STATISTICS WORKSHEET -1

Q1-Q9

Ans1 = A (True)

Ans2 = B

Ans3 = B

Ans4 = D

Ans5 = C

Ans6 = B(false)

Ans7 = B

Ans8 = C

Q10-Q15

Ans10 = Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the new mean. Values are equally likely to plot either above or below the mean. It is also known as Gaussian distribution or Standard normal distribution. The most common type of distribution arises in statistics through random sampling techniques.

Ans11 = Missing data reduces the statistical power of the analysis, which can distort the validity of the results. It appears when no value is available in one or more variables of an individual.

The best imputation technique is Mean or Median imputation:

When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. However, there can be multiple reasons why this may not be the most feasible option:

- There may not be enough observation with non-missing data to produce a reliable analysis.
- In predictive analytics, missing data can prevent the prediction for those observations which have missing data.
- External factors may require specific observations to be part of the analysis

Ans12 = A/B testing is basically statistical hypothesis testing, in other words, statistical inference. It is an analytical method for making decisions that estimate population parameters based on sample statistics. A process whereby a hypothesis is made about the relationship between two data sets and those data

sets are then compared against each other to determine whether there is a statistically significant relationship or not.

Ans13 = true, imputing the mean preserves the mean of the observed data. so if the data is missing completely at random, the estimates of the mean remain unbiased

Ans14 = linear regression analysis is used to predict the value of the variable based on the value of another variable. the variable you want to predict is called the dependent variable. The variable you used to predict the value of another variable is called the independent variable.

Ans15 = there are three real branches of statistics:-

1. **Data collection: there are many ways to collect data.**

2. **Descriptive statistics: descriptive statistics are used to provide data to describe the population either through numerical calculation or with the help of graphs or tables**

3. **Inferential statistics: inferential statistics is one of the two main branches of statistics. It uses a random sample of data taken from a population to describe and make inferences about the population**