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

Answer: B) Total Variation = Residual Variation +RegressionVariation

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

Answer: C) Binomial

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

Answer: A) 2

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

Answer: A) Type-I 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

Answer: A) Power of the Test

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

Answer: B) Increase

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

Answer: B) Hypothesis

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

Answer: 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

Answer:A) 0

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

10. What Is Bayes' Theorem?

Answer: The Bayes' theorem, commonly referred to as the Bayes' rule, is a mathematical formula used to calculate the conditional probability of events in statistics and probability theory. The Bayes theorem basically describes the likelihood of an event based on knowledge of potential relevant conditions in advance.

The following formula represents the Bayes' theorem:

Formula for Bayes' Theorem

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

Where:

P(A|B) is the likelihood that event A will occur after event B has already occurred.

P(B|A) is the likelihood that event B will occur given the occurrence of event A.

P(A) represents the likelihood of event A.

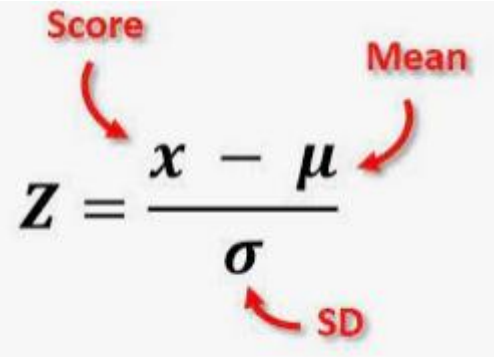
P(B) represents the likelihood of event B.

11. What is z-score?

Answer: A raw score's position relative to its standard deviation—or distance—from the mean is described by the z-score. If a value is above the mean, the z-score is positive; if it is below the mean, it is negative.

Because it standardises the distribution, it enables comparison of scores on many types of variables, hence the name "standard score." A normally shaped distribution with a mean of 0 and a standard deviation (SD) of 1 is known as a standard normal distribution (SND).

Formula:



- Where:
  Z=standard score
  x=observed value
  μ =mean of the sample
  σ =standard deviation of the sample

12. What is t-test?

Answer: The t-test is an inferential statistic used to assess the relationship between two groups' means and assess whether there is a significant difference between them. When data sets contain unknown variances and a normal distribution, such as the data set obtained from tossing a coin 100 times, t-tests are utilised.

In order to evaluate statistical significance, the t-test, a test used for hypothesis testing in statistics, uses the t-statistic, the values of the t-distribution, and the degrees of freedom.

Formula

$t = (\bar{x} – \mu) / (s / \sqrt{n})$

where:

$\bar{x}$ = Observed Mean of the Sample

μ = Theoretical Mean of the Population

s = Standard Deviation of the Sample

n = Sample Size

13. What is percentile?

Answer: A percentile is a score that compares a specific score to the scores of the remainder of the group. It displays the proportion of other scores that a given score outperformed.

Explanation with the Example:

Although the term "percentile" is frequently used, it lacks a common definition. The most typical definition of a percentile is a number below which a predetermined proportion of scores fall. You may be aware that you received a test score of 70 out of 90. But unless you know which percentile you fall into, that number has no real meaning. If you know that you scored in the 85th percentile, it signifies that you outperformed 85% of test-takers.

14. What is ANOVA?

Answer: An ANOVA test is a sort of statistical analysis that checks for variance-based mean differences to see if there is a statistically significant difference between two or more category groups.

The independent variable is divided into two or more groups by ANOVA, which is another important component. For instance, one or more groups might be predicted to have an impact on the dependent variable, whereas another group might be employed as a control group and not be predicted to have an impact.

Conditions for ANOVA

1. Only when there is no relationship between the subjects in each sample can an ANOVA be performed. In other words, in independent samples or between-groups studies, subjects from the first group cannot also be found in the second group.

2. Equivalent sample sizes must be used for all groups and levels.

3. The dependent variable must be regularly distributed in order for an ANOVA to be performed, meaning that the middle scores are most common and the extreme scores are least often.

4. Variances in the population must be equal (i.e. homoscedastic). The term "homogeneity of variance" refers to the similarity of the standard deviation or range of scores across populations.

15. How can ANOVA help?

Answer: When examining three or more variables, an ANOVA is useful. It resembles numerous two-sample t-tests. But it produces fewer type I errors and is suitable for a variety of problems. ANOVA includes dispersing the variation among many sources and groups differences by comparing the means of each group. It is used with test groups, subjects, as well as between and within groups.

The Formula for ANOVA is:

$$F = \frac{MST}{MSE}$$

where:
F=ANOVA coefficient
MST=Mean sum of squares due to treatment
MSE=Mean sum of squares due to error