

Designed for biotechnology students
Biostatistics- FIRST
by
Khalid al jasimee
(Master of AIS)

1. DEFINITIONS, SCOPE AND LIMITATIONS

1.1 Introduction:

In the modern world of computers and information technology, the importance of statistics is very well recognised by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

1.2 Origin and Growth of Statistics:

The word 'Statistics' and 'Statistical' are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of scientific method is of comparatively recent growth. Research particularly into the mathematical theory of statistics is rapidly proceeding and fresh discoveries are being made all over the world.

1.3 Definitions:

Statistics is defined differently by different authors over a period of time. In the olden days statistics was confined to only state affairs but in modern days it embraces almost every sphere of human activity. Therefore a number of old definitions, which was confined to narrow field of enquiry were replaced by more definitions, which are much more comprehensive and exhaustive. Secondly, statistics has been defined in two different ways – Statistical data and statistical methods. The following are some of the definitions of statistics as numerical data.

1. Statistics are the classified facts representing the conditions of people in a state. In particular they are the facts, which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement.

Statistics are measurements, enumerations or estimates of natural phenomenon usually systematically arranged, analysed and presented as to exhibit important inter-relationships among them. Thus statistics of malaria cases in one of malaria detection and treatment include fever cases, number of positives obtained, sex and age distribution of positive cases, etc.

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

1.4.1 Definition by Croxton and Cowden:

Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis. It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one. According to this definition there are four stages:

1. Collection of Data: It is the first step and this is the foundation upon which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of correct method.

2. Presentation of data: The mass data collected should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

3. Analysis of data: The data presented should be carefully analysed for making inference from the presented data such as measures of central tendencies, dispersion, correlation, regression etc.,

4. Interpretation of data: The final step is drawing conclusion from the data collected. A valid conclusion must be drawn on the basis of analysis. A high degree of skill and experience is necessary for the interpretation.

1.5 Functions of Statistics:

There are many functions of statistics. Let us consider the following five important functions.

1. Condensation:

Generally speaking by the word 'to condense', we mean to reduce or to lessen. Condensation is mainly applied at embracing the understanding of a huge mass of data by providing only few observations. If in a particular class in primary School, only marks in an examination are given, no purpose will be served. Instead if we are given the average mark in that particular examination, definitely it serves the better purpose. Similarly the range of marks is also another measure of the data. Thus, Statistical measures help to reduce the complexity of the data and consequently to understand any huge mass of data.

2. Comparison: ii

Classification and tabulation are the two methods that are used to condense the data. They help us to compare data collected from different sources. Grand totals, measures of central tendency measures of dispersion, graphs and diagrams, coefficient of correlation etc provide ample scope for comparison.

If we have one group of data, we can compare within itself. If the rice production (in Tonnes) in Tanjore district is known, then we can compare one region with another region within the district. Or if the rice production (in Tonnes) of two different districts within Tamilnadu is known, then also a comparative study can be made. As statistics is an aggregate of facts and figures, comparison is always possible and in fact comparison helps us to understand the data in a better way.

3. Forecasting:

By the word forecasting, we mean to predict or to estimate before hand. Given the data of the last ten years connected to rainfall of a particular district in Tamilnadu, it is possible to predict or forecast the rainfall for the near future. In business also forecasting plays a dominant role in connection with production, sales, profits etc. The

analysis of time series and regression analysis plays an important role in forecasting.

4. Estimation:

One of the main objectives of statistics is drawn inference about a population from the analysis for the sample drawn from that population. The four major branches of statistical inference are

1. Estimation theory
2. Tests of Hypothesis
3. Non Parametric tests
4. Sequential analysis

In estimation theory, we estimate the unknown value of the population parameter based on the sample observations. Suppose we are given a sample of heights of hundred students in a school, based upon the heights of these 100 students, it is possible to estimate the average height of all students in that school.

5. Tests of Hypothesis:

A statistical hypothesis is some statement about the probability distribution, characterising a population on the basis of the information available from the sample observations. In the formulation and testing of hypothesis, statistical methods are extremely useful. Whether crop yield has increased because of the use of new fertilizer or whether the new medicine is effective in eliminating a particular disease are some examples of statements of hypothesis and these are tested by proper statistical tools.

- **Scope of Statistics:**

Statistics is not a mere device for collecting numerical data, but as a means of developing sound techniques for their handling, analysing and drawing valid inferences from them. Statistics is applied in every sphere of human activity – social as well as physical – like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics cannot be applied. We now discuss briefly the applications of statistics in other disciplines.

Statistics and Agriculture:

Analysis of variance (ANOVA) is one of the statistical tools developed by Professor R.A. Fisher, plays a prominent role in agriculture experiments. In tests of significance based on small samples, it can be shown that statistics is adequate to test the significant difference between two sample means. In analysis of variance, we are concerned with the testing of equality of several population means.

For an example, five fertilizers are applied to five plots each of wheat and the yield of wheat on each of the plots are given. In such a situation, we are interested in finding out whether the effect of these fertilisers on the yield is significantly different or not. In other words, whether the samples are drawn from the same normal population or not. The answer to this problem is provided by the technique of ANOVA and it is used to test the homogeneity of several population means.

Statistics and Medicine:

In Medical sciences, statistical tools are widely used. In order to test the efficiency of a new drug or medicine, t - test is used or to compare the efficiency of two drugs or two medicines, t- test for the two samples is used. More and more applications of statistics are at present used in clinical investigation.

Limitations of statistics:

Statistics with all its wide application in every sphere of human activity has its own limitations. Some of them are given below.

- 1. Statistics is not suitable to the study of qualitative phenomenon:** Since statistics is basically a science and deals with a set of numerical data, it is applicable to the study of only those subjects of enquiry, which can be expressed in terms of quantitative measurements. As a matter of fact, qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed numerically and any statistical analysis cannot be directly applied on these qualitative phenomena. Nevertheless, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.
- 2. Statistics does not study individuals:** Statistics does not give any specific importance to the individual items, in fact it deals with an aggregate of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.
- 3. Statistical laws are not exact:** It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.
- 4. Statistics table may be misused:** Statistics must be used only by experts; otherwise, statistical methods are the most dangerous tools on the hands of the inexperienced. The use of statistical tools by the inexperienced and untrained persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. As King says aptly 'statistics are like clay of which one can make a God or Devil as one pleases'
- 5. Statistics is only, one of the methods of studying a problem:** Statistical method do not provide complete solution of the problems because problems are to be studied taking the background of the country's culture, philosophy or religion into consideration. Thus the statistical study should be supplemented by other evidences.