

# BIO STATISTICS

Shahbaz Baig

**Abstract:** The word “Statistics” which comes from the Latin word status, meaning a political state, originally meant information useful to the state, for example, information about the sizes of populations and armed forces. But this word has now acquired different meanings. Science of collecting summarizing, organizing, analysis, interpretation, presentation and dissemination of data pertaining to vital events. It is the branch of statistics that make the use of statistical methods and general rules in the investigation on specific problem. A collection of facts from which conclusion may be drawn is referred as data.

**Key words:** Statistics, Variable, Ordinal, Nominal, Frequency, Scale.

The word “Statistics” which comes from the Latin word status, meaning a political state, originally meant information useful to the state, for example, information about the sizes of populations and armed forces. But this word has now acquired different meanings.

The earliest writing on statistics was found in a 9th century book entitled: “Manuscript on Deciphering Cryptographic Messages”, written by Al-Kindi (801–873 CE). In his book, Al-Kindi gave a detailed description of how to use statistics and frequency analysis to

decipher encrypted messages, this was the birth of both statistics and cryptanalysis<sup>1,2</sup>.

Some scholars pinpoint the origin of statistics to 1663, with the publication of Natural and Political Observations upon the Bills of Mortality by John Graunt<sup>2</sup>. Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence its stat-etymology. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and the natural and social sciences.

Its mathematical foundations were laid in the 17th century with the development of probability theory by Blaise Pascal and Pierre de Fermat. Probability theory arose from the study of games of chance. The method of least squares was first described by Carl Friedrich Gauss around 1794. The use of modern computers has expedited large-scale statistical computation, and has also made possible new methods that are impractical to perform manually.

**Article Citation:** Baig S, Bio Statistics. Indep Rev Jan-Jun 2017;19(1-6): 53-60.

**Date received:** 18/04/2017

**Date Accepted:** 28/05/2017

**Correspondence Address:**

**Dr. Shahbaz Baig**, MBBS, MPH

Assistant Professor of Community Medicine  
Independent Medical College, Faisalabad.

**1. Dr. Shahbaz Baig**, MBBS, MPH  
Assistant Professor Community Medicine,  
Independent Medical College, Faisalabad.

## What is Statistics?

It is the science of systematic collecting, summarizing, organizing, analysis, interpretation, presentation and dissemination of data.

A.L. Bowel defined statistics as

### “Statistics is the science of counting”.

This definition places the entire stress on counting only. He has also defined statistics as “Science of averages” like, average birth rates, average increase in population, average in per capita income, average increase in standard of living and cost of living.

Prof. Boddingtons has defined statistics as

### “Science of estimates and probabilities”

This definition covers a major part of statistics. It is close to the modern some statistics.

## CHARACTERISTICS OF STATISTICS

The definition stated above indicates that statistics is a subject in its own right. It may therefore be desirable to know the characteristics features of statistics in order to appreciate and understand. Its general nature. Some of its important characteristics are given below.

- Statistics deals with the behavior of aggregates or large groups of data. It has nothing to do with what is happening to a particular individual or object of the aggregate.
- Statistics deals with variability that obscure underlying patterns. No two objects in this universe are exactly alike. If they were, rather than isolated figures.
- Statistics deals with variability that obscure underlying patterns. No two objects in this universe are exactly alike.

If they were, there would have been no statistics problem.

- Statistics deals with uncertainties as every process of getting observations whether controlled or uncontrolled, involves deficiencies or chance variation. That is why have to talk in terms of probability.
- Statistics deals with those characteristics or aspects of things which can be described numerically either by counts or by measurements.
- Statistics deals with those aggregates which are subject to a number of causes such as race, ancestry, age, diet, habits, climate and so forth.
- Statistics laws are valid on the average or in the long run. There is no guarantee that a certain law will hold in all cases. Statistics inference is therefore made in the face of uncertainty.
- Statistics results might be misleading and incorrect if sufficient care in collecting, processing and interpreting the data is not exercised or if the statistical data are handled by a person who is not well versed in the subject matter of statistics.

Are aggregates of facts,

Are numerically expressed,

Are collected for a predetermined purpose

Are estimated according to reasonable standard of accuracy and

Are collected systematic manner<sup>6</sup>.

## STATISTIC

A value concerning some characteristic of a sample is called statistic.

Any measure of the sample is called statistic. The statistics depends upon the sample. Different samples have, in general, different values of the statistic. The symbol  $\bar{x}$  is used for sample mean and  $S$  is used for the sample

variance. These are the statistics (plural for statistic.) The sample statistics are used to draw conclusions about the population parameters. The process of drawing these conclusion is called inference or inferential statistics<sup>6</sup>.

## TYPES OF STATISTICS

- Descriptive statistics
- Inferential statistics
- Applied statistics

### Descriptive Statistics

This deals with collection of data, its presentation of various forms such as tables, graphs and diagrams and finding averages and other measures which would describe the data.

The purpose of descriptive statistics is to present the information in such a way as can be readily help the decision maker.

### Inferential Statistics

This deals with techniques used for analysis of data, making the estimates and drawing conclusions from limited information taken on the sample basis and testing the reliability of estimate.

In drawing conclusions, the decision maker may used of theory of probability to minimize risks.

Modern statistics is mainly concerned with drawing of conclusion from sample about the population.

### Applied Statistics

It is the branch of statistics that make the use of statistical methods and general rules in the investigation on specific problem.

It is used in the field of agriculture, education, health etc.

### Statistical Methods

Statistical methods are those ways that are used to collect, present, analyze, and interpret quantitative data.

### OBSERVATION

A method of data collection in which the situation of interest is watched and the relevant facts, actions and behaviors are recorded<sup>4</sup>.

The value of a variable for an element is called an observation or measurement.

It may be defined as any sort of recording of information is called an observation.

### Population

A population consists of all elements – individuals, items, or objects – whose characteristics are being studied.

### Study Units

The individuals may be called the study units. The study units may be living or non-living things.

A population may be finite or infinite.

It is called finite if its individuals can be counted. For example the population of income-tax payers, the population of smokers in a country.

The population is called infinite when its individuals cannot be counted. For example the number of sun-rays and the number of dust particles on the surface of earth. The number of fish in an ocean may be very large but countable. The numbers of stars in a galaxy may be countable. Such populations

are called countable infinite (very large but countable). The size of the population is denoted by N.

### SAMPLE

A portion of the population selected for study is referred to as a sample.

Sample is any part of the population which is selected from the population. The sample size is denote by n.

### VARIABLE:

A measurable quantity which can vary from one individual or object to another is called variable. A variable is also some time called a variant. A variable is a characteristic under study that assumes different values for different elements. In contrast to a variable, the value of a constant is fixed. Height and weight of an individual, pieces of goods, no of children in a family.

A variable is the item which changes position from time to time, person to person, place, to place and situation to situation..

### TYPES

Variables may be classified as

- 1) QUANTITATIVE
  - a) Discrete
  - b) Continuous
- 2) QUALITATIVE
  - a) Nominal
  - b) Ordinal

### QUANITATIVE

A variable is called a quantitative variable when a characteristic can be expressed numerically such as age, weight or no of children

### Discrete Variable

A variable that can take or assume only certain or some specific value with in a given range. e.g.

No. of houses in a street

No. of children in a family

No. of students in a class

### Continous Variables

A variable that can take or assume any value (typically between certain limits) with in a given range. Most bio-medical variable are continuous. e.g. Height of mercury in a thermometer, speed of a car, height weight of an individual.

### QUALITATIVE VARIABLE

A variable is called a qualitative variable when a characteristic is expressed non numerically such as education, gender or eye colour.

### NOMINAL

These variables, also called “attribute variables” or “categorical variables,” classify observations into a small number of categories. A good rule of thumb is that an individual observation of a nominal variable is usually a word, not a number.

Examples of nominal variables include sex (the possible values are male or female), genotype (values are AA, Aa, or aa), or ankle condition (values are normal, sprained, torn ligament, or broken). Nominal variables are often used to divide individuals up into classes, so that other variables may be compared among the classes.

Nominal variables are often summarized as proportions or percentages.

### Ordinal

Ranked variables, also called ordinal variables, are those for which the individual

observations can be put in order from smallest to largest, even though the exact values are unknown.

Examples of ordinal variables include intensity of pain such as mild, moderate & severe.

## UNIVARIATE AND BIVARIATE VARIABLE

Univariate: one variable,

Bivariate: two variables

UNIVARIATE means “one variable” (one type of data)

Example: Travel Time (minutes): 15, 29, 8, 42, 35, 21, 18, 42, 26. The variable is Travel Time. Age of patient.

BIVARIATE means “two variables”, in other words there are two types of data With bivariate data you have two sets of related data that you want to compare:

So with bivariate data we are interested in comparing the two sets of data and finding any relationships. Example, sex as male or female.

We can use Tables, Scatter Plots, Correlation, Line of Best Fit, and plain old common sense<sup>9</sup>.

## DEPENDENT AND INDEPENDENT VARIABLE

Variable may classified in to dependent and independent variable

### Confounding Variable

It is which is associated both exposure and disease and is distributed unequally in the study and control groups and can distort the study result e.g. Age could be biggest confounding factor in many studies.

## PARAMETER

Any measure of the population is called a parameter.

A value (known/unknown) concerning some characteristic of a population is called parameter. It is a fixed quantity and always to be estimated.

Average age of patients admitted in hospital at a certain time is parameter. It is usually unknown. The symbol  $\mu$  (mue) is use for population mean and the symbol  $\sigma^2$  (sigma-square) is used for population variance. These symbols pertain to the population and are called the parameters. The parameters are usually unknown and are estimated through samples.

## DATA

A collection of facts from which conclusion may be drawn is referred as data. It may defined as

The facts and figures that are collected, analyzed and interpreted.

## TYPES OF DATA

### Primary Data:

The Data initially collected, has lot of unnecessary irrelevant and unwanted information and has not undergone any sort of statistical treatment. This data is called raw data of primary data.

### Secondary Data:

The data, which has undergone at least one sort of statistical treatment is called secondary data. It is data that has been processed profaned. To make secondary data the raw data (Original data) is by removing unwanted irrelevant and unnecessary information and presented in satirical Shane

such as polentas. Graph are table etc.

### **Cooked Data:**

The data that has been collected genuinely and is fictitious is called cooked data

## **COLLECTION OF DATA**

The first step in any investigation is collection of data. The data may be collected for whole population or for a sample only. It is mostly collected on sample and basis. Collection of data is not an easy task for this job, the staff is trained and work is assigned to them.

## **COLLECTION OF PRIMARY DATA**

Primary data are collected by the following method.

### **Direct Personal investigations**

In this method the enumerator goes to the respondents and interviews them. The information supplied by the informants is recorded by the enumerator. The information collected by this method is considered accurate and complete. The quality of the information (data) depends upon the two persons i.e., the respondent and the enumerator. This method of collecting data is very costly because a team of enumerators is required to get the observations. It is suitable for small local inquiries. It is not suitable if the area of inquiry is wide. Lot of funds and human resources are required for an extensive field of inquiry.

### **Indirect Personal Interviews**

This method of collecting primary data is adopted when the respondents do not supply the correct information. They do not cooperate with the enumerators. If people are asked to supply information about their income, their houses and other holdings, they are quite likely to give wrong information. In this method some agency other than the

respondent is selected to supply the data. For example if data is required from shop keepers, their trade unions can be consulted to supply the data about their members. Indirect questions can also be asked from the respondent so that he speaks out the correct figures. The respondent can be asked about the size of his house, his facilities, and other services like gas, telephone, and electricity etc. By indirect questions the respondent can be forced to give the correct answer.

### **Questionnaire Method**

Questions is a list of questions relating to the field of inquiry. This list of questions is sent by mail to the respondents with a request that they send it back after duly filling the entries. This is a cheap method but this method is not successful in areas where people are not literate and cooperative. It is possible that the people do not return the questions because they are afraid of giving information in writing. The framing of questionnaire is quite difficult. Only an expert can frame a good questionnaire. A questionnaire should have the following characteristics.

- Its language should be simple.
- Questions should have an objective answer like 'yes' or 'no'.
- Number of questions should not be very large.

This method is very popular and is being used in big cities where most of the people are literate.

### **Collection through Enumerators**

This method is used to overcome the difficulties of the illiterate people in filling the questionnaire. The trained enumerators go to the informants and they assist them in filling the questionnaire correctly. This is a very costly inquiry and can be carried out by

the Govt. only.

### Collection through Local Sources

These are the reports of the local representative sent to the investigator. The data is not formally collected by local representatives.

They only send the reports. In our country the Patwari sends regular reports to the Govt. regarding the crop conditions in our country. These reports are only the estimates based on the judgments.

### Computer interviews

Respondents enter data directly into computer in response to question presented on the monitor

## COLLECTION OF SECONDARY DATA

The secondary data are not collected from individual units but are taken from different sources called secondary sources. Some of these sources are as follows:

### Internal Secondary Data

Data generated within the organization itself such as no of admission or discharge patients in hospital.

### External Secondary Data

#### Official Statistics

All the government departments maintain their data and publish it annually as official statistics. The important publications are those of the Statistics Division and Bureaus of Statistics in different provinces.

#### Semi Official

The data available with PIA, railways, State Bank of Pakistan, district councils etc. are called semi official.

#### Others

Secondary data can also be obtained from

magazines, news papers and journals of the universities[6].

## MEASUREMENT SCALES

By measurement, usually mean the assigning of numbers to observations or objects and scaling is a process of measuring. The four scales of measurement of briefly mentioned below.

### Nominal Scale

The classification or grouping of the observations into mutually exclusive qualitative categories or classes is said to constitute a nominal scale. For example students are classified as male and female. Number 1 and 2 may also be used to identify these two categories. Similarly, rainfall may be classified very heavy, moderate and light. We may used number 1,2 and 3 two denote the three classes of rainfall. The numbers when they are used only two identify the categories of the given scale, carry no numerical significance and there is no particular order for the grouping.

### Ordinal and Ranking Scale:-

It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurements. For example, the performance of students (or players) is rated has excellent, good, fair or poor, etc. Number 1,2,3,4, etc. are also used to indicate ranks. The only relation that holds between any pair of categories is that of greater than ( or more preferred).

### Interval scale

A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale. Temperature measured on either the Celsius or the Fahrenheit ale in a outstanding example of

interval scale because the same difference exists between the 20°C (68°F) and 30°C (86°F) as between 5°C (41°F) and 15°C (59°F). It cannot be said that a temperature of 40 degrees is twice as hot as a temperature of 20 degree, i.e. is ratio 40/20 has no meaning. The arithmetic operation of addition, subtraction, etc. Are meaningful.

### Ratio Scale

It is a special kind an interval scale where the scale of measurement has a true zero point as its origin. The ratio scale is used to measure weight, volume, length, distance, money, etc. the key to differentiating interval and ratio scale is that the zero point meaningful for ratio scale.[10]

## BIOSTATISTICS

Science of collecting summarizing, organizing, analysis, interpretation, presentation and dissemination of data pertaining to vital events.

Branch of statistic which concern itself with the data and laws relating to vital events of human life.

### Uses of biostatistics

- It is uses to test whether difference between two populations regarding a particular attribute is real or chance of occurrence.
- It is also used to study the correlation/ association between two or more attributes in the same population.
- It is used a evaluate the efficiency of vaccines, sera etc by control studies.

- It is used to locate, define and measure the extent of morbidity and mortality in the community/population.
- To evaluate the achievements of public health programs.
- To fix priorities in public health program.

## REFERENCE

1. Singh, Simon (2000). The code book : the science of secrecy from ancient Egypt to quantum cryptography (1st Anchor Books ed.). New York: Anchor Books. ISBN 0385495323.^ Ibrahim A. Al-Kadi
2. "The origins of cryptology: The Arab contributions", Cryptologia, 16(2) (April 1992) pp. 97–126.
3. Willcox, Walter (1938) The Founder of Statistics. Review of the International Statistical Institute 5(4):321–328 Laura
4. L, observation.Guide<http://marketing.about.com/od/marketingglossary/g/observationdef.htm> McDonald, J.H. 2009.
5. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.
6. Hanif M,Ahmad M.Biostatistics for health students with manual on software applications(1st edition). Islamic society of statistical sciences Pakistan. ISBN969-8858-008.
7. Pierce, Rod. "Univariate and Bivariate Data" Math Is Fun. Ed. Rod Pierce. 4 Aug 2011. 27 Feb 2012 <http://www.mathsisfun.com/data/univariate-bivariate.html>.
8. Pasha GR. A quick approach to statistics(2011 edition).Ilmi kitab khana Lahore.