STATISTICS WORKSHEET-1

1. A
2. A
3. D
4. D
5. C
6. B
7. B
8. A
9. C
10. What do you understand by the term Normal Distribution?

Ans 10: Normal distribution, also known as the Gaussian 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 mean. In graph form, normal distribution will appear as a bell curve.

* A normal distribution is the proper term for a probability bell curve.
* In a normal distribution the mean is zero and the standard deviation is 1.
* Normal distributions are symmetrical, but not all symmetrical distributions are normal.
* In reality, most pricing distributions are not perfectly normal.

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

Ans 11: When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations.

Complete Case Analysis(CCA) is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.

1. What is A/B testing?

Ans 12: A/B testing refers to the experiments where two or more variations of the same webpage are compared against each other by displaying them to real-time visitors to determine which one performs better for a given goal.

1. Is mean imputation of missing data acceptable practice?

Ans 13: Yes, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the **mean remains unbiased**. That's a good thing.

Outliers data points will have a significant impact on the mean and hence, in such cases, **it is not recommended to use the mean for replacing the missing values**. Using mean values for replacing missing values may not create a great model and hence gets ruled out.

1. What is linear regression in statistics?

Ans 14: linear regression **uses one independent variable to explain or predict the outcome of the dependent variable Y**, while multiple linear regression uses two or more independent variables to predict the outcome.

1. What are the various branches of statistics?

Ans 15 : There are four main branches of statistics.

1. Mathematical or theoretical statistics
2. Statistical methods or functions
3. Descriptive statistics
4. Inferential ststistics