UNIT ONE

Introduction to Statistics

1.1. Learning Objectives

After completing this unit, the trainee will be able to:

- Define statistics
- Enumerate the importance of statistics
- Understand the limitations of statistics

1.2. Introduction

This section deals with the basic concepts and definitions of statistics and related terms.

Definition: The term statistics is used to mean either statistical data or statistical methods.

A. Statistical data: When it means statistical data it refers to numerical descriptions of things. These descriptions may take the form of counts or measurements. Thus statistics of malaria cases in one of the health posts of Ethiopia include fever cases, number of positives obtained, sex and age distribution of positive cases, etc.

NB   Even though statistical data always denote figures (numerical descriptions) it must be remembered that all 'numerical descriptions' are not statistical data.

3

Characteristics of statistical data

In order that numerical descriptions may be called statistics they must possess the following characteristics:

- They must be in aggregates – This means that statistics are 'number of facts.' A single fact, even though numerically stated, cannot be called statistics.

- They must be affected to a marked extent by a multiplicity of causes – This means that statistics are aggregates of such facts only as grow out of a ' variety of circumstances'. Thus the explosion of an outbreak of malaria is attributable to a number of factors, Viz., Human factors, parasite factors, mosquito and environmental factors. All these factors acting jointly determine the severity of the outbreak and it is very difficult for any one to assess the individual contribution of any one of these factors.

- They must be enumerated or estimated according to reasonable standard of accuracy. This means that if aggregates of numerical facts are to be called 'statistics' they must be reasonably accurate. This is necessary because statistical data are to serve as a basis for statistical investigations. If the basis happens to be incorrect the results are bound to be misleading.

- They must have been collected in a systematic manner for a predetermined purpose. Numerical data can be called statistics only if they have been compiled in a properly

4

planned manner and for a purpose about which the enumerator had a definite idea. Facts collected in an unsystematic manner and without a complete awareness of the object, will be confusing and cannot be made the basis of valid conclusions.

- They must be placed in relation to each other. That is, they must be comparable. Numerical facts may be placed in relation to each other either in point of time, space or condition. The phrase, 'placed in relation to each other' suggests that the facts should be comparable.

B. Statistical methods: When the term 'statistics' is used to mean 'statistical methods' it refers to a body of methods that are used for collecting, organising, analyzing and interpreting numerical data for understanding a phenomenon or making wise decisions. In this sense it is a branch of scientific method and helps us to know in a better way the object under study.

C. Functions of Statistics: The proper function of statistics is to enlarge our knowledge of complex phenomena, and to give precision to our ideas that would otherwise remain vague and indeterminate. It widens our knowledge because of the following services that it renders:

5

1. It presents facts in a definite form. Statements (facts) given numerically are definite and hence more convincing than facts stated qualitatively.

Example: a) We have recorded more malaria patients this year than the previous year.

   b) We have recorded 2500 malaria patients this year compared to 1500 of the previous year.

The facts given in part 'b' are definite and more convincing.

2. Statistics simplifies huge and complex mass of data. The raw data are usually huge and should be reduced to some simpler form so that we can understand the main features of the data very easily. The complex data may be reduced to totals, averages, percentages, etc. and presented either in tabular or diagrammatic forms. For example, if the list of the ages of patients who visited a given health facility in the last ten years is given, it will be difficult to understand the age groups which were highly affected. Therefore, these data should be reduced to something simpler (eg., average age) so that we can easily point out the most affected age group.

3. Statistics classifies numerical facts. The procedure of classification helps us to have a better understanding of the variable under investigation.

6

Example: a) the variable sex could be classified as male or female.

b) marital status could be classified as single, married, divorced or widowed.

c) etc.

4. Statistics furnishes a technique of comparison. The facts, having been once classified, are now in a shape when they can be used for purposes of comparisons. Certain facts, by themselves, may be less meaningful unless they are capable of being compared with similar facts at other places or at other periods of time.

5. Statistics endeavours to interpret conditions. Based on the existing facts (eg. disease conditions, access to safe water, availability of latrines, etc.) we can interpret the situation to a greater extent and develop mechanisms which would alleviate the given problem.

D. Importance of statistics: The need for statistics in the smooth functioning of an undertaking in any sector is very high. Students who are interested to know the uses of statistics in many other fields could refer to books listed at the end of this module. The uses of statistics in the health service are described below.

Health Service Statistics: Health statistics are very useful to improve the health situation of the population of a given country. For example, the following questions could not be answered correctly unless the health statistics of a given area is consolidated and given due emphasis.

a) What is the leading cause of death in the area? Is it malaria, tuberculosis, etc.?

b) At what age is the mortality highest, and from what disease?

c) Are certain diseases affecting specified groups of the population more than others? (This might apply, for example, to women or children, or to individuals following a particular occupation.)

d) In comparison with similar areas, is this area healthier or not?

e) Are the health institutions in the area able to cope with the disease problem?

f) Is there any season at which various diseases have a tendency to break out? If so, can these be distinguished?

g) What are the factors involved in the incidence of certain diseases, like malaria, tuberculosis, etc.?

8

The functions/uses of health statistics are enormous. A short list is given below.

Health service statistics are used to:

- describe the level of community health

- diagnose community ills

- discover solutions to health problems and find clues for administrative action

- determine priorities for health programmes

- promote health legislation

- determine the met and unmet health needs

- disseminate information on the health situation and health programmes

- determine success or failure of specific health programmes

- demand public support for health work

E. **limitations of statistics:** Some of the important limitations are given below.

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

9

- It deals with aggregates of facts and no importance is attached to individual items .

- Statistical data are only approximately and not mathematically correct. For example, the age of an individual could be 40 years and 8 months and 10 days, etc. However, we use an approximate value of 41 years.

1.3. Definition of terms

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

- **Descriptive statistics -** One branch of statistics which deals with the description of data in a clear and informative manner using tables and graphs. Also, it refers to the methods used to summarize a body of data with one or two meaningful figures. These are the types of statistics most commonly heard over the radio.

- **Vital statistics** - One branch of descriptive statistics of special relevance in public health is that of vital statistics (the recording of vital events as they occur). The most important vital events are: births, deaths, marriages, divorces, migration and the occurrence of particular diseases. They are used to characterize the health status of a population. Coupled with

10

results of periodic censuses and other special enumeration of populations, the data on vital events relate to an underlying population and yield descriptive measures such as birth rates, morbidity rates, mortality rates, life expectancies, and disease incidence and prevalence rates.

- **population (also called source population or target population or reference population) -** the entire group of interest, to which the investigators would like to generalize the results of the study, and from which a representative sample is to be drawn.

- **sample** - A sample is a part of the population.

Example: A representative sample of 400 under-five children was taken from a population of 2500 under-five children in a certain district to study the level of malnutrition of the area.

11

Exercises

1. What is the difference between statistics and biostatistics?

2. Write a short note on the uses of health service statistics.

3. Discuss about a population and a sample by giving your own examples.

4. What are the common vital events in your area?

5. Mention some of the advantages of recording these vital events.