INFERENTIAL STATISTICS

Varnika Mulay

Main concepts of inferential statistics:

- Population: The entire dataset or values which is required for our study or analysis is called population.
- Sample: A sample is a subset of the population that is used to represent the entire population and make inferences about the population.

Inferential statistics is a subfield of statistics that deals with inferring a population's characteristics from a sample collected from it. Inferential statistics goes beyond descriptive statistics by finding correlations between variables, testing hypotheses, and making predictions. On the other hand, descriptive statistics concentrates on summarising and characterizing the features of a dataset.

Inferential statistics involves selecting a sample from a population to perform analysis on. This process is called “Sampling”.

Sampling is a process in which a group or a sample of data is selected from the population which accurately defines or represents the population so as to perform analysis on it. This helps to save time and space as it requires a huge amount of time and space to analyze the entire population.

Types of sampling:

1. **Simple Random Sampling**

   This is a type of sampling in which every data value of the population has an equal chance of being selected in the sample. It is the most simple and straightforward method of sampling.

2. **Stratified Sampling**

   Stratified sampling is a method in which the population is divided into different subgroups also known as strata. These subgroups are created by clustering similar data values in one group. After the subgroups are created, the data points are randomly selected from each stratum to be considered in

the sample. This basically ensures that each subgroup has a representative in the main sample.

3. Systematic Sampling

As the name suggests, systematic sampling is a method that involves selecting the data points in a predefined order. The first member of the sample is selected randomly, and the others are selected at regular intervals, i.e., every n-th data value of the population is selected.

Central Limit Theorem (CLT)

The Central Limit Theorem (CLT) is a statistical theorem that states that the "the distribution of the sample mean will approach a normal distribution as the sample size becomes larger, regardless of the population's distribution, provided the samples are independent and identically distributed."

Estimation

Estimation is a process in inferential statistics used to infer the value of a population parameter based on sample data. It includes point estimation and interval estimation.

Hypothesis and Hypothesis Testing

A hypothesis is a statement about the chosen population that is to be analyzed or tested using statistical analysis. To do this, we need to perform hypothesis testing which is a procedure used to decide whether to accept or reject the hypothesis based on sample data.