BIOSTATISTICS

Statistic or datum means a measured or counted fact or piece of information stated as a figure such as height of one person, birth weight of a baby, etc.

Statistics or data would be plural of the same, stated in more than one figures such as height of 2 persons, birth weight of 5 babies, etc.

How they are collected? from experiments, records and surveys, in all walks of life such as economics, politics, education, industry, business, administration, etc. Medicine too, including Preventive Medicine and Public Health, is one such field.

Statistics though apparently plural, when used in a singular sense, is a science of figures. It is a field of study concerned with techniques or methods of collection of data, classification, summarizing, interpretation, drawing inferences, testing of hypotheses, making recommendations, etc. when only a part of data is used. In any book on statistical methods the word statistics is used both as a plural of statistic and as a science of figures.

BIOSTATISTICS

Biostatistics is the term used when tools of statistics are applied to the data that is derived from biological sciences such as medicine. Everything in medicine be it research, diagnosis or treatment, depends on counting or measurement.

High or low blood pressure has no meaning, unless it is expressed in figures.

Incidence of tuberculosis or death rate in typhoid is stated in figures.

Enlargement of spleen is measured in fingers' breadth.

Thus medical statistics or biostatistics can be called quantitative medicine.

In nature, blood pressure, pulse rate, action of a drug or any other measurement or counting varies not only from person to person but also from group to group. The extent of this variability in an attribute or a character, whether it is by chance, i.e. biological or normal, is learnt by studying statistics as a science. Comparison of a variable in two or more groups is of great importance in applied scientific practice of medicine, e.g. infant mortality rate in developing countries like Pakistan and in developed countries like UK, USA due to external factors like socioeconomic advancement, better application of scientific knowledge in medicine or improved health services.

Rise in pulse rate noted after an injection of a drug may be by chance or due to the effect of drug. Variation more than natural limits may be pathological, i.e. abnormal due to the play of certain external factors. Hence, biostatistics may also be called a science of variation.

The data after collection, lying in a haphazard mass are of no use, unless they are properly sorted, presented, compared, analyzed and interpreted.

A medical student should not depend on a statistician for the statistical analysis. For professional interpretation of his results, he should learn the application of methods himself which do not require knowledge of mathematics higher than what he or she had acquired at school. However, he or she should take the guidance of a qualified statistician right from the beginning of any scientific study till drawing the conclusions.

Medical Statistics go under different names when applied in different names when applied in different fields such as:

•• Health statistics in public health or community health.

•• Medical statistics in medicine related to the study of defect, injury, disease, efficacy of drug, serum and line of treatment, etc.

•• Vital statistics in demography pertaining to vital events of births, marriages and deaths.

These terms are overlapping and not exclusive of each other.

Application and Uses of Biostatistics as a Science

In physiology and anatomy

i. To define what is normal or healthy in a population and to find limits of normality in variables such as weight and pulse rate—the mean pulse rate is 72 per minute but up what limits it may be normal on either side of mean has to be established with certain appropriate techniques.

ii. To find the difference between means and proportions of normal at two places or in different periods. The mean height of boys in Gujrat is less than the mean height in Sargodha. Whether this difference is due to chance of a natural variation or because of some other factors such as better nutrition playing a part, has to be decided. To find the correlation between two variables X and Y such as height and weight—whether weight increases or decreases proportionately with height and if so by how much, has to be found.

In pharmacology

i. To find the action of drug—a drug is given to animals or humans to see whether the changes produced are due to the drug or by chance.

ii. To compare the action of two different drugs or two successive dosages of the same drug.

iii. To find the relative potency of a new drug with respect to a standard drug.

In medicine

i. To compare the efficacy of a particular drug, operation or line of treatment—for this, the percentage cured, relieved or died in the experiment and control groups, is compared and difference due to chance or otherwise is found by applying statistical techniques.

ii. To find an association between two attributes such as cancer and smoking or filariasis and social class—an appropriate test is applied for this purpose.

iii. To identify signs and symptoms of a disease or syndrome. Cough in typhoid is found by chance and fever is found in almost every case. The proportional incidence of one symptom or another indicates whether it is a characteristic feature of the disease or not.

In community medicine and public health

i. To test usefulness of sera and vaccines in the field—percentage of attacks or deaths among the vaccinated subjects is compared with that among the unvaccinated ones to find whether the difference observed is statistically significant.

ii. In epidemiological studies—the role of causative factors is statistically tested. Deficiency of iodine as an important cause of goiter in a community is confirmed only after comparing the incidence of goiter cases before and after giving iodized salt.

In public health, the measures adopted are evaluated. Lowering of morbidity rate in typhoid after pasteurization of milk may be attributed to clean supply of milk, if it is statistically proved.

Fall in birth rate may be the result of family planning methods adopted under National Family Welfare

Programme or due to rise in living standards, increasing awareness and higher age of marriage.

<mark>Thus, by learning the methods in biostatistics, a student learns to evaluate articles published in medical journals or papers read in medical conferences. He understands the basic methods of observation in his clinical practice or research.</mark>

Application and Uses of Biostatistics as Figures

Health and vital statistics are essential tools in demography, public health, medical practice and community services.

Recording of vital events in birth and death registers and diseases in hospitals is like book keeping of the community, describing the incidence or prevalence of diseases, defects or deaths in a defined population.

Such events properly recorded form the eyes and ears of a public health or medical administrator, otherwise it would be like sailing in a ship without compass.

Thus, biostatistics as a science of figures will tell:

a. What are the leading causes of death?

b. What are the important causes of sickness?

c. Whether a particular disease is rising or falling in severity and prevalence?

d. Which age group, sex, social class of people, profession or place is affected the most?

e. The levels or standards of health reached.

f. Age and sex composition of population in a community.

g. Whether a particular population is rising, falling, aging or ailing?

h. Which health program should be given priority and what will be the requirements for the same?

Common Statistical Terms

One should remember before learning the methods in biostatistics, some terms used, their symbols and notations and refer as and when needed.

<mark>1. Variable:</mark> A characteristic that takes on different values in different persons, places or things. A quantity that varies within limits such as height, weight, blood pressure, age, etc. It is denoted as X and notation for orderly series as X_1, X_2, X_3… X_n. The suffix n is symbol for number in the series.

Σ (sigma) stands for summation or results or observation.

<mark>2. Constant:</mark> Quantities that do not vary such as π = 3.1416, e = 2.7183. They do not require statistical study. In biostatistics, mean, standard deviation, standard error, correlation coefficient and proportion of a particular population are considered as constant.

<mark>3. Observation:</mark> An event and its measurements such as blood pressure (event) and 120 mmHg (measurement).

<mark>4. Observational unit:</mark> The source that gives observations such as object, person, etc. In medical statistics the term individuals or subjects is used more often.

<mark>5. Data:</mark> A set of values recorded on one or more observational units. Data are raw materials of statistics.

<mark>6. Population:</mark> It is an entire group of people or study elements— persons, things or measurements for which we have an interest at a particular time.

Populations are determined by our sphere of interest. It may be infinite or finite. If a population consists of fixed number of values, it is said to be finite. If population consists of an endless succession of values. The population is an infinite one.

It has to be fully defined such as all human beings, all families joint or nuclear, all women of 15–45 years

of age or only married women, all patients, all doctors in service or in practice and so on. Such a population invariably gives qualitative data. If it is finite or limited in number it can easily be counted. A statistical population may also be birth weights, hemoglobin levels, readings of a thermometer, number of RBCs in the human body, etc. Such a population mostly gives quantitative data. It is finite or small in number or infinite or unlimited in number that cannot be easily counted.

7. Sampling unit: Each member of a population.

8. Sample: It may be defined as a part of a population. It is a group of sampling units that form part of a population, generally selected so as to be representative of the population whose variables are under study. There are many kinds of sample that can be selected from a population. Various methods.

9. Parameter: It is a summary value or constant of a variable that describes the sample such as its mean, standard deviation, standard error, correlation coefficient, proportion, etc. This value is calculated from the sample and is often applied to population but may or may not be a valid estimate of population. Though not desirable, parameter and statistic are often used as synonyms.

11. Parametric test: It is one in which population constants as described above are used such as mean, variances, etc. and data tend to follow one assumed or established distribution such as normal, binomial, Poisson, etc.

12. Nonparametric tests: Tests such as χ^2 test, in which no constant of a population is used. Data do not follow any specific distribution and no assumptions are made in nonparametric tests, e.g. to classify good, better and best you allocate arbitrary numbers or marks to each category.

Notations for a Population and Sample Values
Roman letters are used for statistics of samples and Greek for parameters of population.
Common notations are:

Summary measures	Sample statistics	Population parameters
Mean	$\bar{X}$	μ
Standard deviation	s	S
Variance	s2	$\sigma2$
Proportion	p	P or π
Complement of proportion	q	Q or $(1-\pi)$

Other symbols commonly used are:
= : Equal to
> : Greater than
$\geq$: Greater than or equal to
< : Less than
$\leq$: Less than or equal to
Z : The number of standard deviations from the mean or standard normal deviate/variate
% : Percent
r : Pearson's correlation coefficient
p : Spearman's correlation coefficient
O : Observed number

E : Expected number
d.f. or df : Degrees of freedom
k : Number of groups or classes
P (A) : Probability of event A