STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is the correct formula for total variation?

b) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

c) binomial

3. How many outcomes are possible with Bernoulli trial?

a) 2

4. If Ho is true and we reject it is called

a) Type-I error

5. Level of significance is also called:

a) Power of the test

6. The chance of rejecting a true hypothesis decreases when sample size is:

b) Increase

7. Which of the following testing is concerned with making decisions using data?

b) Hypothesis

8. What is the purpose of multiple testing in statistical inference?

d) All of the mentioned

9. Normalized data are centred at and have units equal to standard deviations of the original data.

a) 0

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Ans: Bayes' theorem is a mathematical formula used to determine the conditional probability of events. Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event. Some of the applications include but are not limited to, modelling the risk of lending money to borrowers or forecasting the probability of the success of an investment.

11. What is z-score?

Ans: A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

12. What is t-test?

Ans: A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related. T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times. The t-test is a test used for hypothesis testing in statistics and uses the t-statistic, the t-distribution values, and the degrees of freedom to determine statistical significance.

13. What is percentile?

Ans: A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall.

14. What is ANOVA?

Ans: Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

15. How can ANOVA help?

Ans: ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.