Introduction

The term **statistics** is used to mean either statistical data or statistical methods. **Statistical data** refers to numerical descriptions of things. These descriptions may take the form of counts or measurements. Thus statistics of malaria cases in one of malaria detection and treatment posts include fever cases, number of positives obtained, sex and age distribution of positive cases, etc.

Statistical methods refers to a body of methods that are used for collecting, organizing, analyzing and interpreting numerical data for understanding a phenomenon or making wise decisions.

The branch of modern statistics that is most relevant to public health and clinical medicine is **statistical inference.** This branch of statistics deals with techniques of making conclusions about the population. Inferential statistics builds upon **descriptive statistics**. The inferences are drawn from particular properties of sample to particular properties of population. These are the types of statistics most commonly found in research publications.

When the different statistical methods are applied in biological, medical and public health data they constitute the discipline of **Biostatistics**.

Rationale of studying statistics

- More and more things are now measured quantitatively in medicine and public health.
- There is a great deal of intrinsic (inherent) variation in most biological processes.
- Public health and medicine are becoming increasingly quantitative. As technology progresses, the physician encounters more and more quantitative rather than descriptive information.
- The planning, conduct, and interpretation of much of medical research are becoming increasingly reliant on statistical technology. Is this new drug or

procedure better than the one commonly in use? How much better? What, if any, are the risks of side effects associated with its use? In testing a new drug how many patients must be treated, and in what manner, in order to demonstrate its worth? What is the normal variation in some clinical measurement? How reliable and valid is the measurement? What is the magnitude and effect of laboratory. and technical error? How does one interpret abnormal values?

- Statistics pervades the medical literature. As a consequence of the increasingly quantitative nature of public health and medicine and its reliance on statistical methodology, the medical literature is replete with reports in which statistical techniques are used extensively.

"It is the interpretation of data in the presence of such variability that lays at the heart of statistics."

Limitations of statistics:

It deals with only those subjects of inquiry that are capable of being quantitatively measured and numerically expressed.

1. It deals on aggregates of facts and no importance is attached to individual items–suited only if their group characteristics are desired to be studied.

2. Statistical data are only approximately and not mathematically correct.

Scales of measurement

Any aspect of an individual that is measured and take any value for different individuals or cases, like blood pressure, or records, like age, sex is called a ***variable.***

It is helpful to divide variables into different types, as different statistical methods are applicable to each. The main division is into **qualitative** (or categorical) or **quantitative** (or numerical variables).

Qualitative variable: a variable or characteristic which cannot be measured in quantitative form but can only be identified by name or categories, for instance place of birth, ethnic group, type of drug, stages of breast cancer (I, II, III, or IV), degree of pain (minimal, moderate, severe or unbearable).

Quantitative variable: is one that can be measured and expressed numerically and they can be of two types (**discrete** or **continuous**). The values of a discrete variable are usually whole numbers, such as the number of episodes of diarrhoea in the first five years of life. A continuous variable is a measurement on a continuous scale. Examples include weight, height, blood pressure, age, etc.

Although the types of variables could be broadly divided into categorical (qualitative) and quantitative , it has been a common practice to see four basic types of data (scales of measurement).

1-Nominal data: Data that represent categories or names. There is no implied order to the categories of nominal data. In these types of data, individuals are simply placed in the proper category or group, and the number in each category is counted. Each item must fit into exactly one category.

The simplest data consist of unordered, dichotomous, or "either - or" types of observations, i.e., either the patient lives or the patient dies, either he has some particular attribute or he does not.

e.g. Nominal scale data: survival status of propranolol - treated and control patients with myocardial infarction.

Status 28 days after hospital admission	Propranolol treatment Patient	Control Patients
Dead	7	17
Alive	38	26
Total	45	46
Survival rate	84%	63%

The above table presents data from a clinical trial of the drug propranolol in the treatment of myocardial infarction. There were two group of myocardial infarction (MI). There were two group of patients with MI. One group received propranolol; the other did not and was the control. For each patient the response was dichotomous; either he survived the first 28 days after hospital admission or he succumbed (died) sometime within this time period.

With nominal scale data the obvious and intuitive descriptive summary measure is the proportion or percentage of subjects who exhibit the attribute. Thus, we can see from the above table that 84 percent of the patients treated with propranolol survived, in contrast with only 63% of the control group.

Some other examples of nominal data:

Eye color - brown, black, etc.

Religion - Christianity, Islam, Hinduism, etc.

Sex - male, female

2-Ordinal Data: have order among the response classifications (categories). The spaces or intervals between the categories are not necessarily equal.

Example: 1. strongly agree 2. Agree 3. no opinion 4. Disagree 5. strongly disagree. In the above situation, we only know that the data are ordered.

3-Metric Data

a- Interval Data: In interval data the intervals between values are the same. For example, in the Fahrenheit temperature scale, the difference between 70 degrees and 71 degrees is the same as the difference between 32 and 33 degrees. But the scale is not a **Ratio Scale**. 40 degrees Fahrenheit is not twice as much as 20 degrees Fahrenheit.

b- Ratio Data: The data values do have meaningful ratios, for example, age is a ratio data, someone who is 40 is twice as old as one who is 20.

Both interval and ratio data involve measurement. Most data analysis techniques that apply to ratio data also apply to interval data.

4- Numerical Data

a- **Numerical discrete:** data occur when the observations are integers that correspond with a count of some sort. Some common examples are: the number of bacteria colonies on a plate, the number of cells within a prescribed area upon microscopic examination, the number of heart beats within a specified time interval, a mother's history of number of births (parity) and pregnancies (gravidity), the number of episodes of illness a patient experiences during some time period, etc.

b- **Numerical continuous:** The scale with the greatest degree of quantification is a numerical continuous scale. Each observation theoretically falls somewhere along a continuum. One is not restricted, in principle, to particular values such as the integers of the discrete scale. The restricting factor is the degree of accuracy of the measuring instrument most clinical measurements, such as blood pressure, serum cholesterol level, height, weight, age etc. are on a numerical continuous scale.

Exercises: Identify the type of data (nominal, ordinal, interval and ratio) represented by each of the following. Confirm your answers by giving your own examples.

1. Blood group
2. Temperature (Celsius)
3. Ethnic group
4. Number of heart attacks
5. Calendar year
6. Serum uric acid (mg/100ml)
7. Number of accidents in 3 - year period
8. Number of cases of each reportable disease reported by a health worker