



A Gateway to all Postgraduate Courses



An MHRD project under its National Mission on Education through ICT (NME-ICT)

Subject: Statistics

Paper: Biostatistics

Module1:Introduction to Biostatistics

Principal investigator: *Dr. Bhaswati Ganguli, Professor,
Department of Statistics, University of Calcutta*

Paper co-ordinator: *Dr. Sugata SenRoy, Professor, Department of
Statistics, University of Calcutta*

Content writer: *Dr. Atanu Bhattacharjee, Assistant Professor,
Centre for Cancer Epidemiology, The Advanced Centre for
Treatment, Research and Education in Cancer (ACTREC)
Tata Memorial Centre*

Content reviewer: *Department of Statistics, University of Calcutta*

What is Biostatistics?

It is the application of statistics to a wide range of topics in biology ,medical research or public health.

The methods used in dealing with statistics in the fields of medicine, biology and public health for planning, conducting and analyzing data which arise in investigations of these branches.

What is Biostatistics?

It is the application of statistics to a wide range of topics in biology ,medical research or public health.

The methods used in dealing with statistics in the fields of medicine, biology and public health for planning, conducting and analyzing data which arise in investigations of these branches.

Importance of Biostatistics

(I) Medicine is becoming increasingly quantitative.

(II) The planning, conduct and interpretation of much of medical research are becoming increasingly reliant on the statistical methodology.

(III) Statistics pervades the medical literature.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

- (I) Design of study.
- (II) Sample size Determination.
- (III) Epidemiology.
- (IV) Clinical Trial.
- (V) Survival Analysis.
- (VI) Data Management.
- (VII) Choice of Proper Statistical tool for Analysis.

Cohort Study

It is straightforward design to allow direct comparisons of exposed and unexposed persons and can provide measures of effects for various outcome events, like e.g. different endpoints (morbidity, mortality, pre-morbidity) and or different diseases.

Case Control Study

(I) Cases Number of people with the disease under study. (II) Controls Number of people who are free of the disease. The cases and controls are then explored to observe the risk factors that differ between them.

Cross-sectional study

A cross-sectional study examines the relationship between disease (or other health related state) and other variables of interest as they exist in a defined population at a single point in time or over a short period of time.

Sample size calculation

Sample size calculation is usually performed through controlling the type I and type II errors.

Precision Analysis

Sample size is calculated in such a way that there is a desired precision at a fixed confidence level (i.e., fixed type I error). It is simple and easy to perform.

Pre-Study Power Analysis

It provides a sample size for achieving a desired power for detecting a clinically/scientifically meaningful difference at a fixed type I error rate.

Sample size calculation

Sample size calculation is usually performed through controlling the type I and type II errors.

Precision Analysis

Sample size is calculated in such a way that there is a desired precision at a fixed confidence level (i.e., fixed type I error). It is simple and easy to perform.

Pre-Study Power Analysis

It provides a sample size for achieving a desired power for detecting a clinically/scientifically meaningful difference at a fixed type I error rate.

Sample size calculation

Sample size calculation is usually performed through controlling the type I and type II errors.

Precision Analysis

Sample size is calculated in such a way that there is a desired precision at a fixed confidence level (i.e., fixed type I error). It is simple and easy to perform.

Pre-Study Power Analysis

It provides a sample size for achieving a desired power for detecting a clinically/scientifically meaningful difference at a fixed type I error rate.

Survival Analysis

Survival Analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs.

Time

Time refers to the time elapsed from some suitably chosen baseline when the event of interest occurs.

Event

By event we mean death, disease incidence, relapse from remission, recovery (e.g., return to work) or any designated experience of interest that may happen to an individual.

Clinical Trial [Pocock ,1983]

Any form of planned experiment which involves patients and is designed to elucidate the most appropriate treatment of future patients under a given medical conditon.

Epidemiology

Epidemiology is the "study of distribution and determinants of health-related states among specified populations and the application of that study to the control of health problems."

Requirement of Biostatistics in Epidemiology

- (I) Explain how disease is measured in populations, calculate, interpret and communicate measures of association and difference.
- (II) Critically analyse the strengths and weaknesses of different epidemiological study designs.
- (III) Interpret confidence intervals, p-values and sample size and communicate their meaning.

Requirement of Biostatistics in Epidemiology

- (IV) Application of biostatistical principles to critically evaluate and interpret and communicate findings from epidemiological research.
- (V) Explain and contextualise the concepts of population, sampling, measurement, bias, confounding and causation;

Data Managment

The objective of data managment is to main the data quality.

The quality of the data will affect the quality of the analyses performed on the data.

At the close of the study, there is a particularly strong emphasis on checking the quality of the data.

The measures with Biostatistical tool like outlier detection, consistency check and other queries raised through application of biostatistical tool enrich the data quality.

The application of Biostatistics completely unavoidable for datamangment.

Data Managment

The objective of data managment is to main the data quality. The quality of the data will affect the quality of the analyses performed on the data.

At the close of the study, there is a particularly strong emphasis on checking the quality of the data.

The measures with Biostatistical tool like outlier detection, consistency check and other queries raised through application of biostatistical tool enrich the data quality. The application of Biostatistics completely unavoidable for datamangment.

Data Managment

The objective of data managment is to main the data quality. The quality of the data will affect the quality of the analyses performed on the data.

At the close of the study, there is a particularly strong emphasis on checking the quality of the data.

The measures with Biostatistical tool like outlier detection, consistency check and other queries raised through application of biostatistical tool enrich the data quality. The application of Biostatistics completely unavoidable for datamangment.

Data Managment

The objective of data managment is to main the data quality. The quality of the data will affect the quality of the analyses performed on the data.

At the close of the study, there is a particularly strong emphasis on checking the quality of the data.

The measures with Biostatistical tool like outlier detection, consistency check and other queries raised through application of biostatistical tool enrich the data quality.

The application of Biostatistics completely unavoidable for datamangment.

Data Managment

The objective of data managment is to main the data quality. The quality of the data will affect the quality of the analyses performed on the data.

At the close of the study, there is a particularly strong emphasis on checking the quality of the data.

The measures with Biostatistical tool like outlier detection, consistency check and other queries raised through application of biostatistical tool enrich the data quality.

The application of Biostatistics completely unavoidable for datamangment.

Statistical tool for Analysis

- (I) Descriptive Statistics.
- (II) Inferential Statistics.

Descriptive statistics

The extent to which the observations cluster around a central location is described by the central tendency and the spread towards the extremes is described by the degree of dispersion.

Measures of central tendency

The measures of central tendency are mean, median and mode. The mean (or the arithmetic average) is the sum of all the scores divided by the number of scores. Mean may be influenced profoundly by the extreme variables.

Normal distribution or Gaussian distribution

Most of the biological variables usually cluster around a central value, with symmetrical positive and negative deviations about this point. The standard normal distribution curve is a symmetrical bell-shaped.

Skewed distribution

It is a distribution with an asymmetry of the variables about its mean

Inferential statistics

In inferential statistics, data are analysed from a sample to make inferences in the larger collection of the population.

The purpose is to answer or test the hypotheses.

A hypothesis is a proposed explanation for a phenomenon.

Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects.

Inferential statistics

In inferential statistics, data are analysed from a sample to make inferences in the larger collection of the population.

The purpose is to answer or test the hypotheses.

A hypothesis is a proposed explanation for a phenomenon.

Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects.

Inferential statistics

In inferential statistics, data are analysed from a sample to make inferences in the larger collection of the population.

The purpose is to answer or test the hypotheses.

A hypothesis is a proposed explanation for a phenomenon.

Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects.

Inferential statistics

In inferential statistics, data are analysed from a sample to make inferences in the larger collection of the population.

The purpose is to answer or test the hypotheses.

A hypothesis is a proposed explanation for a phenomenon.

Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects.

Inferential statistics

Probability is the measure of the likelihood that an event will occur.

Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

In inferential statistics, the term 'null hypothesis' (H_0) denotes that there is no relationship (difference) between the population variables in question.

Alternative hypothesis (H_1) denotes that a statement between the variables is expected to be true.

Inferential statistics

Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

In inferential statistics, the term 'null hypothesis' (H_0) denotes that there is no relationship (difference) between the population variables in question.

Alternative hypothesis (H_1) denotes that a statement between the variables is expected to be true.

Inferential statistics

Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

In inferential statistics, the term 'null hypothesis' (H_0) denotes that there is no relationship (difference) between the population variables in question.

Alternative hypothesis (H_1) denotes that a statement between the variables is expected to be true.

Inferential statistics

Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

In inferential statistics, the term 'null hypothesis' (H_0) denotes that there is no relationship (difference) between the population variables in question.

Alternative hypothesis (H_1) denotes that a statement between the variables is expected to be true.

Inferential statistics

(I) Parametric Test.

(II) Non-Parametric Test.

Inferential statistics

(I) Parametric Test.

(II) Non-Parametric Test.

Inferential statistics

(I) Parametric Test.

(II) Non-Parametric Test.

Inferential statistics-Parametric tests

(I) Student's t-test.

(II) Analysis of variance.

(III) Repeated measures analysis of variance.

Inferential statistics-Parametric tests

(I) Student's t-test.

(II) Analysis of variance.

(III) Repeated measures analysis of variance.

Inferential statistics-Parametric tests

- (I) Student's t-test.
- (II) Analysis of variance.
- (III) Repeated measures analysis of variance.

Inferential statistics-Non-Parametric tests

(I) Sign test.

(II) Wilcoxon's signed rank test.

Inferential statistics-Non-Parametric tests

(I) Sign test.

(II) Wilcoxon's signed rank test.