Quadrant II – Transcripts and Related Matter

Programme: Bachelor of Science (Third Year)

Subject: BOTANY

Paper Code: BOD -102

Paper Title: Research Methodology, Bioinformatics and Biostatistics

Unit: 6

Module Name: Basic Principles of Biostatistics

Name of the Presenter: Roxiette Heromina Siqueira

Statistics is a branch of science which deals with: collection of data, organizing and summarizing the data, analyses of data and making inferences, or decisions and predictions (Montgomery and Runger).

Biostatistics is the application of statistical methods to the problems of biology, Including human biology, medicine, public health, agriculture, veterinary and genetics.

Biostatistics is also known as Biometry which means 'biological measurement'.

Biometry is derived from two Greek words ('*Bios*' means life and '*metron*' means measured.

Biostatistics can be divided into 2 subcategories: Descriptive and Inferential Biostatistics

HISTORY

The science of biostatistics was initiated by Sir Francis Galton in 1911.

Biostatistics was in its developing phase when Mendel published the details of his Breeding experiments.

The real advancement in this field took place in the first half of twentieth Century.

Karl Pearson and his coworkers established a school of Biometry in the University of London.

Sir Ronald, A. Fisher, S. Wright, K. Mather have made commendable contributions in the field of Biometry.

APPLICATIONS OF BIOSTATISTICS

Biostatistics has wide applications in various fields of biology.

Statistical methodologies are now an essential part of Research in Genetics, Plant Breeding, Ecology, Agriculture, Forestry, Biogeography, Taxonomy and other disciplines of Life Science.

The knowledge of biostatistics is important for biologists for the following reasons:

It helps in understanding the nature of variability,

It helps in deriving general laws from small samples

It helps in making predictions for future planning.

DATA COLLECTION

Data is a collection of observations expressed in numerical figures

Data collection is the process of gathering the desirable information carefully, with least possible distortion, so that the analysis may provide answers that are credible and stand to logic (Sapsford &Jupp).

Data collection is a systematic process of gathering observations or measurements. (Census methods, Sampling Methods)

ORGANISATION OF DATA

Tabulation may be defined as presentation of classified data in a scientific manner and in an orderly sequence so as to bring out its essential features and main characteristics.

PRESENTATION OF DATA

The representation of quantitative data suitably through charts and diagrams is known as Graphical Representation of Statistical Information.

Graph includes both charts and diagrams.

- Graphical representation
    - Line graph
    - Bar diagram
    - Pie chart
    - Pictogram
    - Histogram

Central Tendency

The primary aim of statistical methods is to condense the raw data in a way that it is able to convey some meaningful conclusion at the first sight.

Summarization, condensation and classification simplify the data and improve the understanding of main characteristics of the data.

Sometimes, data is condensed to a single value. Such a single value expression or presentation of data is called central value.

The values of variable tend to concentrate around the central value. Therefore, the central value is also called the central tendency.

The measures devised to calculate the central tendency are known as measures of **central tendency**.

Measures of Central Tendency refer to all those methods of statistical analysis which are used to estimate or calculate the average of a set of data.

Three common measures of Central Tendency are Mean, Median and Mode.

A **measure of dispersion** presents the deviation of value of individual observations around the central value in a set of data.

60,60,60,60 mean=60

30, 50,85,75,60 mean=60

10, 60,90,90,50 mean =60

- Absolute Measures
    - Range
    - Quartile Deviation
    - Mean Deviation
    - Standard Deviation
- Relative Measures
    - Co-efficient of Variation
    - Co-efficient of Quartile Deviation
    - Co-efficient of Mean Deviation

Correlation

The statistical tool for measuring the degree of relationship between two variables is known as correlation.

The change in one variable results in positive or negative change in the other variable or a greater change in one variable results in greater or smaller change in the other variable

- Based on direction of change in the value of 2 variables
    - Positive Correlation
    - Negative Correlation
    - Partial Positive Correlation
    - Partial Negative Correlation

When two variables are correlated, it is quite possible to estimate or predict the value of one variable for a given value of another variable.

The statistical tool which helps to estimate the unknown value of one variable from the known value of the related variable is called **regression**.

The term was first coined by Sir Francis Galton

The regression analysis can be of two types:

Simple Regression-regression analysis confined to the study of only two variables at a time

Multiple Regression: regression analysis confined to the study of more than two variables at a time

Frequency Distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side (Croxton and Gowden).