**Statistic Worksheet-1**

**Answer’s of Questions**

Ans.1 option A

Ans.2 option A

Ans.3 option B

Ans.4 option D

Ans.5 option C

Ans.6 option A

Ans.7 option A

Ans.8 option A

Ans.9 option D

10. What do you understand by the term Normal Distribution

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

11. How do you handle missing data? What imputation techniques do you recommend?

Analyze each column with missing values carefully to understand the reasons behind the missing of those values, as this information is crucial to choose the strategy for handling the missing values.

There are 2 primary ways of handling missing values:

1. Deleting the Missing values
2. Imputing the Missing Values

And I will recommend imputing the missing methods first then go accordingly.

12. What is A/B testing?

A/B testing (also known as split is a methodology for comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13. Is mean imputation of missing data acceptable practice

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.

15. What are the various branches of statistics?

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.