FLIP ROBO

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?
    a) Total Variation = Residual Variation – Regression Variation
    b) Total Variation = Residual Variation + Regression Variation
    c) Total Variation = Residual Variation * Regression Variation
    d) All of the mentioned
       Answere:- (B)

2. Collection of exchangeable binary outcomes for the same covariate data are called_____outcomes.
    a) random
    b) direct
    c) binomial
    d) none of the mentioned
       Answere:- (C)

3. How many outcomes are possible with Bernoulli trial?
    a) 2
    b) 3
    c) 4
    d) None of the mentioned
       Answere:- (A)

4. If Ho is true and we reject it is called
    a) Type-I error
    b) Type-II error
    c) Standard error
    d) Sampling error
       Answere:- (A)

5. Level of significance is also called:
    a) Power of the test
    b) Size of the test
    c) Level of confidence
    d) Confidence coefficient
       Answere:- (B)

6. The chance of rejecting a true hypothesis decreases when sample size is:
    a) Decrease
    b) Increase
    c) Both of them
    d) None
       Answere:- (B)

7. Which of the following testing is concerned with making decisions using data?
    a) Probability
    b) Hypothesis
    c) Causal
    d) None of the mentioned
       Answere:- (B)

8. What is the purpose of multiple testing in statistical inference?
   a) Minimize errors
   b) Minimize false positives
   c) Minimize false negatives
   d) All of the mentioned
      Answere:- (D)

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

   a) 0
   b) 5
   c) 1
   d) 10
      Answere:- (A)

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

10. What Is Bayes' Theorem?
    Bayes' theorem is a mathematical formula that describes how the probability of an event (A) changes based on new information (B). It is named after the 18th-century mathematician and statistician Thomas Bayes. The theorem is represented mathematically as:
    $P(A|B) = P(B|A) * P(A) / P(B)$
    $P(A|B)$ is the conditional probability of event A occurring given that event B has occurred (i.e. the probability of A given B)
    $P(B|A)$ is the conditional probability of event B occurring given that event A has occurred (i.e. the probability of B given A)
    $P(A)$ is the prior probability of event A occurring (i.e. the probability of A before any new information is taken into account)
    $P(B)$ is the prior probability of event B occurring (i.e. the probability of B before any new information is taken into account)

11. What is z-score?
    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 defined as:
       $z\text{-score} = (x - \mu) / \sigma$
    x is the value of the observation or data point
    μ is the mean of the distribution
    σ is the standard deviation of the distribution

12. What is t-test?
    A t-test is a statistical procedure used to determine whether there is a significant difference between the means of two groups of data. There are several types of t-tests, but the most common are the Student's t-test and the independent samples t-test.

    The Student's t-test is used to compare the means of two groups of data when the sample size is small (typically less than 30) or when the population standard deviation is unknown. It compares the difference between the sample mean and the hypothesized population mean to the standard error of the mean.
    The independent samples t-test is used to compare the means of two groups of data when the samples are independent and the variances of the two groups are equal. It is also known as two-sample t-test or unpaired t-test.

13. What is percentile?
    A percentile is a measure of the relative position of a value within a set of data. It indicates the percentage of values in a data set that are less than or equal to a given value.
    For example, if a value is at the 50th percentile, it means that 50% of the values in the data set are less than or equal to that value. Similarly, if a value is at the 75th percentile, it means that 75% of the values in the data set are less than or equal to that value.

14.What is ANOVA?

ANOVA stands for Analysis of Variance, it is a statistical method used to compare the means of two or more groups of data. It is used to determine whether there is a significant difference in the means of the groups being compared.

There are different types of ANOVA depending on the number of groups and the number of factors being studied. The most common types of ANOVA are:
One-way ANOVA: This is used when there is one independent variable with two or more levels and one dependent variable. It is used to compare the means of two or more groups.
Two-way ANOVA: This is used when there are two independent variables with two or more levels each and one dependent variable. It is used to compare the means of two or more groups and to determine the interaction effect of the two independent variables.
N-way ANOVA: This is used when there are more than two independent variables with two or more levels each and one dependent variable.

15.How can ANOVA help?
ANOVA can help in several ways:

Identifying significant differences in means: ANOVA can be used to determine whether there is a significant difference in the means of two or more groups of data. This can be useful in many fields such as psychology, economics, biology, and engineering to make inferences about population means based on sample data.
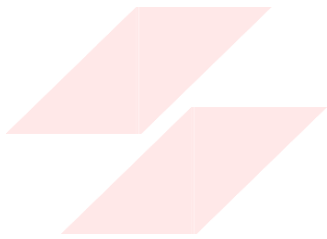
Comparing multiple groups: ANOVA allows for the comparison of means among multiple groups, which can be useful in identifying patterns and trends in data.
Investigating the effect of multiple independent variables: When there are multiple independent variables with two or more levels each, ANOVA allows for the investigation of the effect of each variable on the dependent variable, as well as the interaction effect among different independent variables on the dependent variable.
Identifying outliers: ANOVA can be used to identify outliers in a dataset and compare the means of different groups. This can be useful in identifying observations that may be skewing the data and in understanding the behavior of the data.
Identifying the source of variation: ANOVA can be used to identify the source of variation among different groups of data. This can be useful in identifying which factors are contributing most to the differences in means between groups.

10. What Is Bayes' Theorem?
11. What is z-score?
12. What is t-test?
13. What is percentile?
14. What is ANOVA?
15. How can ANOVA help?