

# STATISTICS WORKSHEET-3

**Q1 to Q9 have only one correct answer. Choose the correct option to answer your question**

Q1 = b) Total Variation = Residual Variation + Regression Variation

Q2 = c) binomial

Q3 = a) 2

Q4 = a) Type-I error

Q5 = b) Size of the test

Q6 = b) Increase

Q7 = b) Hypothesis

Q8 = d) All of the mentioned

Q9 = a) 0

**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly**

Q10: What Is Bayes' Theorem?

Ans: In statistics and probability theory, the Bayes' theorem is also known as the Bayes' rule. It is used to determine the conditional probability of events. The Bayes' theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event. The theorem is named after English Statistician, Thomas Bayes, who discovered the formula in 1763. It is considered the foundation of the special statistical inference approach called the Bayes' inference.

## **Formula for Bayes' Theorem**

The Bayes' theorem is expressed in the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

- $P(A|B)$  – the probability of event A occurring, given event B has occurred
- $P(B|A)$  – the probability of event B occurring, given event A has occurred
- $P(A)$  – the probability of event A
- $P(B)$  – the probability of event B

Q11: What is z-score?

Ans: Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of Standard Deviation from the mean. If a Z-score is 0, it indicates that the data point score is identical to the mean score. Z-score also known as the Standard Score gives us an idea of how far a data point is from the mean.

The formula to find Z-score

$$z = \frac{x - \mu}{\sigma}$$

$\mu$  = Mean

$\sigma$  = Standard Deviation

X = data point

Q12: What is t-test?

Ans: t-test is also known as Student's t-test. A t-test is inferential statistics used to determine if there is a significant difference between the means of two groups and how they are related. It is used when data sets are normal distribution having unknown variances. T-test is a test for hypothesis testing in statistics.

In the t-test we have:

1. one sample t-test: one set of a data

## 2. two-sample t-test: we have 2 sample t-tests

1. independent sample
2. paired sample

Q13: What is percentile?

Ans: A Percentile is a term that describes how a score compares to other scores from the same set. The percentage of values in a set of data scores that fall below a given value. The general rule is that if a value is in the  $k^{\text{th}}$  percentile, it is greater than K percent of the total values.

$$P_x = \frac{x(n + 1)}{100}$$

$P_x$  = The value at which x percentage of data lie below that value

$n$  = Total number of observations

Q14: What is ANOVA?

Ans: ANOVA (Analysis of Variance) is an analytical tool used in statistics that splits an observed aggregate variability found inside a data set into two parts. ANOVA is used to compare differences in means among more than 2 groups. it does this by looking at variations in the data and where that variation is found (hence its name). Specifically, ANOVA compares the amount of variation between groups with the amount of variation within groups.

For example :

## We are trying to predict a company's sales based on its promotions.  
low (on radio , newspaper) , medium(billboard in public places) , high(tv adds) . promotions are categorical in nature.

## The Null Hypothesis for ANOVA says that, the average of the dependent variable are same for all the given groups.

# Alternate Hypothesis says . mean of the dependent variable are not same for the given groups

#  $H_0 = \mu_A = \mu_B = \mu_C$

#  $H_a$  = all mean or all are not equal

# step1 : Calculate the means of each school .

# step2 : Calculate the grand mean .  
# step3 : Variations between ,within .  
# step4 : Calculate the mean squared variance of between and within .  
# step5 : Calculate F statistics and corresponding p\_value .

Q15: How can ANOVA help?

Ans: ANOVA is helpful for testing three or more variables. ANOVA groups difference by comparing the means of each group and includes spreading out the variance into diverse sources. ANOVA can help to find the differences between the means of Independent variables. When it is understood that how each variables' mean is different from the others, then we can begin to understand which of the feature has a connection with the dependent variable