Statistics Worksheet – 3

Q1) B) Total variation = Residual Variation + Regression Variation

Q2) C) Binomial

Q3) A) 2

Q4) A) Type1 Error

Q5) C) Level of Confidence

Q6) B) Increased

Q7) B) Hypothesis

Q8) D) All of Above

Q9) A) 0

Q10) Bayes Theorem is used to determine conditional probability of an event A when event B has already occurred. $P(A|B) = P(B|A)P(A) / P(B)$

Q11) Z-Score gives an idea of how far a data point is from the mean. In other words it is no of standard deviations a given point is from the mean. Z-Score lies between -3 to 3.

Q12) T-Test is one of the tests used for hypothesis testing. It is a type of inferential statistics used to determine if there is significant difference between the mean of 2 groups

Q13) Percentile are the values below which a certain percentage of data in a dataset is found. For eg 90 Percentile means 90% of data points are below this point.

Q14) Analysis of variance also termed as ANOVA is a statistical method that compares variance between population. It is of 2 types One Way and 2 Way.

Q15) Anova can be helpful in following ways:

- Allows to compare more than 2 groups simultaneously to test if the relationship exists between them.
- Determine the variability of the samples and within the samples
- Helps in improving the data accuracy of data prediction and analysis.

Machine Learning Assignment – 3

Q1) D) All of the Above

Q2) D) None

Q3) C) Reinforcement and Unsupervised

Q4) B) The tree representing how close data points are to each other

Q5) D) None

Q6) C) K nearest neighbour is same as k-means

Q7) D) 1,2,3

Q8) A) Only 1

Q9) A) 2

Q10) B)

Q11) A

Q12) B

Q13) Clustering is the most common form of unsupervised learning.

- Helps in determining the internal structure of the data.
- Helps in understanding the natural grouping in a dataset
- Used in outlier detection