WITH NOTES

Introduction to research

Research Design
The 8 steps model

A. **Steps in planning research study:**

Step 1: Formulate a **research problem.**

Step 2: Research **design.**

Step 3: Construct **tools** for data collection.

Step 4: Select a **sample (size & method).**

Step 5: **Write research protocol (proposal).**

B. **Steps in conducting study:**

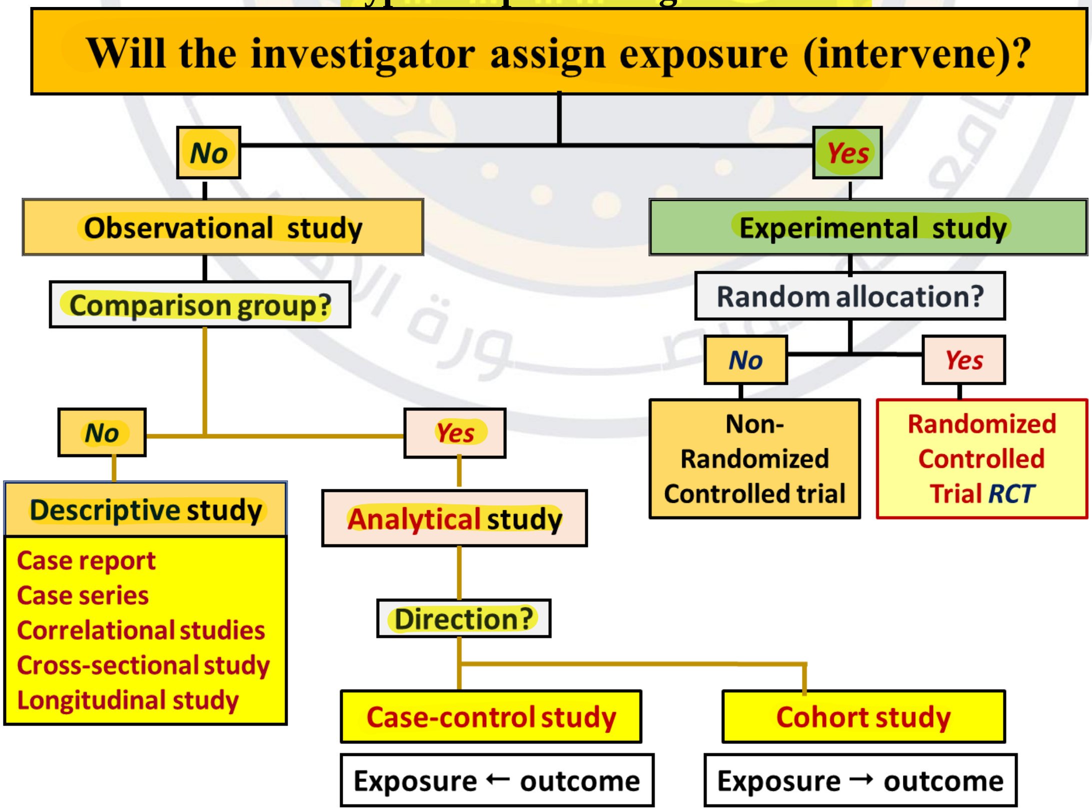
Step 6: **Collecting** data.

Step 7: **Processing** data.

Step 8: **Writing** research report.

Epidemiological Study :
- Study of:
Distribution of disease or health problem → Descriptive study
Determinants of disease or health problem → Analytical study
- Purpose: Application of this study for prevention & control

Types of epidemiological studies.


Will the investigator assign exposure (intervene)?

No
Observational study
Comparison group?
No
Descriptive study
Case report
Case series
Correlational studies
Cross-sectional study
Longitudinal study
Yes
Analytical study
Direction?
Case-control study
Exposure ← outcome
Cohort study
Exposure → outcome

Yes
Experimental study
Random allocation?
No
Non-Randomized Controlled trial
Yes
Randomized Controlled Trial RCT

Descriptive studies

Uses

☒ First phase in the epidemiological investigation.

☒ Describes the **pattern**, **characteristics** and **distribution** of a disease or health problem in the population.

☒ Give data about:
- When the disease occurs (Time).
- Where the disease occurs (Place).
- Who is getting the disease (Person).

☒ **Formulating (not testing) research hypotheses** (It is the 1st step in the search for determinants or risk factors).

Types of descriptive studies:

OSPE + MCQ :
- Case scenario ويسأل عن نوع الدراسة
- Outcome of this study ونحسبها

1- Case Reports:

Presentation of a single case that is newly reported or has unique finding e.g.

☒ Newly described disease.

☒ Unexpected or new therapeutic effect.

☒ Link between diseases.

2- Case Series:

☒ Describe a number of similar cases with a given disease in one report.

☒ May describe unusual variations of a disease and May indicate the start of an epidemic.

☒ A major trigger for further research.

3- Ecological studies:

☒ **Looking for** associations (**correlation)** between **exposures & outcomes** in population rather than in individuals.

☒ **Use already collected population data** (e.g. vital statistics, censuses and national health surveys).

☒ **Comparing populations in different places at the same time or in a time series by comparing the same population in one place at different times.**

☒ **Examples:**

- Mortality from CHD & per capita sales of cigarettes.
- High incidence of MI & consumption of dietary fat & fast food.
- Negative correlation between access to efficient ANC & maternal mortality rate.

4- Cross-Sectional Studies (Prevalence studies):

☒ **It is observational** study that **carried out once** (snapshot of a population) at **a single point in time.**

☒ Both **exposure** (risk factors) and **outcome** (diseases) **are present** (we cannot determine if exposure preceded disease or not).

سؤال مهم أوي MCQ

☒ It measures **prevalence**, not incidence of disease.

☒ Used to study conditions that are relatively *frequent with long duration* of expression (non-fatal, chronic conditions).

$$\text{Prevalence rate} = \frac{\text{Number of new and preexisting cases of disease during specified period}}{\text{Population examined during the specified period}} \times 10^{n}$$

جملة generate hypothesis مهمة أوي
وموجودة في كل أنواع الـdescriptive studies

الجدول دا حفظ
SAQ + MCQ + OSPE

Advantages of Cross-sectional Studies	Disadvantages of Cross-sectional Studies
☒ Used to study conditions that are relatively frequent with long duration (chronic conditions). ☒ Good for generating hypotheses about the cause of disease. ☒ Can estimate prevalence rates and exposure proportions in the population. ☒ Relatively easy, quick and inexpensive. ☒ No follow up, relatively easy, quick and inexpensive. ☒ It is the first step to develop evidence for causal association.	☒ It is not useful for studying: • Acute diseases. • Diseases with seasonal variations. • Highly fatal diseases. • Rare diseases. ☒ Can't estimate incidence rate. ☒ It gives very little information about the natural history of diseases. ☒ Cannot determine if exposure preceded disease or not. ☒ Not differentiate between causes of disease & factors associated with disease. ☒ Not provide solid evidence for causal association as it does not determine if really exposure preceded disease or not.

5- Longitudinal (incidence) studies:

- ☒ Repeated observations (follow-up) in same community over prolonged period to identify new cases of disease.
- ☒ Follow up and re-examination have the following problems:
  1. Loss to follow-up.
  2. Difficulty in maintaining standards and stability of clinical and laboratory examination over a long period of time.
- ☒ It is used to measure:
  1. Incidence rate.
  2. Natural history of dis. & its final outcome (case fatality, survival).
  3. Risk factors of disease.

Analytical studies

- Basic Question in Analytic Epidemiology: Are exposure & outcome (disease) linked?

مهمة جدااااا MCQ

☒ These studies are used to test an etiologic hypothesis such as smoking and Lung cancer; excess carbohydrates and obesity

= prove or disprove

☒ Analytical studies always require the use of control group.

- **Types of analytical studies:**

A. Case-control studies (Retrospective Studies)

☒ **Definition:**

It is an “observational” in which we assess the frequency of exposure to specific risk factor (suspected etiological factors) in patients who have developed a disease

☒ It is compared with that of controls who do not have the disease

☒ Case-control studies provide a relatively simple way to investigate causes of diseases, especially rare diseases

☒ The investigator is looking backward from the disease to a possible cause (retrospectively) by direct questioning and or extracted from clinical records

☒ **Steps to conduct case control study:**

1. **Identify the study group (cases)**: Define case & criteria for inclusion & exclusion of cases
2. **Identify controls:** (needed for comparison)

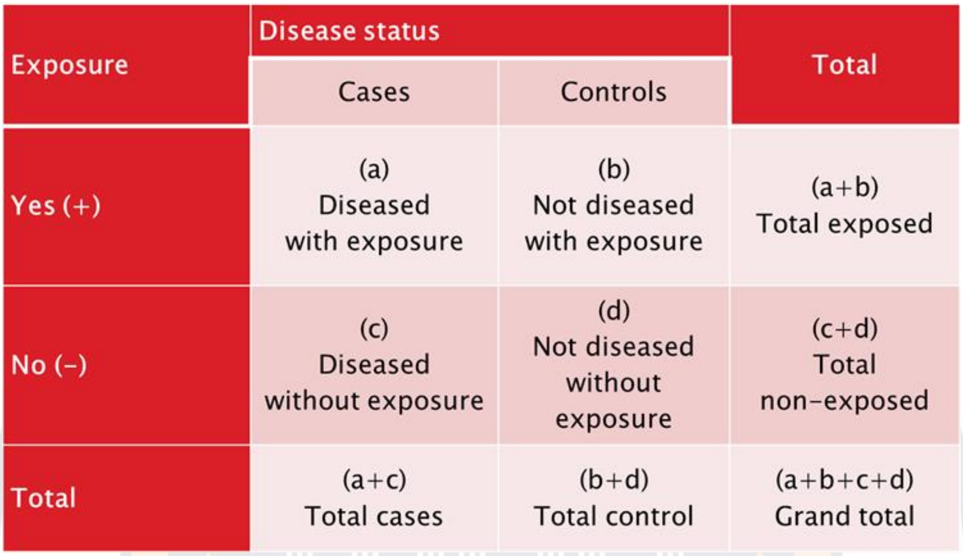
الشروط اللي لازم تتوفر في الcontrol group مهمة
وممكن تيجي سؤال

- Controls must be free from the studied disease.
- Controls must be matched with cases for certain characteristics known to influence the outcome of the disease (confounding factors) e.g. age, sex, social class …

(same range)

- Sources of controls:
    - ○ General population
    - ○ Hospital controls
    - ○ Special control series (Family members- friends- neighbors)

3. Summarize your data in 2x2 (association) tables:
    - Summarize frequencies of disease & exposure
    - Calculate association



Exposure	Disease status		Total
	Cases	Controls	
Yes (+)	(a) Diseased with exposure	(b) Not diseased with exposure	(a+b) Total exposed
No (−)	(c) Diseased without exposure	(d) Not diseased without exposure	(c+d) Total non-exposed
Total	(a+c) Total cases	(b+d) Total control	(a+b+c+d) Grand total

- Proportion of the exposed among cases (P1)= a/a + c
- Proportion of the exposed among control (P2) = b/b + d
- **Relative contribution = P1 – P2**

  It represents relative contribution of the suspected cause to the total frequency of the disease

مهمة جداااا

Outcome of case control :
- Odds ratio

- Odds Ratio = odd of exposure among cases / odd of exposure among controls

  Odds Ratio = a/c / b/d = ad/bc

  It is the indirect estimation of the risk

- **Interpretation of the odds Ratio (OR):**
    - **OR = 1** Exposure is not associated with outcome or disease.
    - **OR > 1** Increased exposure accompanies increased outcome
    - **OR < 1** Increased exposure accompanies decreased outcome.

      → Protective factor

☒ Advantages of case control studies: مهمة جدااا

1. Easy to carry out.
2. Quick & cheap.
3. Can be used in rare diseases.
4. Allows the study of several risk factors.
5. Useful in the study of disease with a long latency.
6. Does not require large samples.
7. Can prove hypothesis (Exposure & Disease are related).
8. Can estimate risk (odds ratio).

☒ Disadvantages of Case Control Study: مهمة جدااا

1. Cannot calculate prevalence or incidence rates.
2. Not useful in rare exposure.
3. Liable to bias or mistakes

N.B: Bias: Any systematic error in the design, conduct or analysis of a study that can result in wrong results. تعريف ال**bias** مهم

Analytic studies II

Cohort study

Definition:

☒ Cohort is a group having a common characteristic, e.g. a smoker's cohort means all are smokers in that group.

☒ An observational prospective (longitudinal or follow up) study in which we compare

☒ exposed group (individuals with a risk factor) with

☒ non exposed group (others without the risk factor)

☒ as regards the incidence of a disease over time.

Steps of cohort:

1. First we exclude cases of disease under investigation.
2. The free cohort, divided into 2 groups:
   - **Exposed group:** individuals exposed to risk factor.
   - **Control group:** individuals not exposed to this factor.
3. Both groups are followed up over a sufficient period of time. Therefore the cohort should be stable, cooperative & accessible to the investigator.
4. If the incidence of disease among exposed group is higher than its incidence among non exposed group, this supports the etiological hypothesis

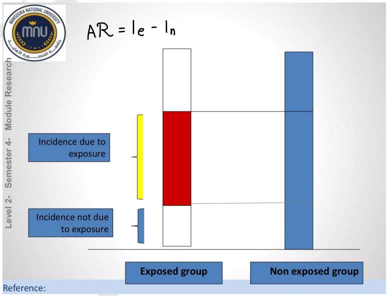
= test = prove or disprove

Basic analysis of cohort study:

Exposure	Disease status		Total
	Present	Absent	
Yes (+)	(a) with exposure with disease	(b) with exposure without disease	(a+b) Total exposed
No (-)	(c) Without exposure with disease	(d) without exposure Without disease	(c+d) Total non-exposed
Total	(a+c) Total with disease	(b+d) Total without disease	(a+b+c+d) Grand total

☒ **Basic analysis involves: calculation of:**

1. Overall incidence = a + c / a+b+c+d
2. Incidence rate among the exposed (Ie) =(a/a+b)
3. Incidence rate among the non-exposed (In) =(c/c+d)
4. Relative risk (RR) = Incidence among exposed (Ie)/ Incidence among non exposed (In)



RR answers the question: "How many times exposed person is at risk of developing disease compared to non-exposed?" مهمة جداااا

5. Attributable risk (AR) = I e – In

AR answers the question: "How much of the studied disease can be attributed to exposure“. “**Proportion of disease in a population that would be eliminated if risk factor is eliminated**”. مهمة جداااا

تعريف الAR مهم جداااا

☒ **Interpretation of Relative Risk (RR):**

- **RR = 1:** No association between exposure & disease.
- **RR > 1:** Positive association (increased risk) i.e. exposed group has higher incidence than non-exposed group.
- **RR < 1:** Negative association (protective effect) i.e. non-exposed group has higher incidence.

Example

	Developed CHD	Do not Develop CHD	Total
Smokers	**60**	**40**	**100**
Non-smokers	**20**	**80**	**100**

Answer

- ☒ Incidence in smokers = 60/100
- ☒ Incidence in non-smokers = 20/100
- ☒ Relative risk = 60/20 = 3 (smokers are at a higher risk of developing CHD 3 times than non-smokers).
- ☒ Attributable risk = 60-20/100 = 40/100 (40 out 100 of CHD cases among smokers is attributed to their smoking).

الجدول دا مهم جدا!!!
ممكن يجي كله على بعضه SAQ

Differences between case-control & cohort studies

Case-control studies	**Cohort studies**
Proceed from outcome to cause (from disease to risk factor)	Proceed from cause to outcome (from risk factor to disease)
Compares people with disease &those without disease	Compares exposed with non exposed
Retrospective	Prospective
☒ **Aims** to prove or disprove that suspected cause occurs more frequently in diseased than non diseased	to prove or disprove that suspected disease occurs more frequently in exposed than non exposed.
☒ **Advantages** 1. Cheap &quickly done. 2. Does not require large sample. 3. Useful in studying rare diseases. 4. Can study several risk factors. 5. Can estimate risk (odds ratio)	1. Less bias in selection of control. 2. Methods can be standardized. 3. Study several outcomes. 4. Valuable in rare exposure. 5. incidence rate and relative risk can be calculated
☒ **Drawbacks** 1. Liable to bias. 2. Not useful in rare exposures. 3. Uncertain data due to incomplete records of past events& unstandardized observation. 4. Difficulty to be sure that the association is causal or not.	1. Expensive and time consuming. 2. Needs a very large sample even with common diseases. 3. Delayed results if latent period is long. 4. Prolonged follow up can cause drop out of cases and loss of standardization.

Summary and wrap-up

- ☒ Cohort study is a type of **observational analytic study** designs.
- ☒ The participants **do not have** the outcome of interest to begin with. They are selected based on the exposure status of the individual.
- ☒ If the exposure is rare, then a cohort design is an efficient method .
- ☒ Losses during Follow-up of the study participants is very important source of bias.
- ☒ These studies are used to estimate the incidence rate and relative risk.

Experimental studies

MCQ + OSPE
- Case scenario ويسأل عن :
Experimental group ? اللي اتدخلت وغيرت فيه
Control group ? اللي هقارن بيه

Types of the studies

☒ **Non-intervention studies:**

- **Descriptive studies.**
- **Comparative (analytical) studies.**

☒ **Intervention studies:**

- **Experimental studies.**

What is intervention study?

A prospective study comparing the effect and value of intervention (s) against a control in human being. It confirms etiological hypothesis & assess effectiveness of preventive measures & new therapies

مهمة جداااا

Two approaches

1. **Addition of possible causal agent (therapeutic)** e.g. testing new drug, implantation of organ. May be dangerous or fatal for human, for practical and ethical reasons.
2. **Protection from causative agent:** by removing agent from environment (smoking) or administering a protective measure (preventive) e.g. fluoridation of water supplies or vaccination. These are safe and done on human.

Characteristics of experimental study

☒ **Manipulation:** the researcher does something to one group of subjects in the study.

☒ **Control:** the researcher introduces one or more control group(s) to compare with the experimental group.

مهمة جدااا

☒ **Randomization:** the researcher takes care to randomly assign subjects to the control and experimental groups. (Each subject is given an equal chance of being assign in either group) + Matching

N.B : Matching is mainly done in case-control studies, but it can also be used in cohort & experimental studies .

Intervention studies

Investigator determines which individuals are exposed to factor of interest (intervention arm) and which are unexposed (control arm).

Stages and phases of clinical trials:

مهم نعرف كل phase والهدف منه SAQ + MCQ + OSPE

A. Stage 1 (preclinical studies = pre-phase I): involves lab. animals

B. Stage 2: involves human participants

1. Phase I trials
2. Phase II trials
3. Phase III trials
4. Phase IV trials (post-marketing surveillance)

OSPE : Case scenario (دراسة في المعمل على فأر) in which phase ? Stage 1 phase 0

Stage 1 (preclinical studies = phase 0 = pre-phase I) :

☒ **In vitro and lab. animals. We look at five things:**

- Pharmacokinetics.
- Pharmaco-dynamics.
- Drug metabolism.
- Lethal dose (LD50).
- Teratogenic effects. مهمة جدا

Stage 2: involves human participants

☒ Phase I trials (20-100 subjects):

- Not randomized, volunteers. ← على ناس سليمة
- Drugs with serious SE can be tested on seriously ill patients who have failed to respond to current established therapy.
- To assess safety, pharmacokinetics, and safe dose range.

☒ Phase II trials (100-200 subjects): ← على ناس عندها المرض

- Therapy is still promising after phase I.
- Objectives are to set & test dose necessary for pharmaco-dynamic effects, to evaluate potential effectiveness (preliminary efficacy) and to determine optimal method of administration.

☒ Phase III trials (the classical phase) (500-1500 subjects): ← أهم phase

- Randomized double blind controlled trial with adequate sample size & power.
- Aim to assess efficacy and additional safety
- Used to evaluate whether a new product should be licensed for public use.
- Provide decision makers with scientific evidence about relative effectiveness and safety of competing treatments.

MCQ : Scientific evidence about relative effectiveness and safety of competing treatments & bases for market use provided by which phase of? Phase 3 "Classical phase"

☒ Phase IV trials :

Conducted after treatment is approved for general use. Aim to assess:

- Effectiveness
- Drug safety: long term effects, surveillance for rare SE.
- Drug interactions with other drugs or diets.
- Pharmaco-epidemiology: distribution & determinant of drug use.
- Pharmaco-economics: cost-effectiveness of drug.
- Benefits & harms in presence of comorbidity.

Conducting a randomized Clinical trial

A. Formulate a hypothesis.
B. Select participants and get informed consent.
C. Allocate subjects to comparison groups.
D. Administer treatment and measure outcome.
E. Analyze data

Allocate subjects to comparison groups

☒ By randomization into: study & control groups. (means each subject in reference group has an equal chance to be present in either groups)

☒ Study group exposed to intervention.

☒ Control group: No treatment or Placebo

☒ Both groups must be matched

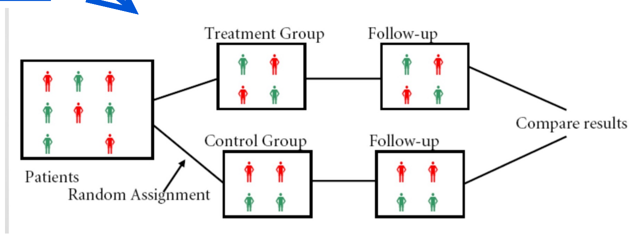
Blindness MCQ (Case scenario - type of blindness) ? - OSPE

☒ Means ensuring that a person "investigator, data collector, or analyst" remains unaware of which arm a subject has been allocated to.

- **Single-blind:** the subject participating in the trial.
- **Double-blind:** the subject & investigators (clinician, interviewers, laboratory personnel).
- **Triple blind:** the subject, investigators & the data analysts.

MCQ : The best study design in experimental study is :
Randomized double blind controlled

OSPE : Type of study ? Experimental study



Medical statistics 1

Definition

It is the study of methods of collecting, presenting (descriptive statistics), analysing and evaluating conclusions from data (inferential statistics).

Importance:

☒ It presents facts

☒ It simplifies mass of figures

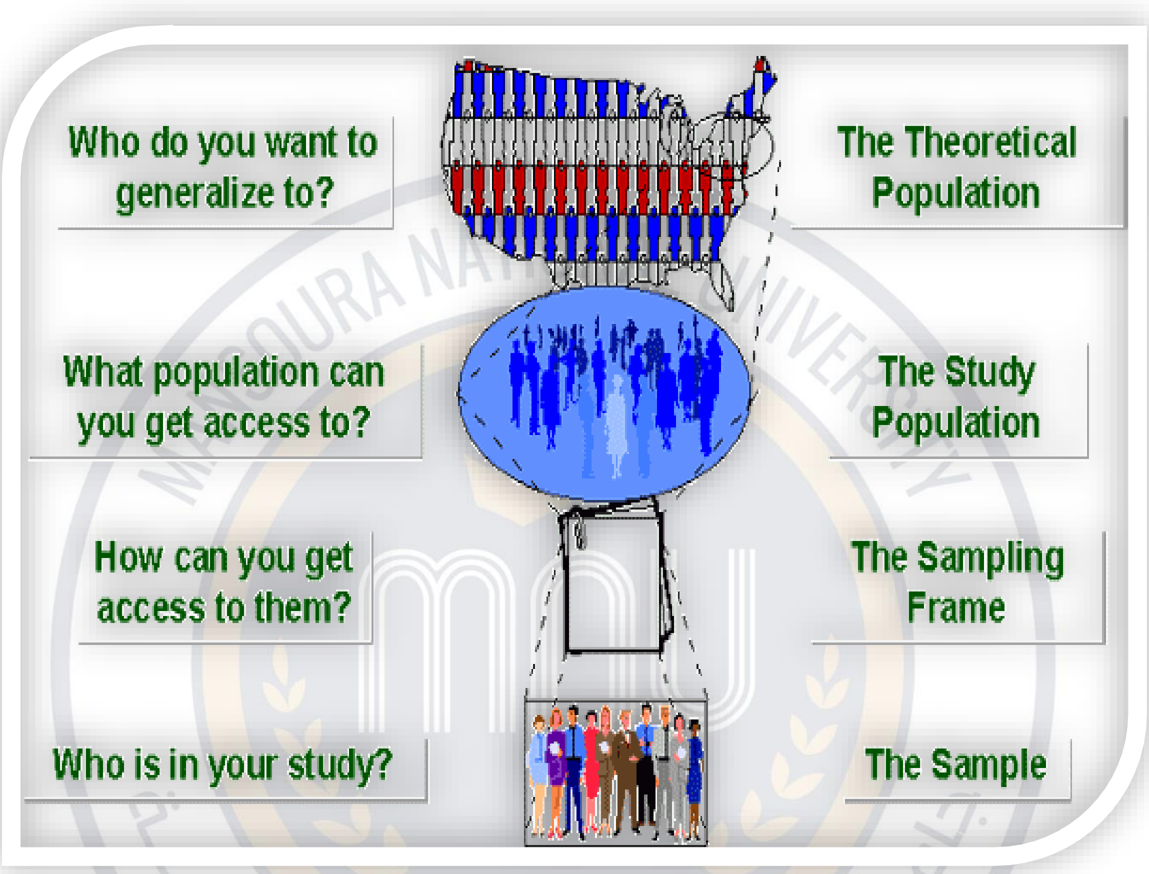
☒ It reduces the volume of data

☒ It facilitates comparison

☒ It helps in:

- formulating and testing hypothesis
- formulation of suitable policies.
- measuring the standard of health

Population & samples & sampling techniques


Who do you want to generalize to?
The Theoretical Population
What population can you get access to?
The Study Population
How can you get access to them?
The Sampling Frame
Who is in your study?
The Sample

- **Sample:**

is a subset of population that is used to gain information about the entire population. A good sample : representative-adequate-unbiased

- **Why Sampling?**

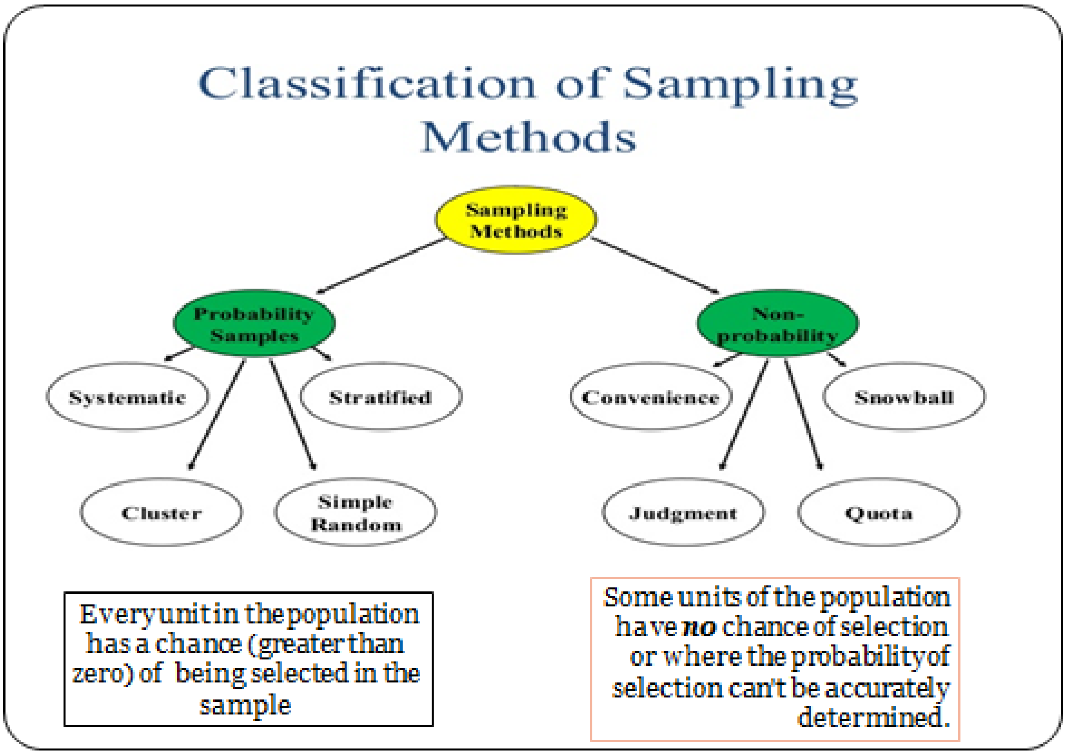
☒ Lower cost

☒ Saves time

☒ Provides more intensive and accurate investigations and information.

What happens when there is no sampling?

Selection Bias (non-representative sample): systematic difference between the characteristics of the people selected for a study and those who are not.



Classification of Sampling Methods

Sampling Methods

Probability Samples: Systematic, Stratified, Cluster, Simple Random

Non-probability: Convenience, Snowball, Judgment, Quota

Every unit in the population has a chance (greater than zero) of being selected in the sample

Some units of the population have ***no*** chance of selection or where the probability of selection can't be accurately determined.

Sampling techniques :

☒ Non-Probability

☒ Probability

OSPE : Case scenario أو صورة

Population characteristics	Appropriate sampling technique
I. Population is a homogeneous mass of individuals	**Simple Random Sample**
II. Population is heterogeneous, consists of definite strata each of which is different, characteristics	**Stratified Random Sample**
III. Sample unit is a group not an individual • They are selected randomly from all groups of same type • All members of selected group will be included in the study	**Cluster Sample**
IV. Population is a confined community Select sample units at regular intervals from this list every 3rd or 5th or 10th <<<<The start is randomly >>>>	**Systematic Random Sample**
V. Population is distributed over a large geographical area as in national surveys	**Multi-stage Random Sample**

Data & Information

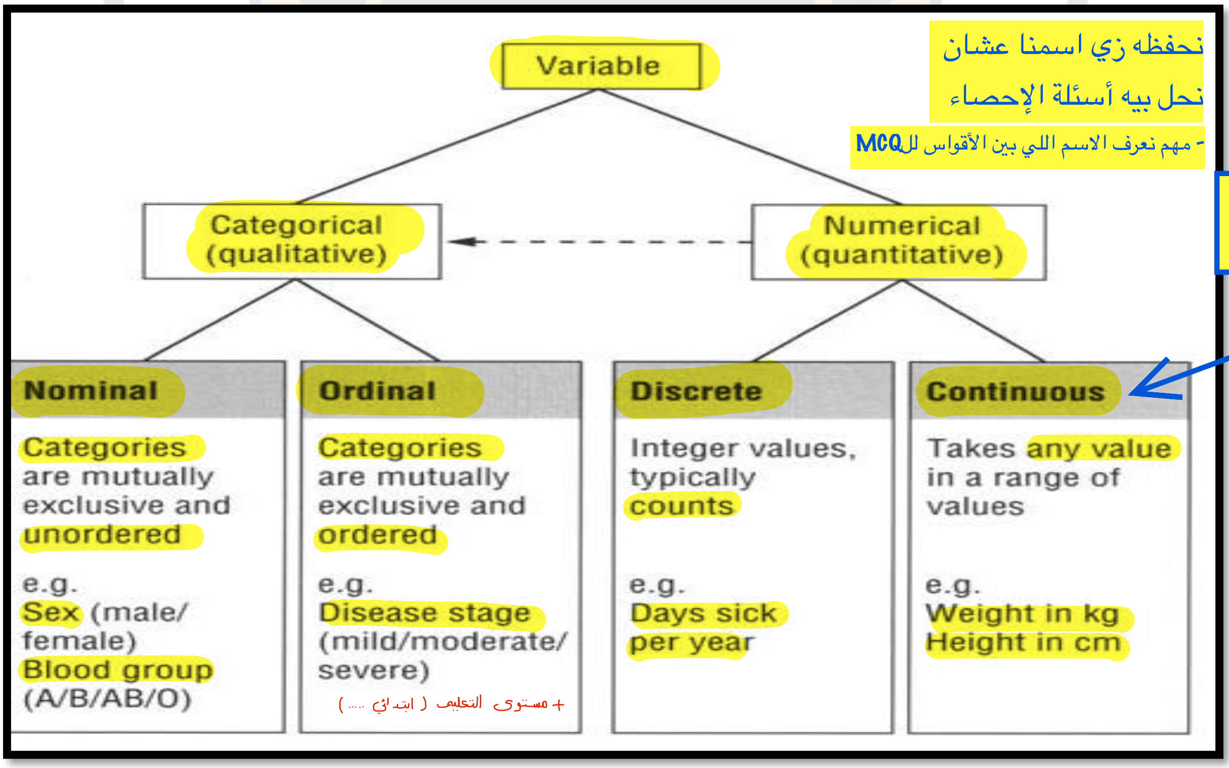
☒ Data consist of discrete observations of variables that carry no or little meaning when considered alone.

☒ Data need to be transformed (manually or by computer programs) into information by reducing them and adjusting them for variations in age and sex and others. Information support decision-makers, policy makers and planners to take proper action in their works.

Sources of Data:

- ☒ Population Census
- ☒ Registration of vital events e.g. Births and deaths, marriage
- ☒ Notification of diseases (Disease Registers) Communicable and non-communicable diseases.
- ☒ Hospital Records
- ☒ Epidemiological surveillance
- ☒ Health Service records
- ☒ Environmental Health data
- ☒ Health Surveys
- ☒ Published articles and reports

Definition of variable: Is a characteristic or attribute that vary from person to person, from time to time and from person to person



Variable

Categorical (qualitative) ← Numerical (quantitative)

Nominal
Categories are mutually exclusive and unordered
e.g.
Sex (male/ female)
Blood group (A/B/AB/O)

Ordinal
Categories are mutually exclusive and ordered
e.g.
Disease stage (mild/moderate/ severe)
+ مستوى التعليم (ابتدائي .....)

Discrete
Integer values, typically counts
e.g.
Days sick per year

Continuous
Takes any value in a range of values
e.g.
Weight in kg
Height in cm

نحفظه زي اسمنا عشان
نحل بيه أسئلة الإحصاء
- مهم نعرف الاسم اللي بين الأقواس للMCQ

ليها presentation
و analysis غير
باقي الأنواع

* Numerical → Categorical (فئات) → فئتين (nominal)
→ ٣ فئات أو أكثر(ordinal)

Analysis of quantitative data

☒ **Measures of central tendency or averages.**

- Mean
- Median
- Mode

☒ **Measures of dispersion (spread)**

- Range
- Mean deviation
- Variance
- Standard deviation

☒ **Measures of location**

- Percentile
- Quartile

مهم نعرف طريقة الحساب
والadvantages & disadvantages
OSPE : Calculate 2 central tendency?

Measures of central tendency or averages.

☒ **Mean (Average):**

- Is obtained as sum of all values divided by the no. of values Q.
- Mean = Σ x/n
- **Advantages:** used in quantitative continuous data (normally distributed)
- **Disadvantages:**
  - affected by extreme values
  - it should not be used for non parametric or skewed data
- **Importance:** Best summarizing value for normally distributed data

☒ **Median:**

- The median is the value that lies in the middle of the ordered observations.

A. When sample size is odd number:

- The observations are ordered according to an ascending or descending magnitude.
- Determine the rank of the median given by **((n + 1)/2)**
- Using the obtained rank and referring back to the ordered or arranged observations and find the value of median.

B. When sample size is even:

- In a distribution with even no. of total values: Such a distribution has 2 middlemost values; median is the average of two middlemost values when arranged in an ascending or descending order of values.
- Median = Mean (average) of (n /2)th and (n/2 + 1)th value in ascending order

- **Advantages :**
  - It can be used with quantitative & qualitative ordinal variables (e.g. median number of patients in cancer stages).
  - It is useful for summarizing data with extreme values as it is not affected by extreme values.
- **Disadvantages**
  - It cannot be used with qualitative nominal variables.
  - 2- It is not easy to be used in statistical analysis

☒ **Mode:** خلي بالكوا : لو السؤال فيه رقمين متكررين لازم نكتب الاتنين لأن الدرجة هتتوزع على كل رقم

- Most frequent or most commonly occurring value in a distribution
- This is done by finding the observation which has the highest frequency.
  - e.g. weight of five children as follows : 9, 8, 12, 7, 8 kg.
  - It is seen that eight is the observation of highest frequency.
  - The mode = 8 kg
- A similar procedure can be used for finding the mode from qualitative data.
- **Advantages:**
  - It can be used in all types of variables
  - It is not affected by extremes or out-lying observation
- **Disadvantages:**
  - Sometimes the mode cannot be determined, this happens when all observation have the same frequency (i.e. uniform distribution).
  - Sometimes we may obtain two modes (bimodal) or more (multimodal) from the same group of data.
  - e.g. 22, 24, 26, 28, 24, 26, Mode= 24 & 26

Measures of dispersion (spread)

- Using measures of central tendency is not enough to describe completely a mass of data.
  - If we have five persons with age 30, 34, 32, 36 and 28 years, the mean age is 32 years.
  - We get the same mean age of 32 years for other five persons have their ages as 12, 30, 8, 62 and 48 years but the two groups are totally different.

Measure of dispersion include :

☒ **Range:**

- It is a simple measure of dispersion and by definition range is difference between the biggest and smallest observation.
- From the above two examples range for first group = 36 - 28 = 8 years and for second group = 62-8 = 54 years.

مش هيجي لنا نحسب mean deviation, variance & SD
بس ممكن يسألك :
OSPE : Mention 2 measures of dispersion ?

☒ **Mean deviation**

$$\frac{\Sigma\, | x - \bar{x} |}{n}$$

☒ **Variance**

$$\frac{\Sigma\, (x - \bar{x})^2}{n - 1}$$

☒ **Standard deviation**

- It is the commonly used measure of dispersion and generally the best.
- It measures the deviation of observations from the arithmetic me
  - obtaining the deviation of each value from the arithmetic me
  - square the deviation from the mean.
  - The squared deviations are summed and divided by the number of observations minus one (n-1) to get the variance (S2)

- The square root of variance (S2) gives us the standard deviation(S).

$$= \sqrt{\frac{\Sigma (x-\overline{x})^2}{n-1}}$$

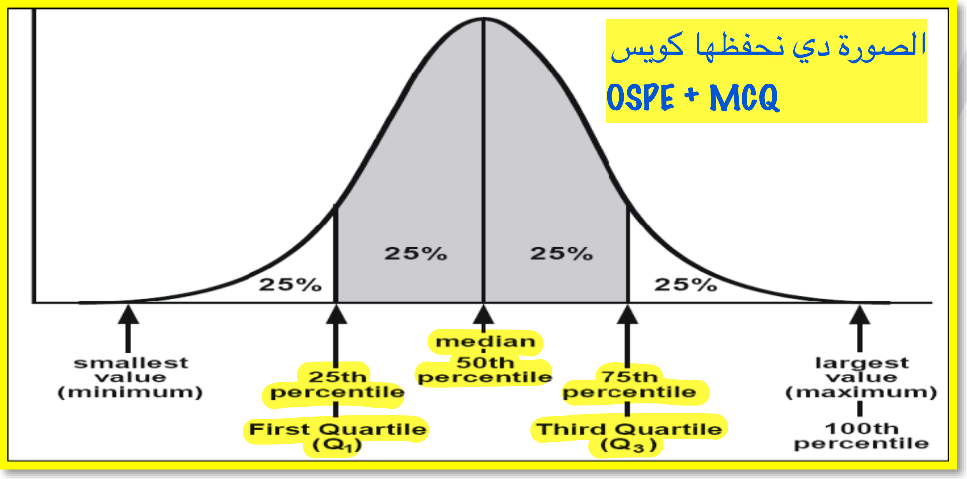
Measures of location

☒ **Quartile:**

Divides a distribution into 4 equal parts, so the number of intercepts required will be 3, i.e. Q1, Q2, Q3

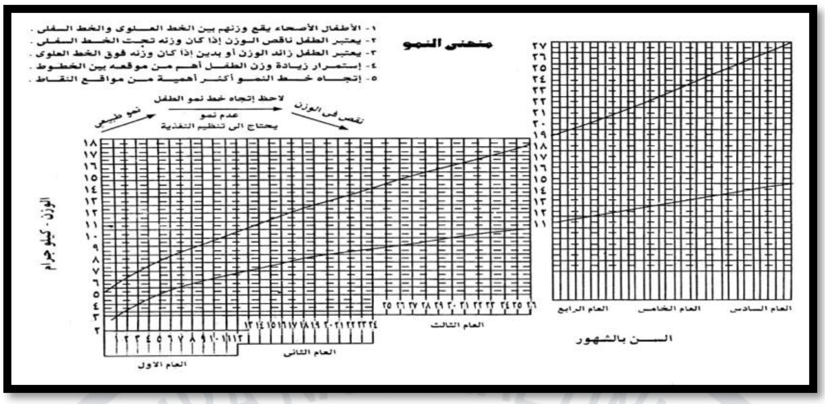
☒ **Percentile:**

Divides a distribution into 100 equal parts, AFTER arranging in an ascending order such that each part/segment has equal number (n/100) of subjects.



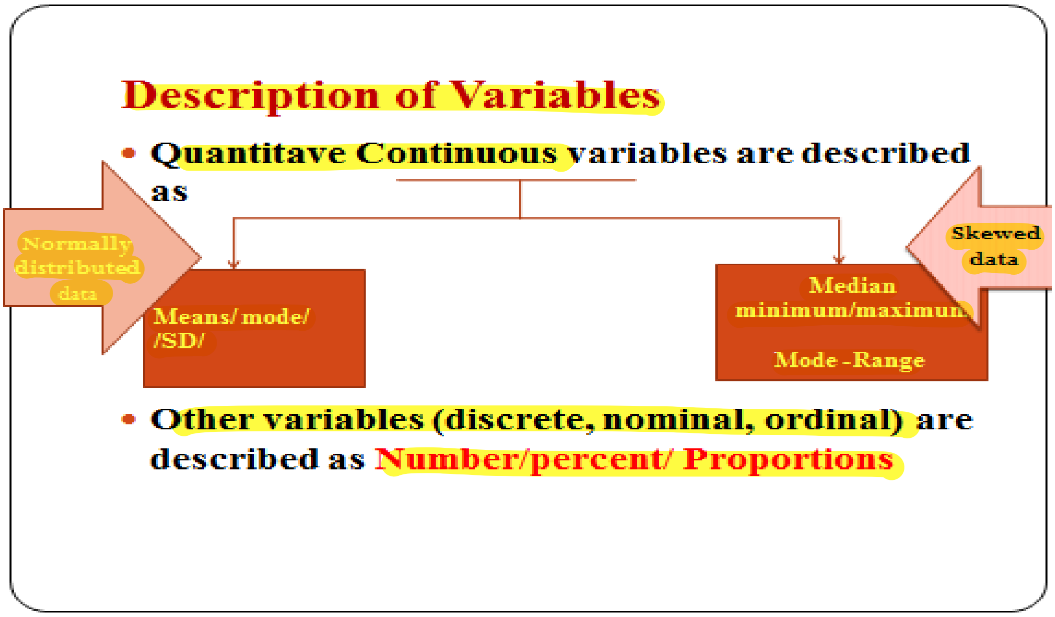
الصورة دي نحفظها كويس
OSPE + MCQ

Quartiles


منحنى النمو

١- الأطفال الأصحاء يقع وزنهم بين الخط العلوى والخط السفلى .
٢- يعتبر الطفل ناقص الوزن إذا كان وزنه تحت الخط السفلى .
٣- يعتبر الطفل زائد الوزن أو بدين إذا كان وزنه فوق الخط العلوى .
٤- إستمرار زيادة وزن الطفل أهم من موقعه بين الخطوط .
٥- إتجاه خط النمو أكثر أهمية من مواقع النقاط .

Percentiles


Description of Variables

- Quantitave Continuous variables are described as

Normally distributed data

Means/ mode/ /SD/

Skewed data

Median minimum/maximum

Mode -Range

- Other variables (discrete, nominal, ordinal) are described as Number/percent/ Proportions

Important note for MCQ - SAQ - OSPE :
* Normally distributed curve :
- Calculate central tendency (mean - mode)
- dispersion (central tendency)
* Skewed data :
- Calculate central tendency (median - mode)
- Calculate dispersion (range)

Medical statistics II:

Data distribution & presentation

Data presentation

- Data presentation can be either tabular or graphical

I . Tabular

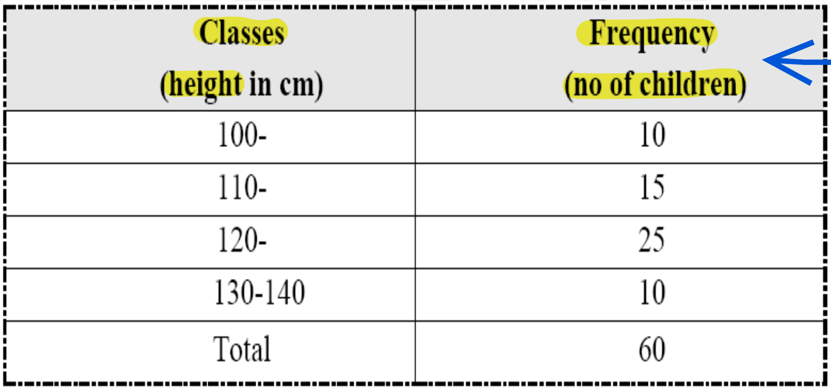
☒ Requirements of tabulation:

- Use a clear and concise title that describes the content of the data in the table.
- Precede the title with a table number.
- Label each row and each column and include the units of measurement for the data (for example, years, mm Hg, mg/dl).
- Show totals for rows and columns, where appropriate.
- Explain any codes, abbreviations, or symbols in a footnote.

☒ Types of Tabular presentation

A. Descriptive

مهم نعرف شكل الجدول **1. Descriptive Table for Quantitative data (frequency distribution table)**



Classes (height in cm)	Frequency (no of children)
100-	10
110-	15
120-	25
130-140	10
Total	60

يشيل الخانة دي ويسألنا عن تكملة الجدول
أو يشيل رقم من اللي تحت واحنا نحسبه

2. Descriptive table summarizing Qualitative data (simple frequency table)

Barriers	Frequency (no)	%
Lack of facilities	20	20
Priority to patients needs	30	30
Fear of dry hands	10	10
Forgetfullness	40	40
Total	100	100

B. Analytic

لو لقينا في الجدول مقارنة بين مجموعتين أو أكتر و p-value و test of significance هنسميه analytic table

يحدد p value ويسأل هل دي
significant ولا ؟
أو هل ال null hypothesis
rejected ولا ؟

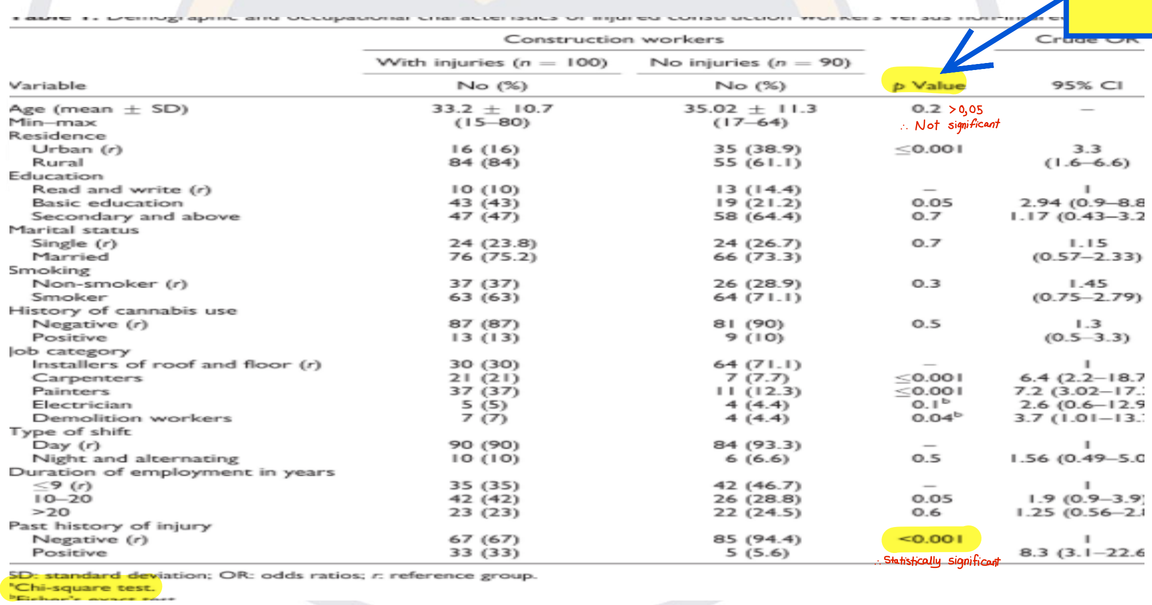


Table 1. Demographic and occupational characteristics of injured construction workers versus non-injured

	Construction workers			Crude OR
	With injuries (*n* = 100)	No injuries (*n* = 90)		
Variable	No (%)	No (%)	*p* Value	95% CI
Age (mean ± SD)	33.2 ± 10.7	35.02 ± 11.3	0.2 >0,05 ∴ Not significant	—
Min–max	(15–80)	(17–64)		
Residence				
Urban (*r*)	16 (16)	35 (38.9)	≤0.001	3.3
Rural	84 (84)	55 (61.1)		(1.6–6.6)
Education				
Read and write (*r*)	10 (10)	13 (14.4)	—	1
Basic education	43 (43)	19 (21.2)	0.05	2.94 (0.9–8.8
Secondary and above	47 (47)	58 (64.4)	0.7	1.17 (0.43–3.2
Marital status				
Single (*r*)	24 (23.8)	24 (26.7)	0.7	1.15
Married	76 (75.2)	66 (73.3)		(0.57–2.33)
Smoking				
Non-smoker (*r*)	37 (37)	26 (28.9)	0.3	1.45
Smoker	63 (63)	64 (71.1)		(0.75–2.79)
History of cannabis use				
Negative (*r*)	87 (87)	81 (90)	0.5	1.3
Positive	13 (13)	9 (10)		(0.5–3.3)
Job category				
Installers of roof and floor (*r*)	30 (30)	64 (71.1)	—	1
Carpenters	21 (21)	7 (7.7)	≤0.001	6.4 (2.2–18.7
Painters	37 (37)	11 (12.3)	≤0.001	7.2 (3.02–17.
Electrician	5 (5)	4 (4.4)	0.1[b]	2.6 (0.6–12.9
Demolition workers	7 (7)	4 (4.4)	0.04[b]	3.7 (1.01–13.
Type of shift				
Day (*r*)	90 (90)	84 (93.3)	—	1
Night and alternating	10 (10)	6 (6.6)	0.5	1.56 (0.49–5.0
Duration of employment in years				
≤9 (*r*)	35 (35)	42 (46.7)	—	1
10–20	42 (42)	26 (28.8)	0.05	1.9 (0.9–3.9)
>20	23 (23)	22 (24.5)	0.6	1.25 (0.56–2.
Past history of injury				
Negative (*r*)	67 (67)	85 (94.4)	<0.001	1
Positive	33 (33)	5 (5.6)	∴ Statistically significant	8.3 (3.1–22.6

SD: standard deviation; OR: odds ratios; *r*: reference group.
[a]Chi-square test.
[b]Fisher's exact test.

☒ Tests of Hypotheses and Significance

- An investigator conducting a study usually has a research idea in mind
  - ○ Research Question ? What are the risk factors (age – Job category –past hx of injuries –drug abuse for occupational injuries among construction workers?
- **Null hypothesis:** there is no association between risk factor and injuries

☒ Alternative hypothesis:

There is association between risk factor and injuries

☒ Test of significance:

MCQ : Case scenario (مقارنة بين أطوال طلبة في سنة خامسة وأطوال طلبة في كجي تو) test of significance: t - test

Quantitive data	**Qualitative data**
t test	**Chi square test**
To compare between different means	**To compare frequencies of categorical variable in different groups**

☒ **The result of the statistical test** either supports or rejects the Null hypothesis.

☒ Probability value (p-value)

- The p-value is the probability of obtaining the effect observed in the study (or one stronger) if the null hypothesis of no effect is actually true.
- The p value gives the probability of any observed difference having happened by chance.
- It is the cutoff for rejecting null hypothesis (p=0.05).
- Interpretation of results:

If p value less than 0.05	If p value more than 0.05
Reject null hypothesis	Accept null hypothesis
Accept alternative hypothesis	Reject Alternative hypothesis

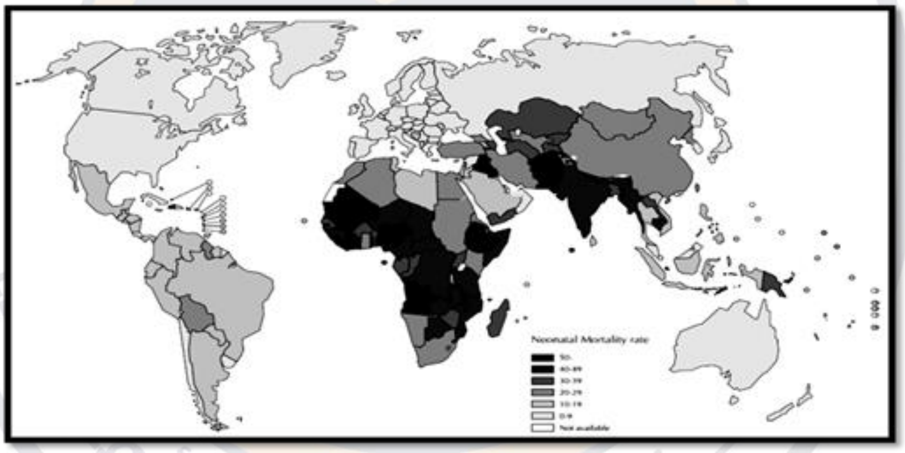
مهم نشوف أول
حاجة الp value
في أي
statistical test

SAQ : Enumerate 2 graphical presentation?

II . Graphical presentation:

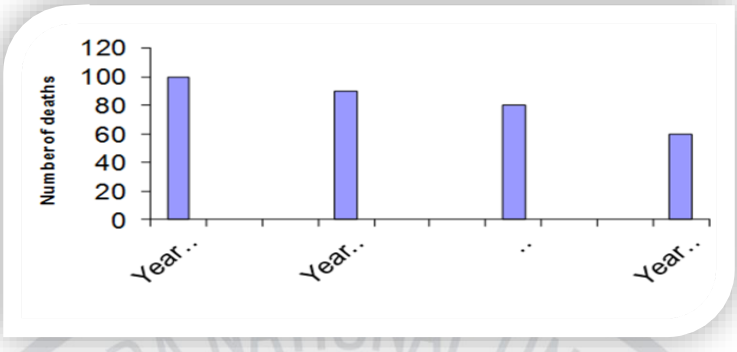
- ☒ **Visual display of data using plots and charts**
- ☒ **Not substitute for tables**
- ☒ **Stress on certain information ,quick idea about situation**
- ☒ **Should be as simple as possible, self-explained without reference to text**

1. **Map diagram/cartogram:**
   - is a map which demonstrates geographic distribution of a particular characteristic or variable
   - e.g. prevalence of certain disease or infant mortality. The following Cartogram demonstrates Neonatal mortality rate, by country, 2000.

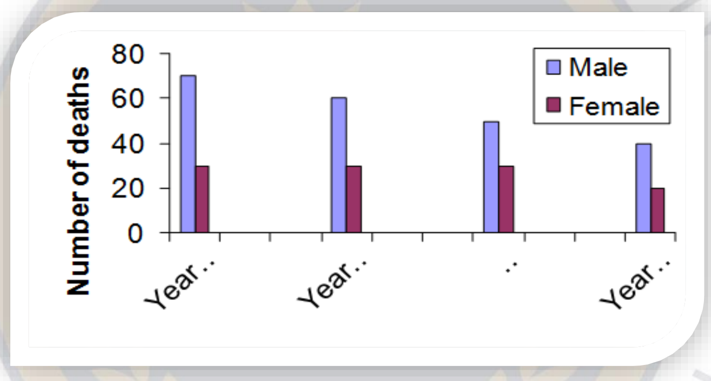


Total in bar chart
مش لازم يكون ١٠٠٪

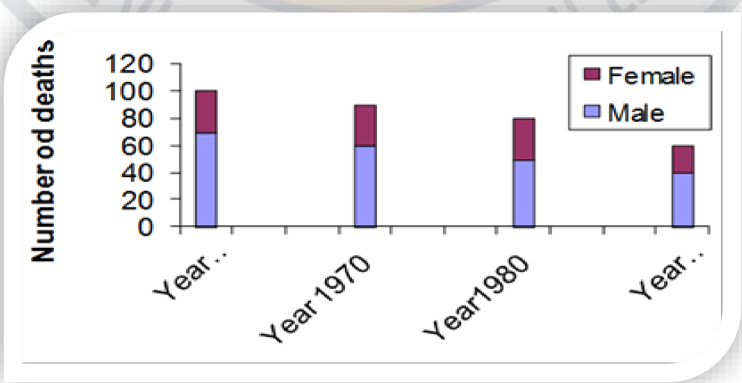
2. **Bar chart**
   - **Simple bar chart:** Different values of a particular variable are illustrated by vertical or horizontal bars to show simple comparison of size. The base line forms the time scale.
   - **Multiple bar charts:** 2 components of data represented by one chart.
   - **Component bar chart:** Each bar stands for a number of components according to their relative proportion.



Simple bar chart showing **number of deaths among patients** admitted to the hospital (X) in years (1960-1990).



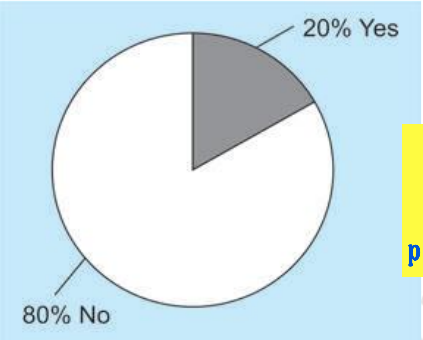
Multiple bar chart showing **males & females deaths among patients** admitted to the hospital (X) in years (1960-1990)



Component bar chart showing **males & females deaths among patients** admitted to the hospital (X) in years (1960-1990)

3. Pie or circular charts:

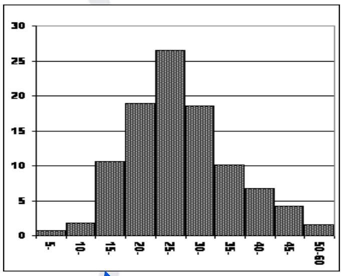
- Is for 'presentation of discrete data or qualitative characteristics'
- All pie categories are mutually exclusive, with a total of 100% (360º).



معلومة مهمة جداااا :
ممكن نحول الpie chart لbar chart
بس مش كل الbar chart ممكن أحولها لpie chart

4. Histogram:

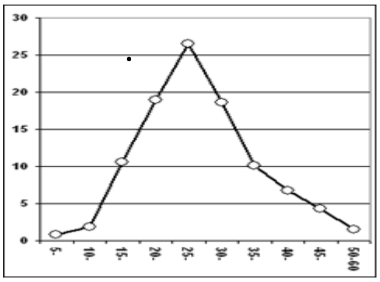
- Observations of frequency distribution table are illustrated on arithmetic paper as rectangles drawn side by side with no spaces between to get a block diagram (histogram). Classes of the frequency distribution are plotted along
- the X axis & height of constructed rectangles is corresponding to the frequency.



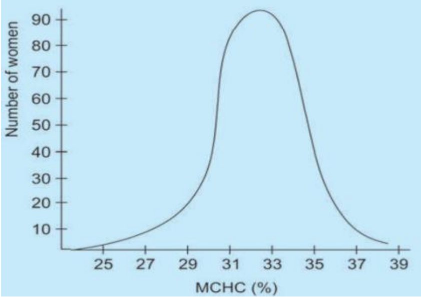
OSPE : type of graph ?
Variable ?
Type of variable ?
Other 2 graphs for same data ?

5. Frequency polygon:

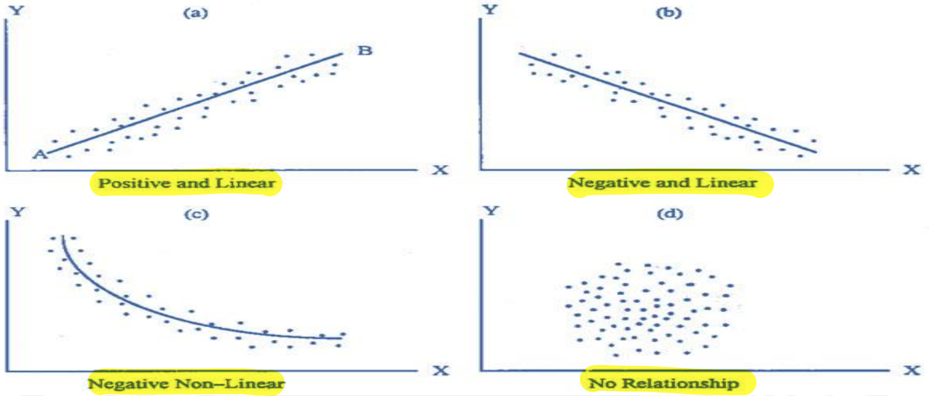
- Mid points of upper bases of rectangles are connected together by series of straight lines



6. Smooth curves: = Normal distribution curve

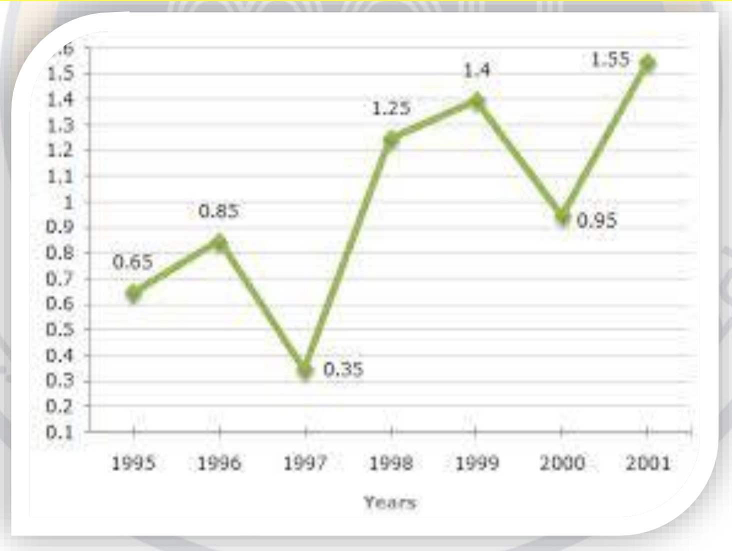


7. **Scatter diagram:** Illustrate the relationship between two continuous variables مهم



8. **Line graph:** Is a frequency polygon presenting variations by line– It shows the trend of an event over a period of time.

OSPE + SAQ + MCQ : Graph show trend over time ? Line graph



Display data over time for example:
prevalence of diseases by time

<u>**Summary**</u> OSPE + SAQ : Enumerate ? مهمة جدا!!!!

* The most common types of graphical presentation for <u>**discrete data**</u> & qualitative data are Bar & Pie charts and map diagrams.

* Histogram, frequency polygon & smooth curves for <u>**continuous data**</u> or grouped data arranged in frequency distribution.

Data distribution

Symmetric distribution curves: Normal (Bell- shaped) curve & U shape of death rate according to age

I. Normal Distribution مهمة جدااا وبيجي عليه أسئلة كتير

☒ **It is bell shaped and symmetric curve**

☒ The curve rises to its peak at the mean where mean = median = mode and it is located at the midpoint of the base

☒ The area under normal curve unity = 100%, each half = 50% MCQ

☒ The area starts from –ve to +ve and the two edges of curve do not meet X line except at infinity

☒ The X axis is divided according to standard deviation into approximately 3 standard deviations

☒ **Mean ± 1 standard deviation = 68.2% and Mean ± 2 standard deviations = 95.45% and mean ± 3 standard deviations = 99.73%**

الأرقام دي مهمة جدااا
وممكن يجي فيها مسألة

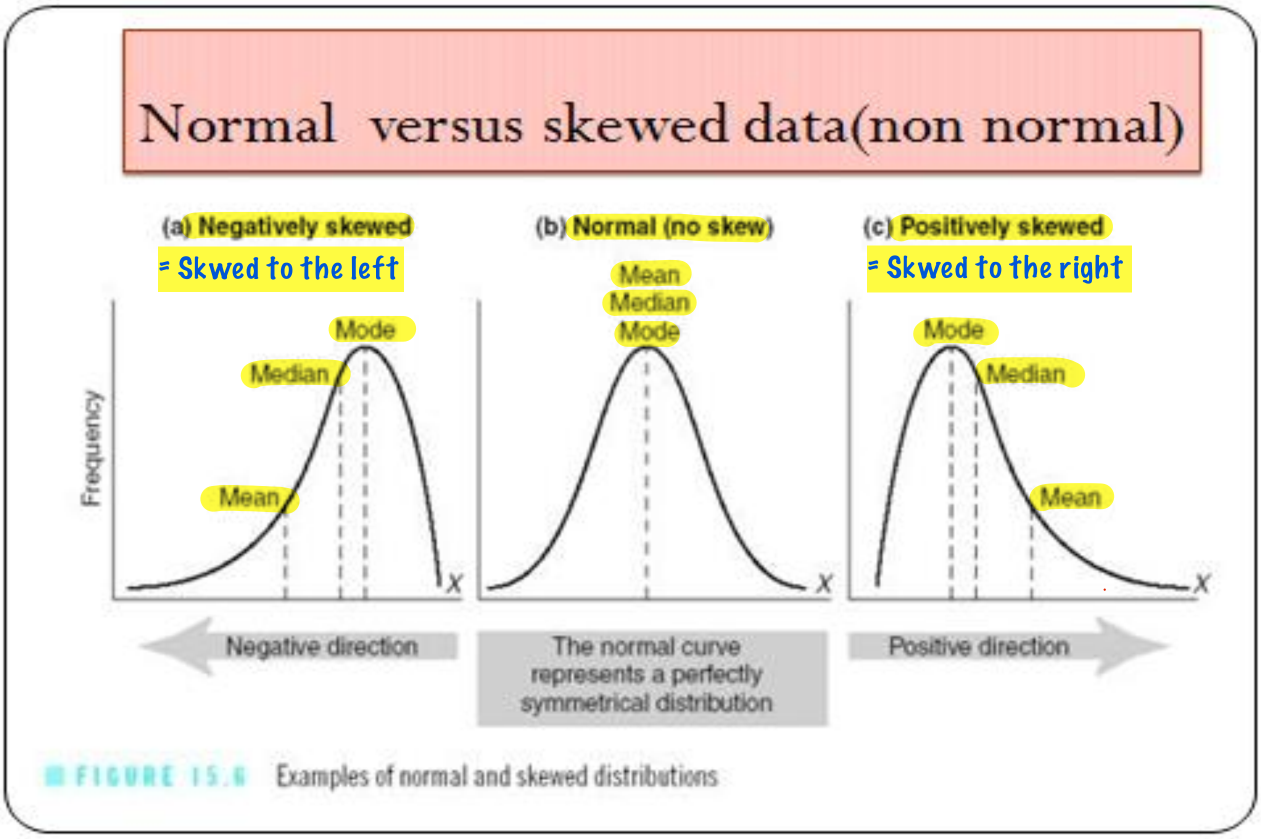
☒ **Example**

- One thousand randomly selected men have a mean systolic blood pressure of 120 mmHg with a SD of 10 mmHg. The population is normally distributed with respect to systolic blood pressure.
- About 68% (680) of the men have systolic blood pressure of 120 ± 10 (mean ± 1 SD) or systolic blood pressure ranging from 110-130 mmHg.
- 95% (950) of men have systolic blood pressure of 120 ± 2 (10) (mean ± 2 SD) or systolic blood pressure ranging from 100 – 140 mmHg and 99.7% their systolic blood pressure is 120 ± 3 (10) (mean ± 3 SD) or men systolic blood pressure ranging from 90 -150mmHg.

II. Skewed (asymmetric) distribution curves: مهمة

☒ **Skewed to the right**: mean>median>mode

☒ **Skewed to the left**: mean<median<mode



Normal versus skewed data(non normal)

FIGURE 15.6 Examples of normal and skewed distributions

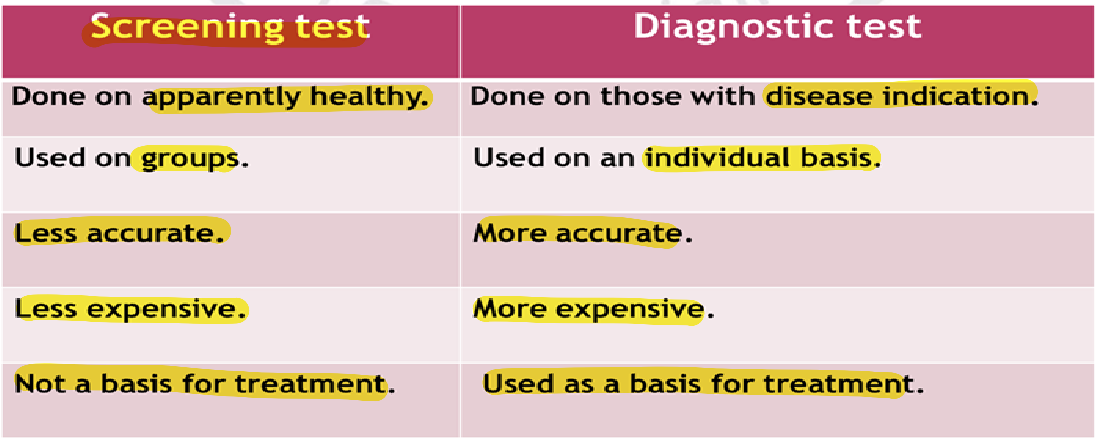
MCQ مهم نعرف إن ال**mean** أكتر قيمة بتتأثر في حالة ال**skewed**

Role of screening tests in disease diagnosis

- **Definition:**
  - ☒ Screening is the investigation of apparently healthy individuals to detect unrecognized cases or individuals with high risk of developing a disease.
  - ☒ Therefore, intervention can be done to prevent occurrence of the disease or to improve its prognosis when it develops.
- **Objective of a screening test:**
  - ☒ **Immediate objective:** Simple test applied on large number to exclude those free from the disease and pick up those possibly suffering of the disease and subjected to detailed investigation to prove or disprove the diagnosis
  - ☒ **Ultimate objective:** to reduce mortality and morbidity

- **Screening test:** a simple test applied on large number to exclude those free from disease & to pick up those possibly suffering from disease & subjected to detailed investigation to prove or disprove the diagnosis (i.e. reference test).

- **Difference between screening test & diagnostic test:** SAQ مهم



Screening test	Diagnostic test
Done on apparently healthy.	Done on those with disease indication.
Used on groups.	Used on an individual basis.
Less accurate.	More accurate.
Less expensive.	More expensive.
Not a basis for treatment.	Used as a basis for treatment.

- **Nature of screening tests: Screening tests may be:**
  - ☒ A clinical step (e.g., breast palpation),
  - ☒ A laboratory (e.g., glucose tolerance test for diabetes mellitus)
  - ☒ Other investigation (e.g., mammography).
- **Types of Screening:** SAQ
  - ☒ **Mass Screening** offered to all individuals, irrespective of the presence of particular risk to the disease in question. This is not a useful preventive measure unless it is backed-up by treatment & follow-up facilities for positive screening.

    زي اللي بتتعمل مع الhypothyroidism
    من كعب الرجل لأي مولود

  - ☒ **High Risk Screening** offered to those with special risk, e.g., screening of close relative of known diabetics (a greater number of cases can be identified at less cost).
  - ☒ **Multiphase screening** for a variety of diseases at one time. This is a well-established procedure in antenatal care & school examinations.
- **Requirements for a screening program:**
  - ☒ Suitable disease
  - ☒ Suitable test
  - ☒ The population to be screened
- **Requirements Of Screening Program regarding (The disease):** سؤال مهم جدا!!!!

1. **Importance of the disease:**

   The disease should be an important health problem, i.e., high frequency and/or bad sequelae, e.g., congenital hypothyroidism, although rare, should be detected early because of its serious sequelae if untreated and because it is treatable.

2. **Adequate understanding of the natural history of the disease:**

   to identify the points at which the disease can be detected by screening with effective intervention before irreversible damage, to evaluate the effectiveness of any intervention

3. A recognized latent period or asymptomatic stage.
4. Can be detected before onset of symptoms and signs
5. At risk individuals can be identified and screened
6. Available facilities for diagnosis and treatment.
7. Agreed policy on whom to treat as patients
8. An effective treatment, available , effective and acceptable
9. Benefits of early detection exceeds risks and costs (money, manpower and equipment).

Requirements of a screening test:

أهم اتنين

☒ Valid

☒ Reliable

☒ Cheap, easily and quickly done.

☒ Safe, not painful.

☒ Objective rather than subjective.

☒ Acceptable by the population

Validity and reliability of screening test:

☒ It is the capacity of a test to give **true results**. Therefore, a valid test is the test which **correctly** detects the presence or absence of a condition.

☒ e.g., glucosuria as a test to detect diabetes mellitus has poor validity compared to glucose tolerance test.

☒ **Validity includes:** مهم نحفظ الdefinitions

- **Sensitivity:** The ability of the test to identify correctly those who have the disease, i.e., it gives few false negative results.
- **Specificity:** The ability of the test to identify correctly those who do not have the disease, i.e., it gives few false positive results.
- **Positive predictive value:** abbreviated PPV or PV+, is the proportion of all people with positive tests who truly have the condition – a / (a+b) in the above table.

- **Negative predictive value:** (NPV or NP-) is the proportion of all people with negative tests who truly do not have the condition – d / (c+d) in the above table.

Results of screening test & the true diagnosis:

موضع سؤال
- يجيب الجدول جاهز واحنا نحسب
- يجيب **case scenario** واحنا نكون الجدول

نوتس على السريع : مهم نعرف نحدد ال**diagnostic**
test في السؤال وهنلاقي قبله كلمات مفتاحية زي :
gold standard test - PCR - Culture media - confirmatory test - HbA1C test

Screening test	Disease		Total
	Present	Absent	
Positive	(TP) A	(FP)B	A+b
Negative	C(FN)	D(TN)	C+d
Total	A+C	B+D	A+B+C+D

☒ **Sensitivity:** = probability of a positive test in people with the disease = a/ (a+c)

☒ **Specificity:** = probability of a negative test in people without the disease = d/ (b+d)

☒ **Positive predictive value:** probability of the person having the disease when the test is positive = a /(a+b)

☒ **Negative predictive value:** probability of the person not having the disease when the test is negative = d/ (c+d).

☒ **Accuracy: =** (a+d) / (a+b+c+d)

☒ **Reliability (Repeatability):**

- It is the level of agreement between repeated measurements; therefore, a technique will give the same values on repeated application on the same individual.

- Disease with high mortality consider that the test is highly specific
مهمة جدااا دكتورة غادة أكدت عليها في المحاضرة واتأكد عليها في المراجعة

Morbidity and mortality Statistics

Definition of Morbidity Statistics: SAQ

- Statistics that enumerate the extent, frequency, or severity of disease in a community.

Types of Morbidity Statistics: SAQ + OSPE : Enumerate examples of morbidity statistics ?

مهم نعرف :
- التعريف (نظري)
المعادلة بتاع كل rate وحسابها (نظري وعملي)

☒ **Incidence rate**

- **Definition:**

The rate of occurrence of new cases in a specified population.

- **Calculation:**

$$\frac{\textbf{Number of reported new cases of a disease in certain Y / } \boldsymbol{L}}{\textbf{At risk population in the same Y/L}} \times 10^{n}$$

☒ **prevalence rate**

- **Definition:**

Frequency of existing cases (old and new) in a defined population.

- **Calculation:**

$$\frac{\textbf{Number of people with a disease or condition (old + new cases)}}{\textbf{Total number of examined population at same locality and time}} \times 10^{n}$$

Constant is 10 ^{n}, where n = 1 or 2, or 3 etc.

OSPE + MCQ :
- Nominator of .....?
- Denominator of ....?

SAQ + OSPE : Enumerate examples of mortality statistics ?

مهم نعرف :
- التعريف (نظري)
المعادلة بتاع كل **rate** وحسابها (نظري وعملي)
ومهم جداااا نكتب المعادلة قبل ما نعوض لأن عليها جزء من الدرجات

Types of Mortality Statistics:

A. Crude (death) mortality rate

$$\frac{\text{Total number of deaths a certain Y \& L}}{\text{Estimated mid year } \mathit{population\ at\ same} \text{ Y \& L}} \times 1000$$

B. Sex Specific Death Rate

$$\frac{\text{Total number of deaths of a certain sex in a certain Y \& L}}{\mathit{Total\ number\ of\ same\ sex\ at\ same} \text{ Y \& L}} \times 1000$$

C. Cause-specific mortality rate

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\mathit{Estimated\ mid-year\ population\ at\ same} \text{ Y \& L}} \times 100000$$

D. Proportionate mortality rates

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\mathit{Total\ deaths\ from\ all\ causes\ at\ same} \text{ Y \& L}} \times 100$$

E. Case fatality rate

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\mathit{Total\ number\ of\ cases\ at\ same} \text{ Y \& L}} \times 100$$

F. Maternal Mortality

- **Definition:**

  Maternal mortality means death among mothers due to causes related to and/or aggravated by pregnancy, labor & puerperium.

- **Rate:** مهم كل حرف في المعادلة نظري وعملي

$$\frac{\text{Total number of maternal deaths in a certain Y \& L}}{\mathit{No.\ of\ female\ in\ childbearing\ period}\ (15-49\ y)\ \mathit{at\ same} \text{ Y \& L}} \times 100000$$

N.B : الرقم اللي بنضرب فيه دا اللي بنعبر بيه لما يطلع الرقم من المعادلة
لو ما فهمتش قصدي افتح سؤال **3** في ال**6 training** صفحة **28**

- **Causes** SAQ

1. Hemorrhage: May occur during pregnancy, labor or puerperium. It forms the most important causes in Egypt.
2. Hypertensive disease of pregnancy (eclampsia & preeclampsia).
3. Puerperal sepsis: most preventable cause.
4. Pre-existing diseases aggravated by pregnancy, labor & puerperium e.g.
   - Rheumatic heart disease.
   - Chronic glomerulonephritis complicated by renal failure.
   - Uncontrolled D.M

- **Age specific mortality rates:**

مهم نعرف :
- التعريف (نظري)
المعادلة بتاع كل rate وحسابها (نظري وعملي)
ومهم جدا!!!!! نكتب المعادلة قبل ما نعوض لأن عليها جزء من الدرجات

A. Still birth

- **Definition:**
  - Stillbirth is the delivery, after the 28th week of pregnancy, of a baby who has died.
  - Loss of a baby before the 28th week of pregnancy is called a miscarriage.
- **Rate:**

$$\frac{\text{Total number of still births in a certain Y \& L}}{No.\ of\ total\ births\ at\ same\ \text{Y \& L}} \times 1000$$

B. Infant Mortality Rate

$$\frac{Number\ of\ deaths\ less\ than\ 1\ year\ of\ age\ in\ \ a\ certain\ Y\ \&\ L}{No.\ of\ \text{live}\ births\ at\ same\ Y\ \&\ L} \times 1000$$

Infant mortality rate is divided into neonatal and post neonatal mortality rates.

C. Neonatal mortality rate

$$\frac{\text{Number of deaths less than 28 days of age in a certain Y \& L}}{No.\ of\ \text{live}\ births\ at\ same\ \text{Y \& L}} \times 1000$$

Causes: SAQ

مترتبين بحسب الimportance والmost common

- Prematurity (preterm & LBW)
- Congenital malformations & Rh incompatibility
- Birth injuries
- Asphyxia neonatorum.
- Infections as congenital infection, tetanus neonatorum, acute respiratory disease &diarrhea.

D. Post-neonatal mortality rate

$$\frac{\text{Number of deaths from 28 days to 1 year in a certain Y \& L}}{No.\ of\ \text{live}\ births\ at\ same\ \text{Y \& L}} \times 1000$$

Causes: SAQ

مترتبين بحسب الimportance والmost common

- Infections in the form of:

  *Acute respiratory diseases

  * Infective diarrheal diseases.

- Other infections e.g. tetanus neonatorum, pertussis & measles.
- Sever nutritional deficiency e.g. PEM.
- Accidents.
- The remaining section of prematurity & congenital malformations

E. Child (1-4years) mortality rate

$$\frac{\text{Number of child } (1-4 \text{ years of age}) \text{ in a certain Y \& L}}{No.\ of\ total\ children\ in\ same\ age\ group\ at\ same\ \text{Y \& L}} \times 1000$$

Causes: SAQ

مترتبين بحسب الـimportance والـmost common

- Infectious diseases as acute respiratory diseases, diarrhea, pertussis, measles & meningitis.
- Accidents.
- Severe nutritional deficiency.
- Other causes as rheumatic heart disease

<u>Best wishes</u>