

MLIS semester II
(paper – 201)
Research Methodology
Topic – **Statistical methods**
e-content

MADHU
Guest Faculty,
Dept. of library & Information Science
Patna University.

Email Id – madhu.gfatm@gmail.com

Statistical methods

Statistics is a term used to summarize a process that an analyst uses to characterize a data set. If the data set depends on a sample of a larger population, then the analyst can develop interpretations about the population primarily based on the statistical outcomes from the sample. Statistical analysis involves the process of gathering and evaluating data and then summarizing the data into a mathematical form.

Statistics is used in various disciplines such as psychology, business, physical and social sciences, humanities, government, and manufacturing. Statistical data is gathered using a sample procedure or other method.

Statistical methods

Two types of statistical methods are used in analyzing data:

- Descriptive statistics
- Inferential statistics.

Descriptive statistics

Descriptive statistics are used to synopsise data from a sample exercising the mean or standard deviation.

Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the mean, median, and mode, while measures of variability include the standard deviation, variance, the minimum and maximum variables.

It gets the summary of data in a way that meaningful information can be interpreted from it. Using descriptive analysis, we do not get to a conclusion however we get to know what in the data is i.e. we get to know the quantitative description of the data.

For instance, consider a simple example in which you must determine how well the student performed throughout the semester by calculating the average. This average is nothing but the sum of the score in all the subjects in the semester by the total number of subjects. This single number is describing the general performance of the student across a potentially wide range of subject experiences.

Whenever we try to describe a large set of observations with a single value, we run into the risk of either distorting the original data or losing any important information. The student average won't determine the strong subject of the student. It won't tell you the specialty of the student or you won't come to know which subject was easy or strong. In spite of these limitations, Descriptive statistics can provide a powerful summary which may be helpful in comparisons across the various unit.

There are two types of statistics that are used to describe data:

- **Measures of central tendency:** In this, a single value attempts to describe the data by using its central position with the given set. They are also classified as a summary set. In order to get the central value, they use averaging(mean), median or mode.
- **The measure of spread:** In this, the data is summarized by describing how well the data is spread out. For example, if the mean score of 100 students is 55 then there will be students whose score will be less than 55 or more than 55. Which means their score will be spread out in a way that their mean is 55. To describe the spread, we can use either of the statistical technique i.e. range, quartiles, variation, standard deviation, and absolute deviation.

Measures of Descriptive Statistics

All descriptive statistics are either measures of central tendency or measures of variability, also known as measures of dispersion. Measures of central tendency focus on the average or middle values of data sets, whereas measures of variability focus on the dispersion of data. These two measures use graphs, tables and general discussions to help people understand the meaning of the analyzed data.

Measures of central tendency describe the centre position of a distribution for a data set. A person analyzes the frequency of each data point in the distribution and describes it

using the mean, median or mode, which measures the most common patterns of the analyzed data set.

Measures of variability, or the measures of spread, aid in analyzing how spread out the distribution is for a set of data. For example, while the measures of central tendency may give a person the average of a data set, it does not describe how the data is distributed within the set. So while the average of the data may be 65 out of 100, there can still be data points at both 1 and 100. Measures of variability help communicate this by describing the shape and spread of the data set. Range, quartiles, absolute deviation and variance are all examples of measures of variability. Consider the following data set: 5, 19, 24, 62, 91, 100. The range of that data set is 95, which is calculated by subtracting the lowest number (5) in the data set from the highest (100).

Inferential statistics

Inferential statistics are used when data is viewed as a subclass of a specific population.

The group of data that contains the information we are interested in is known as population. Inferential Statistics is used to make a generalization of the population using the samples. Where

the sample is drawn from the population itself. It is necessary that the samples properly demonstrate the population and should not be

biased. The process of achieving these kinds of samples is termed as sampling. Inferential Statistics comes from the fact that the sampling naturally incurs sampling errors and is thus not expected to perfectly represent the population.

There are two types of Inferential Statistics method used for generalizing the data:

- Estimating Parameters
- Testing of Statistical Hypothesis