Assignment 3

Statistics

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

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

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

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

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

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

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

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

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

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

10. What Is Bayes' Theorem?
Answer:
   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.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B = events
$P(A|B)$ = probability of A given B is true
$P(B|A)$ = probability of B given A is true
$P(A), P(B)$ = the independent probabilities of A and B

11. What is z-score?
Answer:
   A measure of how many standard deviations below or above the population mean a raw score is called z score. It will be positive if the value lies above the mean and negative if it lies below the mean. It is also known as standard score.

$$Z = \frac{\bar{x} - \mu}{\sigma}$$
$\bar{x} = Sample\ mean$
$\mu = Population\ mean$
$\sigma = Standard\ deviation$

12. What is t-test?
Answer:
   A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
$\bar{x} = Sample\ mean$
$\mu = Population\ mean$
$\sigma = Standard\ deviation$
$n = Sample\ size$

13. What is percentile?
Answer:

A percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed.

Example - If you score 75 points on a test, and are ranked in the 85th percentile, it means that the score 75 is higher than 85% of the scores.

14. What is ANOVA?
Answer:

Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

15. How can ANOVA help?
Answer:

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 provides the overall test of equality of group means. It can control the overall type I error rate (i.e. false positive finding) It is a parametric test so it is more powerful, if normality assumptions hold true.