Advanced Analytics Assignment 02

248518 Gaurav B Raskar

Q.1.

Univariate analysis: studies one variable to understand its distribution and patterns.

Bivariate analysis: examines the relationship between two variables to see how they interact.

Multivariate analysis: explores the connections between three or more variables at the same time.

Q.2.

Central Limit Theorem (CLT) states that the distribution of a sample will approximate a normal distribution (i.e., a bell curve) as the sample size becomes larger.

Q.3.

Data mining is the process of discovering patterns, trends, and useful information from large sets of data using techniques like statistics, machine learning, and database systems. Involves Data Wrangling, Data Cleaning, Data Transformation, Dimension Reduction etc.

Q.4.

Type I Error:

* Represented by α (alpha).
* Occurs when actual Null Hypothesis is False but, Measured as True.

Type II Error:

* Represented by β (beta).
* Occurs when actual Null Hypothesis is True but, Measured as False.

Q.5.

Due to Time Constraint and Cost effectiveness.

Q.6.

1) Simple Random Sampling:

Randomly select samples with given sample size.

2) Stratified Random Sampling:

Randomly selected from each Homogenous stratas/groups of population.

3) Systematic Random Sampling:

If data is extremely homogenous then, periodically selected from ordered list.

4) Cluster Sampling:

Sampling with multiple stages by selecting clusters and sub-clusters. Clusters are Heterogeneous.