STATISTICS WORKSHEET-3

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

1. Which of the following is the correct formula for total variation?
   Answer: Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called_____outcomes.
   Answer: binomial

3. How many outcomes are possible with Bernoulli trial?
   Answer: 2

4. If Ho is true and we reject it is called
   Answer: Type-I error

5. Level of significance is also called:
   Answer: Power of the test

6. The chance of rejecting a true hypothesis decreases when sample size is:
   Answer: Increase

7. Which of the following testing is concerned with making decisions using data?
   Answer: None of the mentioned
   Statistical hypothesis is used

8. What is the purpose of multiple testing in statistical inference?
   Answer: Minimize false positives

9. Normalized data are centred at____and have units equal to standard deviations of the original data
   Answer: 0

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Answer: Bayes' theorem is a mathematical formula that describes the relationship between the probability of an event occurring (the prior probability) and the probability of that event occurring given some new information (the posterior probability).

The formula for Bayes' theorem is as follows:

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

Where:

P(A|B) is the posterior probability of event A occurring given that event B has occurred.

P(B|A) is the likelihood of event B occurring given that event A has occurred.

P(A) is the prior probability of event A occurring.

P(B) is the prior probability of event B occurring.

11. What is z-score?

Answer: A z-score, also known as a standard score, is a measure of how many standard deviations an observation or data point is from the mean of a distribution. It is a dimensionless value that indicates the location of a data point relative to the mean and standard deviation of a distribution.

The formula for calculating a z-score is:

$z = (x - \mu) / \sigma$

Where:

x is the data point or observation
μ is the mean of the distribution
σ is the standard deviation of the distribution

A positive z-score indicates that the data point is above the mean, while a negative z-score indicates that the data point is below the mean. The farther the z-score is from zero, the more extreme the data point is relative to the mean.

Z-scores are useful for comparing data points from different distributions, normalizing data, and identifying outliers in a dataset. They are also used in statistical tests such as the z-test and the t-test, which test hypotheses about population means based on samples.

12. What is t-test?

Answer: A t-test is a statistical test that compares the mean of a sample to a known or hypothesized value, typically the population mean. It is used to determine whether there is a significant difference between the mean of a sample and a known or hypothesized value. There are several types of t-tests including:

One-Sample t-test: This test compares the mean of a sample to a known population mean.

Independent Sample t-test: This test compares the means of two independent samples.

Paired Sample t-test: This test compares the means of two related samples, such as before and after measurements.

One-Way ANOVA: This test compares the means of more than two independent samples.

13. What is percentile?

Answer: A percentile is a measure of the relative standing of a value within a dataset. It represents the value below which a certain percentage of observations in a dataset fall. For example, if a data point has a percentile rank of 90, it means that it is greater than or equal to 90% of the other data points in the dataset. Percentiles are often used to understand and describe the distribution of a dataset.

The formula to calculate the percentile of a data point x in a dataset with n observations is:

$Percentile = (P / 100) * n$

Where P is the percentile rank (e.g. 90 for the 90th percentile) and n is the number of observations in the dataset.

For example, if we have a dataset of 100 observations and we want to find the 90th percentile, we would take $(90 / 100) * 100 = 90$. This means that the 90th percentile is the value that is greater than or equal to 90% of the other observations in the dataset.

Percentile can also be calculated by using Percentile formula which is

Percentile = (P / 100) * (n + 1)

14. What is ANOVA?

Answer: ANOVA (Analysis of Variance) is a statistical technique that is used to test whether there are significant differences between the means of two or more groups. It is used to determine whether the variation among groups is greater than the variation within groups. ANOVA is a generalization of the t-test, which is used to compare the means of two groups.

There are several types of ANOVA, including:

One-Way ANOVA: This test compares the means of two or more independent groups.

Two-Way ANOVA: This test compares the means of two or more groups that have been split into subgroups based on two factors.

Repeated Measures ANOVA: This test compares the means of two or more groups that have been measured multiple times.

Mixed ANOVA: This test is a combination of both one-way and two-way ANOVA.

15. How can ANOVA help?

Answer: ANOVA can help in several ways:

Identifying differences between means: ANOVA can be used to determine whether there are significant differences between the means of two or more groups. It can help you to identify which groups are different from each other and which ones are not.

Understanding the sources of variation: ANOVA can help you to understand the sources of variation in your data, such as the variation among groups and the variation within groups. This can provide insight into the underlying processes that are driving your data.

Comparing multiple groups: ANOVA allows you to compare multiple groups simultaneously, rather than comparing each group one at a time. This can be more efficient and less prone to errors than multiple t-tests.

Identifying interactions: Two-way ANOVA can be used to identify interactions between two factors, which can reveal complex relationships between variables.

Making inferences about populations: ANOVA can be used to make inferences about population means based on samples. It can help you to estimate the population means and to make predictions about future observations.

Making decisions: ANOVA can be used to support decision-making by providing statistical evidence for or against certain hypotheses. It can help you to identify which groups are different and can inform the design of experiments and the selection of treatments.

Overall, ANOVA is a powerful statistical technique that can provide valuable insights into the relationships between variables and help you to make informed decisions based on your data.