Model Aswer

1. **Give an account on types of variables and data presentation**

Types of variables (data):

I- Quantitative (Numerical):

Measurements are expressed in numbers.

A- Discrete variables: variable expressed as a whole number with no fraction.

e.g. - Number of children in family.
- Number of pregnancies.
- Pulse rate.

B- Continuous: there is continuous change in its value, fraction may be present.

e.g. Height, Weight, and Age.

II- Qualitative (categorical):

Measures expressed as description.

A- Nominal: no special arrangement.

e.g. : Sex (Male and Female), Presence of Hypertension (Yes or No) (Dichotomous).
- Blood group (ABO) and race (White, Black, and Hispanic) (Multichotomous).

B- Ordinal : data can be arranged.

e.g. grade of disease (mild, moderate, and severe).

Presentation of Data (tables –graphs- parameters)

Presentation of quantitative (Numerical) variables:

- **Table**

Frequency distribution and relative frequency table;

There is no difficulty if the data set is small, we can arrange those few numbers and write them. For large data sets, a useful device for summarization is the formation of a frequency table or frequency distribution. This is a table showing the number of observations, called frequency, within certain ranges of values of the variable under investigation. For example, taking the variable to be the age at death, the second column of the table provides the frequencies.

An optional but recommended step in the formulation of a frequency distribution is to present relative frequency in addition to frequency for each interval. Relative frequency, defined by:

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total number of observations}}$$

The following table shows this.

- **Graph:** Histogram and the Frequency Polygon;

A convenient way of displaying a frequency table is by means of a histogram and/or a frequency polygon. A histogram presents us with a graphic picture of the distribution of measurements, look at the following figure

Histogram

Frequency Polygon (look at the following figure) as an alternative of filling the whole figure with columns, a point is taken at the middle and top of each column. A line is drown to link each two point together and multiple lines are seen, so the name polygon (many line).

Frequency polygon

If the smooth line passes between the points instead polygon, it is known as frequency curve.

Frequency curve

- B**Parameters:**

In quantitative variables we use two concomitant measures to summarize the data, measures of central tendency (the middle) and measures of dispersion (variability).

I- Measures of Central Tendency:

One of the most useful summary numbers is an indicator of the center of a distribution of observations the middle or average value. The three measures of central tendency used in medicine and epidemiology are the mean, the median, and, to a lesser extent, the mode.

a. Mid Range:

The value that lie mid way between the highest (maximum) and lowest (minimum) values.

Mid range = Maximum + Minimum / 2.

b. The Mean:

The mean is the arithmetic average of the observations. It is symbolized by $\bar{X}$ (called X-bar) and is calculated as follows: add the observations to obtain the sum and then divide by the number of observations. The formula for the mean is,

$$\overline{X} = \frac{\Sigma X}{n}$$

Where Σ (Greek letter sigma) means to add, X represents the individual observations, and n is the number of observations.

c. The Median:

The median is the middle value in a set of data arranged in order of magnitude. It divides the data into 2 equal groups above and below the

median value. It is defined as the mean of the two middle values for an even number of observations.

To calculate Median:

- Arrange the observations from smallest to largest (or vice versa).
- Count in to find the middle value.
- The median is the middle value for an odd number of observations.

d. The Mode:

The mode is the value that occurs most frequently. It is commonly used for a large number of observations when the researcher wants to designate the value that occurs most often. When a set of data has two modes, it is called bimodal. The modal value is the highest bar in a histogram.

II- Measures of Dispersion:

Although the mean provides useful information, you have a better idea of the distribution of data if you know something about the spread, or the variation, of the observations. Several statistics are used to describe the dispersion of data: range, standard deviation, percentile rank, and inter-quartile range coefficient of variation, and minimum and maximum.

a. Minimum and Maximum:

It's the highest and lowest value of the data.

b. The Range:

The range is the difference between the largest and the smallest observation. It is easy to determine once the data have been arranged in rank order.

c. The Standard Deviation:

The standard deviation is a measure of the "average" spread of the observations about the mean. The name of the statistic before the square root is taken is the **variance**, but the standard deviation is the statistic of primary interest. The standard deviation is symbolized as SD, or simply s and its formula is:

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

II.Mention briefly

a. **Validity parameters of screening test**

Validity: is the rate at which a test is capable of differentiating the presence or absence of a disease concerned.

Validity is the rate at which a test is capable of differentiating the presence or absence of a disease concerned

Sensitivity = ability of test to detect people who actually have the disease
(True Positives/All Positives)

Positive predictive value is correct identification of cases among test positive individuals
(True Positives/All test Positives)

Specificity = ability of test to identify correctly people who actually do not have the disease
(True Negatives/All not diseased)

Negative predictive value is correct exclusion of not diseased among test negative individuals
(True Negatives/ all test negative individuals)

Screening Test Versus Diagnostic Test Results

Screening Test		Diagnostic Test +ve	Diagnostic Test -ve	Total
	+ve	A True +ve	B False +ve	A+B
	-ve	C False –ve	D True -ve	C+D
Total		A+C	B+D	A+B+C+D

Validity Parameters of Screening Test:

- Sensitivity % = A/(A+C) * 100
- Specificity % = D/ (B+D) *100
- +ve Predictive value % = A/(A+B) *100
- -ve predictive value % =D/(C+D) *100
- Agreement % = ∑ A+D/ (A+B+C+D) *100

--Case-control study

1. Case/Control study:

It is an "observational" design comparing exposures in disease cases versus healthy controls from same population.

- Exposure data collected retrospectively.
- It is the most feasible design where disease outcomes are rare.

Advantages of Case-Control Study:

- Quick, inexpensive.
- Well-suited to the evaluation of diseases with long latency period.
- Useful in rare diseases.
- Examine multiple etiologic factors for a single disease.

Disadvantages of Case-Control Study:

- Not useful in rare exposure.

- Incidence rates cannot be estimated.
- Selection Bias and recall bias.

Controls should be similar (matched) to cases in different aspect except freedom of the investigated condition. Their source may be hospital relative of the case or community members. Controls are subjected for interviewing & investigation like cases. The data obtained from both groups are compared.

c- Probability sample techniques

1. **Random (probability) Sample :**

Random in statistics means:

- All units of population are known and available for sampling.
- All units have an equal chance to be taken in the sample (probability).
- Unit: is the element of interest (person, house, and place).

Types of Random Sample:

1. **Simple Random Sample:**

It is the simplest type of random technique that satisfy the two condition mentioned before. This could be done by:

- Lottery Methods.
- Table of Random Number.
- In large samples we give number and use computer to draw random numbers.

2. **Cluster Random Sample:**

Cluster Sample when the population falls into naturally occurring subgroups, each having similar characteristics; a cluster sample may be the most appropriate. To select a cluster sample, divide the population into groups, called clusters, and select all of the members in one or more (but not all) of the clusters.

3. **Stratified Random Sample:**

- When it is important for the sample to have members from each segment of the population, you should use a stratified sample.
- Depending on the focus of the study, members of the population are divided into two or more different subsets, called strata, according to certain character (sex classify population into males and females).
- A sample is then randomly selected from each of the strata.
- Using a stratified sample ensures that each segment of the population is represented.

For example, to collect a stratified sample of the number of people who live in Fayoum governorate households, you could divide the households into socioeconomic levels, and then randomly select households from each level.

4. Systematic Random Sample:

- A sample in which each member of the population is assigned a number.
- The members of the population are ordered in some way.
- A starting number is randomly selected, and then sample members are selected at regular intervals from the starting number. For instance, every 3^{rd}, 5^{th}, or 100^{th} member is selected.

Steps of Systematic Random Sample:

- Determine sample fraction by dividing the number of population on sample size
- For example, 100/20= 5.
- By simple random technique 2^{nd} unit is selected.
- By adding 5, the sample units will be 7^{th}, 12^{th}, and so on till complete the sample size, here 20.

D. Data collection tools

Data Collection Tools:

1. Questionnaire.

2. Observation checklist.
3. Data collection forms.
4. Other data collection tools.
    - Photography / Video: provides visually represented information
    - Maps and drawing.

E. Relative risk and its role in determing disease causation

The Relative Risk:

Relative Risk (RR) of exposure to studied factor = (1) / (2) or a / (a + b) divided by c/ (c+ d). It measure incidence of disease among exposed in relation to non-exposed.

Relative risk (RR) answers the question:"How many times an exposed person is at risk of developing disease compared to non-exposed?"

Interpretation of (RR) Relative risk:

- >1 means the exposure is a risk factor.
- = 1 means the exposure is not associated with the disease.
- < 1 means the exposure is protective.