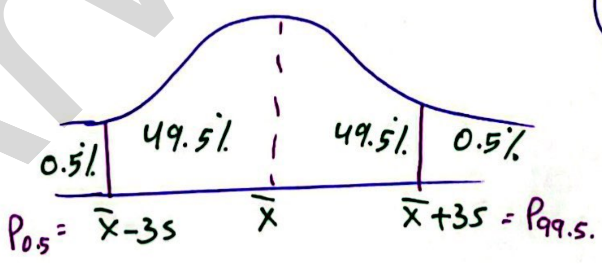
$\boxed{1}$ For $K = 1$ $\rightarrow (\bar{x} - s, \bar{x} + s)$ contains at least 68%



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



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



بعض الكتب بتحكي
99.7% بس مش
مهم :)

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 ....

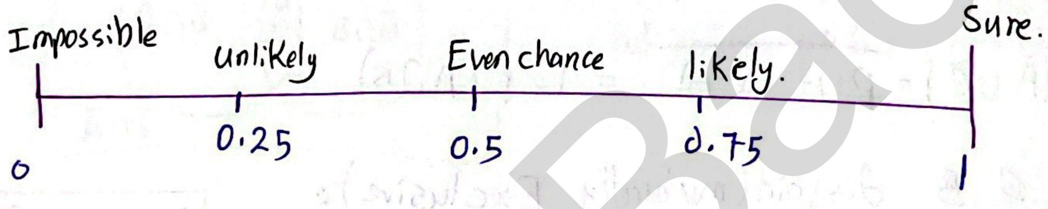
① simple : → one element only.

② Combined → two or more.

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

③ Sure → All the sample space.

④ Impossible. → No elements of S.



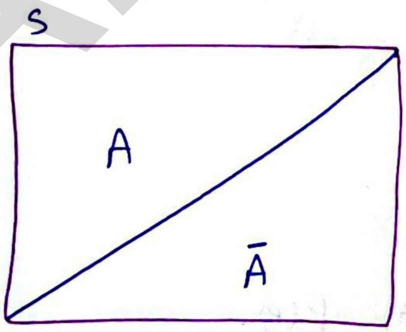
* Note : Any has prob. less than 0.05 called unusual.

→ Facts :

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

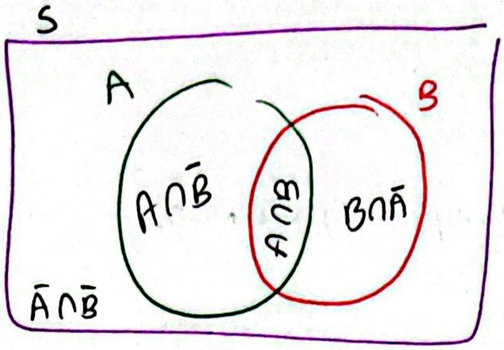
* Rules of probability :-

1)



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

2)



تبديلية

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

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

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

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

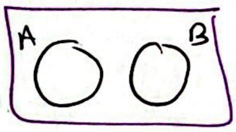
$\boxed{5}$ Demorgan Law's :-

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

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

$\boxed{6}$ A & B disjoint (mutually Exclusive) :

$P(A\cap B) = 0.$ التقاطع صفر.



$\boxed{7}$ A & B Independent :

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

$\boxed{8}$ conditional probability :

$$P(A\underset{\substack{\text{given}\\ \text{if}}}{|}B) = \frac{P(A\cap B)}{P(B)}.$$

* If they're independent :

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

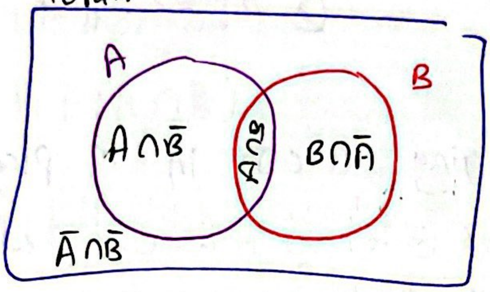
* prob. table :

	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.

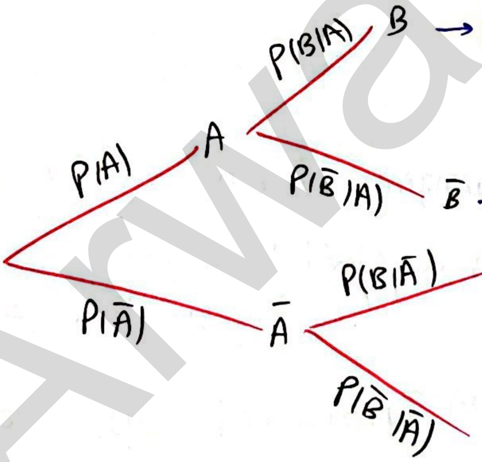
* Venn-diagram:-

total.



لما ينقسم المجموعة إلى جزئين بينهم عناصر مشتركة.

* Tree diagram :



$B \rightarrow P(A\cap B) = P(A) * P(B|A)$

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

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

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

← لما ينقسم الفضاء العيني مرتين

← لما اختار عنصرين

← لما انفذ تجربة على مرحلتين

* Counting Rules:

يا ساتر يارب

1 Multiplication principle :

$\boxed{n_1}$ $\boxed{n_2}$

total $n = n_1 * n_2$.

2 Factorial :

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

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

$n!$: number of ways for arranging n obj in R places.

له ترتيب عناصر في أماكن ، عدد العناصر مساوي لعدد الاماكن
وما عندي شروط على الخانات .

3 Permutation :-

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

${}^{n}P_{r}$: number of ways for arranging n obj in r places.

له ترتيب عناصر في أماكن ، عدد العناصر أكبر من عدد الاماكن وما عندي
شروط على الخانات .

A) Distinct obj مختلفة

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

B) Indistinct obj. متكررة

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

4 combination :

$$^{n}C_r = \binom{n}{r} = \frac{n!}{(n-r)! * r!} = \frac{^{n}P_r}{r!}$$

$^{n}C_r$: number of ways for selecting r obj from n obj

لـ اختيار عناصر بدون ترتيب وبدون إرجاع

* انتبه لنص 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

To Summer up:

- Random variable:

1) Discrete when it has a finite or countable number of possible outcomes that can be listed. قيم معدودة

2) Continuous when it has an uncountable number of possible outcomes, represented by an interval on a number line.قيم غير معدودة

Note: In any probability distribution we have $\sum \mathbf{P(x_i)} = \mathbf{1}$

- **Probability density function (P.D.F):**

A function $f(x) = P(X=x_i)$ is called a p.d.f if:

1. $f(x) \geq 0$, for all $x \rightarrow$ يعني اي احتمالية لازم تكون أكبر من صفر
2. $\sum f(x_i) = 1 \rightarrow 1 =$ مجموع الاحتمالات كلها

اول خطوة بالحل هي تحويل الاقتران الى جدول يحتوي قيم اكس و احتمالاتهم.

Note: The CDF of random variable X is defined as: $F(a) = P(X \leq a)$

- **The expected value:**

The weighted mean of all possible values of the random variable X, denoted by E(X) or μ.

$$\mu = E(X) = \sum X * P(X = x_i)$$

بضرب كل اكس باحتمالاها و بعدها بجمّعهم.

عادي يكون سالب او موجب او صفر.

- The variance:

Measures the distance from the mean.

$Var(x) = \sigma^2 = E(X-\mu)^2$ OR $Var(x) = E(X^2) - (E(X))^2$

Standard deviation= $\sigma = \sqrt{variance}$

ممنوع يكون سالب فقط موجب او صفر.

Note: $E(X^2) = Var(X) + (E(X))^2$ "Extremely important"

Properties of the expected value:	Properties of the variance:
$E(a) = a$	$Var(a) = zero$
$E(aX) = a\,E(X)$	$Var(aX) = a^2\,Var(X)$
$E(X \pm Y) = E(X) \pm E(Y)$	$Var(aX + b) = a^2\,Var(X)$
$E(g(x)) = \sum g(x) * P(X = x_i)$	$VAR(X \pm Y) = VAR(X) \pm Var(Y)$ Only if X and Y are independent.

- **Special discrete distributions:**

Binomial distribution $X \sim Bin(n, p)$	**Poisson distribution** $X \sim poi(\mu)$
Conditions: 1. Two outcomes. 2. Independent trials 3. Prob of success constant each time.	**Conditions:** 1. Occurrences independent 2. Occurrences proportional to the length of time. 3. Prob. Of an event is the same for each period.
n: number of trials p: prob. of success	$\boldsymbol{\mu}$ **: Average of occurrences**
$\mathbf{p.d.f}:\ P(X = x_i) = \binom{n}{x_i} p^{x_i} q^{n-x_i}$	$\mathbf{p.d.f}:\ P(X = xi) = \frac{e^{-\mu} * \mu^x}{x!}$
$\mathbf{P(X \le K)}$ use tables	$\mathbf{P(X \le K)}$ use tables
$\mathbf{E(X)} = \mu = n * p$	$\mathbf{E(X)} = \mu$
$\mathbf{Var(X)} = \sigma^2 = n * p * q$	$\mathbf{Var(X)} = \sigma^2 = \mu$
$\mathbf{Std(X)} = \sigma = \sqrt{variance}$	$\mathbf{Std(X)} = \sigma = \sqrt{\mu}$

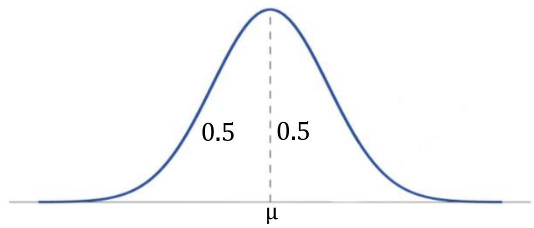
- **Binomial → poisson :**

  If $n \geq 30$ and $P < 0.1$ we approximate Binomial to Poisson by:

  $X \sim Bin(n, p) \rightarrow X \sim Poi(n \times P)$

Normal Distribution:

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

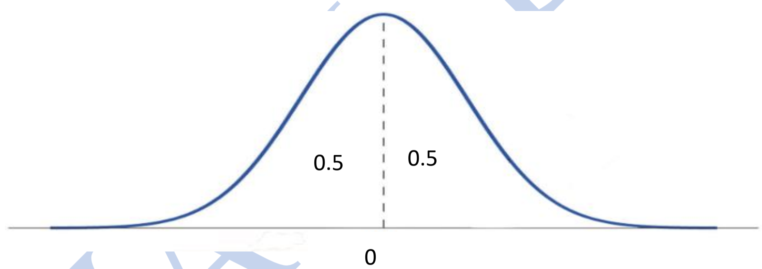


⚠ **A normal distribution has the following properties:**

1) Symmetrical ≡ bell-shaped
2) Mean = Median = Mode and are all located at the canter
3) Probability = area
4) Total area under the curve = 1

- **The standard normal:**

- It is a normal distribution with a mean of zero and variance of 1.
- $X \sim N(\mu, \sigma^2) \;\rightarrow\; Z \sim N(0, 1)$



1. $P(Z \le K) \rightarrow$ *we use tables*
2. $P(Z > K) = 1 - P(Z \le K)$
3. $P(Z > k) = P(Z < -K)$ "لأنه متماثل فبقدر اعكس اشارة الداخل و اقلب المتباينة"
4. $P(Z = k) =$ zero
5. $P(a < Z < b) = P(Z < b) - P(Z < a)$

✅ **Note:** $P(Z < a)$ = area

if area > 0.5 → a = +

if area < 0.5 → a = -

if area = 0.5 → a = 0

Note:

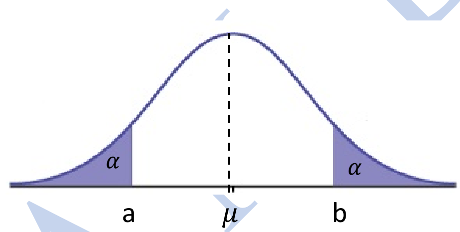
I. Any number more than 3.5, it's probability = 1
II. Any number less than -3.5, it's probability = zero

_ Transformation formulas:

$X \sim N(M, \sigma^2)$ we should transform X to Z	$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ we should transform X to Z , but how?
$Z = \frac{X - \mu}{\sigma} \rightarrow Z \sim N(0,1)$	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow Z \sim N(0,1)$

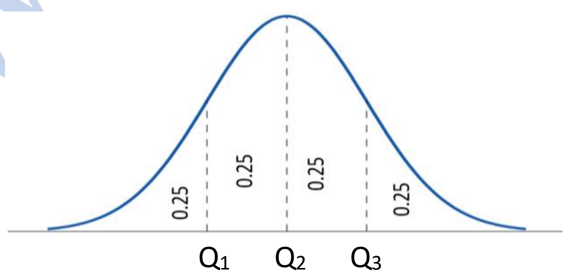
Very important notes:

[1]



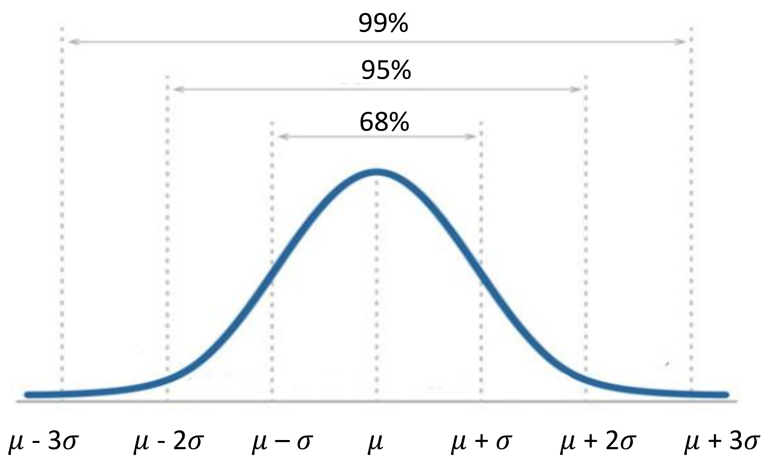
$$\therefore P(X < a) = P(X > b) \rightarrow \mu = \frac{a + b}{2}$$

[2]



- $P(X \le Q1) = 0.25$ "المساحة المحصورة تحت الربع الاول
- $P(X \le Q2) = 0.5$ "المساحة المحصورة تحت الربع الثاني
- $P(X \le Q3) = 0.75$ "المساحة المحصورة تحت الربع الثالث
- $P(X \le Pk) = k\%$ "المساحة المحصورة تحت نسبة معينة

[3]



- The probability that X lies within k standard deviation:

  $\therefore P(\mu - K.\sigma < X < \mu + K.\sigma)$

- **The distribution of sample proportion:**

$\widehat{P} \sim N\left(P, \dfrac{P(1-p)}{n}\right)$, so the mean is p and the variance is $\dfrac{P(1-p)}{n}$

$\therefore\ \widehat{P}$ is distributed as Normal with $\mu = P\ \&\ \sigma^2 = \dfrac{P(1-p)}{n}$

- How to convert $\widehat{P}$ to Z?

$$Z = \frac{\widehat{P} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$$

0798208683 Arwa M. Bader Principles of statistics-JU

Final Summery :

Parameters	Estimators
μ	$\overline{X}$
σ^2	S^2
σ	S
P	$\widehat{P}$

- **If $X \sim (\mu, \sigma^2)$, then we should transform to stndard normal by:**

$$Z = \frac{X - \mu}{\sigma}$$

- **Distribution of sample mean ($\overline{X}$):** $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

Is $\boldsymbol{\sigma}$ given?

Yes

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

No

$n \geq 30$

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

$n < 30$

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

- **Distribution of sample variance (S^2):**

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \rightarrow \sim X^2\,(n-1)$$

with d.f $= n - 1$, where n is the sample size

- **Distribution of sample proportion ($\widehat{P}$):**

$$\widehat{P} \sim N\left(P, \frac{P(1-p)}{n}\right)$$

$$Z = \frac{\widehat{P} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$$

- **Distribution of the difference between two means $(\bar{X} - \bar{Y})$:**

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2 \,, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Are σ_1 & σ_2 given?

Yes

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

No

$\sigma_1 \neq \sigma_2$

$$t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

d.f=Min{n-1,m-1}

$\sigma_1 = \sigma_2$

$$t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S.P\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

d.f = n+m-2

$$S.P = \sqrt{\frac{S_1^2(n-1) + S_2^2(m-1)}{n+m-2}}$$

- **Distribution of difference between two proportions $(\hat{P}1 - \hat{P}2)$:**

$$\hat{P}_1 - \hat{P}_2 \sim N\left(P_1 - P_2 \,, \frac{P_1(1-P_1)}{n} + \frac{P_2(1-P_2)}{m}\right)$$

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n} + \frac{P_2(1-P_2)}{m}}} \sim N(0,1)$$

⚠ ملاحظات هامة:

1) جدول الـ Z بعطي اللي اقل بينما جدول الـ t و الـ X^2 بعطي اللي أكبر.
2) إشارة المساواة مش مهمة بجميع الثلاث جداول.
3) إنتبه إنك مع تستخدم الجداول الصحيحة أثناء الحل، يعني إنتبه لعنوان الجدول t-table, X^2-tables.... 😊

Population parameter	Point estimation	C.I for parameter	Test statistics
μ	$\bar{X}$	**1) σ known** $\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ $E = Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ $S.E = \frac{\sigma}{\sqrt{n}}$ $\bar{X} = \frac{L+U}{2}$ **2) σ unknown:** 1. $n \geq 30$ $\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}$ 2. $n < 30$ $\bar{X} \mp t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ $Z = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$ $t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$
$\mu_1 - \mu_2$ **(Independent)**	$\bar{X} - \bar{Y}$	**1) σ_1 & σ_2 known** $(\bar{X} - \bar{Y}) \mp Z_{\alpha/2} * \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$ **2) σ_1 & σ_2 Unknown:** **1. $\sigma_1 \neq \sigma_2$** $(\bar{X} - \bar{Y}) \mp t_{\alpha/2} * \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$ d.f = Min{n-1,m-1} **2. $\sigma_1 = \sigma_2$** $(\bar{X} - \bar{Y}) \mp t_{\alpha/2} * S.P * \sqrt{\frac{1}{n} + \frac{1}{m}}$ $S.P = \sqrt{\frac{S_1^2(n-1)+S_2^2(m-1)}{n+m-2}}$ d.f = n+m-2	$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$ $t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$ $t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S.P\sqrt{\frac{1}{n} + \frac{1}{m}}}$
$\mu_1 - \mu_2 = \mu_d$	$\bar{d}$	$\bar{d} \mp t_{\alpha/2} * \frac{s.d}{\sqrt{n}}$	$t = \frac{\bar{d} - \mu_d}{s.d/\sqrt{n}}$

		- $\bar{d} = \frac{\sum d_i}{n} = \frac{\sum (X_i - Y_i)}{n}$ - S.d = $\sqrt{\frac{\sum d_i^2}{n-1} - \frac{(\sum d_i)^2}{n*(n-1)}}$ - Error of estimation E = $t_{\alpha/2} * \frac{s.d}{\sqrt{n}}$	
P	$\hat{P}$	$\hat{P} \mp Z_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$	$Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$
P1 – P2	$\hat{P}_1 - \hat{P}_2$	$(\hat{P}1 - \hat{P}2) \mp Z_{\alpha/2} * \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n} + \frac{\hat{P}_2(1-\hat{P}_2)}{m}}$	$Z = \frac{\hat{P}_1 - \hat{P}_2 - (P_1 - P_2)}{\sqrt{P^*(1-P^*)} * \sqrt{\frac{1}{n} + \frac{1}{m}}}$ $P^* = \frac{X+Y}{n+m}$
σ^2	S^2	$\left(\frac{(n-1)S^2}{X^2_{\alpha/2}}, \frac{(n-1)S^2}{X^2_{1-\alpha/2}}\right)$	$X^2 = \frac{S^2(n-1)}{\sigma_0^2}$
σ	S	$\left(\sqrt{\frac{(n-1)S^2}{X^2_{\alpha/2}}}, \sqrt{\frac{(n-1)S^2}{X^2_{1-\alpha/2}}}\right)$	Square the test

Sample size

$$n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 * \hat{P}(1-\hat{P})$$

- **Given by previous study "use it"**
- **Not given, assume $\hat{P} = 0.5$**

$$n = \left(Z_{\frac{\alpha}{2}} * \frac{\sigma}{E}\right)^2$$

⚠ Sample size should be a whole number, so if you got a fraction, round it up.

⚠ Length= 2*Error.