STATISTICS WORKSHEET-3

Q1. Which of the following is the correct formula for total variation?

Ans: b) Total Variation = Residual Variation + Regression Variation

The complementary part of the total variation is called unexplained or residual

Q2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

Ans: c. Binomial

Q3. How many outcomes are possible with Bernoulli trial?

Ans: 2

Having only two outcomes.

Q4. If Ho is true and we reject it is called

Ans: a) Type-I error

A type I error occurs when we Reject Ho is True. In this case, we mistakenly reject a True null hypothesis. A type II Error occurs when we fail Reject Ho when, in fact, Ho is false

Q5. Level of significance is also called:

Ans: b) Size of the test

Q6.

Ans: b) Increase

Q7.

Ans: b) Hypothesis

Q8.

Ans: - d) All of the mentioned

Q9.

Ans: a) 0

Q10. What is Bayes Theorem?

Ans:

Bayes Theorem is used for conditional Probability. Conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances.

Bayes Theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of causes.

Let's know Probability and conditional Probability.

P(A|B) = P (A intersection B)/P(B)

P(A) = The probability of A occurring

P(B) = The probability of B occurring

P(A/B) = The probability of A given B

P(B/A) = The probability of B given A

P(A∩B) = The Probability of A intersection B

Formula for Bayes' Theorem

P(A|B) = P(A∩B)/P(B) =P(A).P(B|A)/P(B)

P(A|B) = Posterior

P(B|A) is Likelihood

P(A) is Prior

P(B) is Marginal

Q11. What is Z-Score?

Ans: In gaussian or standard normal distribution mean is zero and standard deviation towards right is positive 1 till 3 and on left it includes negative -1 to -3. According to empirical formula 68 percentage of the data falls under in 1st standard deviation, 2nd standard deviation has 95 percentage of the data, 3rd standard deviation has 99.7 percentage of data in standard normal distribution.

Zscore=Data Point-Mean/Standard Deviation

Zscore tells us how far data is from the mean ex: if data lies +2 then we get to know data lies 2 std deviation from mean. We use Zscore to calculate how much area that specific z-score is associated with and we can find the exact area by using zscore table.

Example:

Gate score=500,

mean score=390,

standard deviation is 45.

Find score=?

Here X=500, mean=390, std=45

Z=(data-mean)/std

Z=500-390/45

Zscore stands 2.44

Since the result is positive , we will make use of the positive table and check we get 0.99266

We multiply 0.99266*100 and get 99.26%

Q. What is t-test?

Points are as follows.

1. A t-test is a part of inferential Statistics used to determine if there is a significant difference between the two groups and how they are related by comparing the average values of two data sets.

2. T-test are used when the data sets follow a normal distribution and have unknown variances,

3. T-test is a test used for hypothesis testing in stats and uses the t-statistic, the t-distribution values and the degrees of freedom to determine statistical significance.

4. T-test can be dependent or independent

5. Student's is known as t-test, sample size less 30

 Formula: T-test= Mean-Mu/Standard Deviation(S) multiplied by sample size.

Where Standard Deviation is sum of (Mean-X) **2/n-1

Note: For one category feature we use one sample proportion test

For Two Category feature we use chi-square test

For One numerical feature (continuous variable) we use two T test

For Two numerical feature we use correlation.

Q. What is percentile?

Ans: Percentile in statistics is used to understand and interpret data, this indicates the values below which a certain percentage of the data falls in.

 Formula:

 n=(P/100) *N

where n=ordinal rank, P=percentile, N=number of values in data sets.

Ex: Score are [75, 77, 78, 78, 80, 81, 81, 82, 83, 84, 84, 84, 85, 87, 87, 88, 88, 88, 89, 90]

We need to find the 5oth percentile:

P=5o, N=20 find n

Sort the data in ascending order

n= (50/100) *20

n =10,

score is 84

50 percentage of the student lies between those marks.

Q. What is ANOVA?

Ans: This is one part for Inferential statistics, Analysis of Variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set: 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.

The Formula for ANOVA is:

F = MST/MSE Here:

- ➢ F=ANOVA coefficient
- ➢ MST=Mean sum of squares due to treatment
- ➢ MSE=Mean sum of squares due to error

Q. 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.

It is employed with subjects, test groups, between groups and within groups.

A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents