STATISTICS WORKSHEET-3

1   Total Variation = Residual Variation + Regression Variation

2   binomia

3   2

4   Type-I error

5   size of the best

6   Increase

7   Hypothesis

8   All of the mentioned

9   0

10   It is a mathematical formula for determining conditional probability. conditional probability is the likelihood of an out come occurring based on a previous outcome having occurred in similar circumstance

$$P(A/B)=P(B/A)P(A)/P(B)$$

11   Z-Score is a numerical measurement used in statistics of value's relationship to the mean of a group value's ,measured in the term of standard deviation from the mean

12   t-test is any statistical hypothesis test in which the test statistical follows a Student's t- distribution under the null hypothesis .

13   I n statistics, A percentile is a term that describe how a score compare to other score from the same set . While there is no universal definition of percentile , It is commonly expressed as the percentage of values in a set of data score that fall below a given value.

$$Px=x(n+1)/100$$

14   A statistical method for testing whether two or more dependent variable means are equal.

15   Anova is helpful for testing three or more variables .it is similar to multiple two sample

T-tests. How ever it results in fewer type 1 error and is appropriate for a range of issues .

MACHINE LEARNING

1   All of the above

2   None

3 Reinforcement learning

4 . The tree representing how close the data points are to each other

5   None

6   k-nearest neighbour is same as k-means

7 1, 2 and 3

8 Clustering analysis is negatively affected by multicollinearity of features

9 2

10 Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

   Given a database of information about your users, automatically group them into different market segments.

11 a

12   b

13 clustering helps to determine the internal structure of the data . this clustering analysis has been used to for analysis vector region of attraction . clustering helps in understanding the natural grouping in a dataset . their purpous is to make sense to partition the data into some group of logical groupings.

14   Improve the quality of clustering . first,perform a visual check that the cluster look as expected and example that you considere similarly do appear in the same cluster .then check these commonly used described in the following sections.