STATISTICS WORKSHEET-1

1) A

2) A

3) C

4) C

5) C

6) B

7) B

8) A

9) C

10) Normal Distribution is a probability distribution that is symmetric and bell-shaped. It is characterized by its mean and standard deviation. In a normal distribution, the majority of the data points are centered around the mean, and the probabilities of values occurring decrease as they move further away from the mean. The shape of the distribution is defined by the mean and standard deviation.

11) Handling missing data depends on the nature and extent of missingness. Some common techniques for handling missing data include deletion of missing data, imputation (replacing missing values with estimated values), and using advanced methods such as multiple imputation or maximum likelihood estimation. The choice of technique depends on the specific context and assumptions made about the missing data.

12) A/B testing, also known as split testing or bucket testing, is a method used in statistics and data analysis to compare two versions of a variable (typically a web page or app feature) to determine which one performs better. It involves randomly assigning participants or users into two groups (A and B), exposing each group to a different version, and then analyzing the differences in outcomes or metrics to determine which version is more effective.

13) Mean imputation of missing data is commonly used technique, but it has limitations and potential drawbacks. It involves replacing missing values with the mean of the available data. While it can be simple to implement, it may introduce bias and underestimate the variability in the data. It assumes that the missing values have the same mean as the observed values, which may not be true in all cases. Other imputation techniques, such as multiple imputation or regression imputation, may be more appropriate depending on the data and the missingness mechanism.

14) Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and estimates the coefficients that define the slope and intercept of the best-fit line. The goal of linear regression is to find the line that minimizes the sum of squared differences

between the observed data and the predicted values.

15) Statistics can be broadly divided into several branches, including:

*Descriptive Statistics: Involves summarizing and describing data using measures such as mean, median, and standard deviation.

*Inferential Statistics: Deals with making inferences and conclusions about a population based on sample data.

*Probability Theory: Studies the likelihood of events occurring and provides a framework for quantifying uncertainty.

*Statistical Modeling: Involves building mathematical models to represent and analyze complex systems or phenomena.

*Experimental Design: Focuses on designing experiments to investigate the effects of different variables and control for confounding factors.

*Biostatistics: Applies statistical methods to analyze and interpret data in the field of biology and medicine.

*Econometrics: Applies statistical techniques to analyze economic data and test economic theories.

*Bayesian Statistics: Involves updating beliefs and making statistical inferences using Bayes' theorem and prior knowledge.

*Data Mining and Machine Learning: Utilizes statistical methods and algorithms to discover patterns and make predictions from large datasets.