Assignment – 3

Various statistical measures which are used in the analysis of quantitative data

The term Quantitative data analysis refers to analyzing data that is based on numbers or data that can be easily converted into numbers without losing its meaning. This type of data analysis involves statistical analysis methods which is broadly categorized into 2 branches:

1. Descriptive Statistics:
Descriptive statistics are specific methods basically used to calculate, describe, and summarize collected research data in a logical, meaningful, and efficient way. It is used to describe the characteristics(attributes) of a sample or population in total and limits generalization to the particular group (Sample).

2. Inferential Statistics:
Inferential statistics describe the many ways in which statistics derived from observations on samples from study populations can be used to deduce whether or not those populations are truly different.

Therefore, to summarize we can say that **descriptive statistics** focus on describing the **sample**, while **inferential statistics** aim to make predictions about the **population,** based on the findings within the sample.

Descriptive Statistics

Descriptive Statistics helps you understand the **details of your Sample**. Some common statistical measures it covers are:

1. Mean: The mean is the average value of a dataset. It's calculated by adding up all the values in the dataset and dividing by the total number of values.

2. Median: The median is the middle value when the dataset is arranged in ascending or descending order. If there's an even number of values, the median is the average of the two middle values.

3. Mode: The mode is the value that appears most frequently in a dataset. If there is only one value which occurs maximum number of times, then it is called mode. If there are 2 mode distribution then it is called bimodal and in case of more than 2 modes it is called multimodal.

4. Range: The range is the difference between the maximum and minimum values in a dataset. It gives an idea of the spread of values.

5. Variance: Variance measures how spread out the values in a dataset are from the mean. It's the average of the squared differences from the mean.

6. Standard Deviation: It's the square root of the variance and gives a measure of the amount of variation or dispersion in a dataset. A high standard deviation indicates that the values are spread out over a wider range.

7. Average Deviation: Provides the average of different variations from a data set. The purpose of average deviation is to measure the distance of a deviation from the data set's mean or median.

8. Percentiles: Percentiles indicate the value below which a given percentage of data falls. For instance, the 75th percentile represents the value below which 75% of the data falls.

8. Quartiles: Quartiles divide a dataset into four equal parts. The first quartile (Q1) is the 25th percentile, the second quartile (Q2) is the median, and the third quartile (Q3) is the 75th percentile.

9. Skewness: Skewness measures the asymmetry of the distribution of values. A positive skew indicates that the tail of the distribution is longer on the right side, while a negative skew indicates a longer tail on the left side.

10. Correlation Coefficient: It measures the strength and direction of the linear relationship between two variables. The coefficient ranges from -1 to 1, where 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship.

These statistical measures help in summarizing and interpreting quantitative data, providing insights into the central tendency, dispersion, shape of the distribution, and relationships between variables.

Inferential Statistics

Inferential statistics involves using sample data to make inferences or predictions about a larger population. There are several inferential statistical measures and techniques commonly used in the analysis of quantitative data:

1. Hypothesis Testing: This involves making decisions about a population based on sample data. Common tests include:

- **Z Test:** When data has a normal distribution and a sample size of at least 30, the **z test** is applied to the data. When the population variance is known, it determines if the sample and population means are equal.
- **T Test:** Used to compare means between two groups. When the sample size is less than 30, and the data has a student t distribution, a **t test** is utilized. The sample and population mean are compared when the population variance is unknown.
- **F Test:** When comparing the variances of two samples or populations, an **f test** is used to see if there is a difference.
- **Chi-square tests:** Used to assess the independence of categorical variables. It is applied only to discrete data. The data that are counted rather than measured.

2. Regression Analysis: It examines the relationship between one dependent variable and one or more independent variables. It can be used for prediction and to determine the strength and nature of relationships.

3. Correlation Analysis: Measures the strength and direction of the relationship between two continuous variables using correlation coefficients (e.g., Pearson's correlation coefficient).

4. Probability Distributions: Various distributions like normal, binomial, Poisson, etc., are used to model and analyze different types of data.

5. Sampling Techniques: Methods used to select a subset of individuals (sample) from a larger population. Techniques include simple random sampling, stratified sampling, cluster sampling, etc.

6. Non-parametric Tests: Used when assumptions of parametric tests are violated or data do not meet parametric assumptions. Examples include the Wilcoxon signed-rank test, Mann-Whitney U test, etc.

Each of these measures and techniques serves a different purpose in inferential statistics, allowing statisticians and researchers to draw conclusions and make predictions about populations based on sample data. The choice of method depends on the research question, the type of data collected, and the underlying assumptions.