STATISTICS WORKSHEET-3

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial
3. a) 2
4. a) Type-I error
5. b) Size of the test
6. b) Increase
7. b) Hypothesis
8. d) All of the mentioned
9. a) 0
10. Bayes theorem states that the conditional probability of an event, based on the occurance of another event, Is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.
11. Z score is also known as standard score. Z score describes the position of a score in terms of its distance from the mean, when measured with standard deviation units. The Z score is positive if the value lies above the mean, and negative if it lies below the mean.
12. T test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether two groups are different from one another.
13. Percentile is a term that describes how a score compares to other score from the same data set.
14. ANOVA is statistical method used to test differences between two or more means. Analysis of variance splits an observed data set into two parts systematic factors and random factors. Its is used to determine the influence that the independent variables have on the dependant variable in a regression study.
15. ANOVA is helpful for testing three or more variables. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.

MACHINE LEARNING

1. d. All of the above
2. d. None
3. c. Reinforcement learning and Unsupervised learning
4. b. The tree representing how close the data points are to each other
5. d. None
6. c. k-nearest neighbour is same as k-means
7. d. 1, 2 and 3
8. a. 1 only
9. a. 2
10. b.Given a database of information about your users, automatically group them into different market segments.
11. a.
12. b.
13. Clustering is important in machine language because it simplifies the processing of large datasets. It serves as feature data for downstream machine language systems.
14. Clustering performance can be improved by applying unsupervised feature learning to input data using either RICA or SFT.