WORKSHEET - 3

STATISTICS

Q1 answer : b) Total variation = Residual variation + Regression variation

Q2 answer : c)binomial

Q3 answer : a)2

Q4 answer : a)Type -1 erro

Q5 answer : b)size of the test

Q6 answer : b) Increse

Q7 answer : b)Hypothesis

Q8 answer : d) All of the mentioned

Q9 answer : a)0

Q10 answer : a theorem describing how the conditional probability of each of a set of possible causes for a given observed outcome can be computed from knowledge of the probability of each cause and the conditional probability of the outcome of each cause.

Q11 answer : The z score is used to denote the number of standard deviations by which a raw score lies above or below the mean.

Q12 answer : A T-test is a statistical method of comparing the means or proportions of two samples gathered from either the same group or different categories. It is aimed at hypothesis testing, which is used to test a hypothesis pertaining to a given population. It is the difference between population means and a hypothesized value.

Q13 answer : percentiles are the values below which a certain percentage of the data in a data set is found.

Q14 answer : 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.

Q15 answer : ANOVA can help to identify the sources of variation in a data set. This can help to improve the accuracy of data predictions and analyses. Additionally, ANOVA can help to identify relationships between different variables in a data set. This information can be used to improve data models and predictions.

WORKSHEET - 3

MACHINE LEARNING

Q1 answer : d. All of the above

Q2 answer : d. None

Q3 answer : c. Reinforcement learning and Unsupervised learning

Q4 answer : b. The tree representing how close the data points are to each other

Q5 answer : d. None

Q6 answer : c. k-nearest neighbor is same as k-means

Q7 answer : d. 1,2 and 3

Q8 answer : a.1 only

Q9 answer : a. 2

Q10 answer : b. Given a database of information about your users, automatically groups them into different market segments.

Q11 answer : a.

Q12 answer : b.

Q13 answer : clustering helps to determine the internal structure of the data. This clustering analysis has been used for model analysis, vector region of attraction. Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.

Q14 answer : Improve the quality of your clustering. First, perform a visual check that the clusters look as expected, and that examples that you consider similar do appear in the same cluster. Then check these commonly-used metrics as described in the following sections