

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

Ans - Total Variation = Residual Variation + Regression Variation(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

Ans – binomial(c)

3. How many outcomes are possible with Bernoulli trial?

- a) 2**
- b) 3
- c) 4
- d) None of the mentioned

Ans – 2(a)

4. If H_0 is true and we reject it is called

- a) Type-I error**
- b) Type-II error
- c) Standard error
- d) Sampling error

Ans - Type-I error (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

Ans - Size of the test (b)

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

- a) Decrease
- b) ncrease**
- c) Both of them
- d) None

Ans – Increase (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

Ans – Hypothesis (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

Ans - All of the mentioned (d)

9. Normalized data are centered at _ and have units equal to standard deviations of the original data

- a) 0
- b) 5
- c) 1
- d) 10

Ans - 0(a)

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

- 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?

10. Ans -

The Bayes theorem is a mathematical method used in probability theory and statistics to calculate the conditional probability of events. The Bayes theorem basically describes the likelihood of an event based on knowledge of potential relevant conditions in advance. A conditional probability might result in more precise conclusions since it takes into account more conditions, or more data. As a result, conditional probabilities are essential for computing accurate probabilities and predictions in machine learning. It is critical to comprehend how algorithms and techniques like the Bayes Theorem are used in machine learning given that the subject is becoming more and more prevalent across a range of domains. Let E_1, E_2, \dots, E_n be a set of events associated with a sample space S , where all the events E_1, E_2, \dots, E_n have nonzero probability of occurrence and they form a partition of S . Let A be any event associated with S , then according to Bayes theorem,

$$P(E_i | A) = \frac{P(E_i)P(A|E_i)}{\sum_{k=1}^n P(E_k)P(A|E_k)}, \text{ for any } k = 1, 2, 3, \dots, n$$

11. Ans -

The amount of standard deviations a given data point is above or below the mean is represented by the Z-score, also known as the standard score. The mean is calculated by taking the average of all group values, adding them all up, and then dividing the result by the total number of group members. Subtract the mean from each of the individual data points and divide the result by the standard deviation to obtain the Z-score. Results of zero demonstrate that the point and mean are equal. When data points fall below the mean, the Z-score is negative, while a result of one indicates that the point is one standard deviation above the mean. In the majority of large data sets, 99% of the values have a Z-score between -3 and 3, which denotes that they are within three standard deviations above or below the mean.

12. Ans –

A statistical test called a t test is employed to compare the means of two groups. Because all test findings are based on t-values, t-tests are so named. During a hypothesis test, a test statistic is a standardized value that is computed from a sample of data. The data are compared to what would be anticipated if the null hypothesis were true throughout the process of calculating the test statistic. It is frequently employed in hypothesis testing to establish whether a procedure or treatment truly affects the population of interest or whether two groups differ from one another.

13. Ans –

A percentile in statistics refers to a score's position in relation to other scores from the same set. Although percentile has no one fixed meaning, it is frequently described as the proportion of values in a set of data scores that are lower than a particular value. Percentiles demonstrate how one value stands up to others. The percentile formula for a score 'x' can be defined as number of scores that fall under 'x' divided by total number of values in the given population.

The formula of percentile is:

$$\text{Percentile}(x) = (\text{Number of values fall under 'x'}/\text{total number of values}) \times 100$$

$$P = (n/N) \times 100$$

Where,

P – Percentile

n – Number of values below 'x'

N – Total count of population

14. Ans –

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. ANOVA tests come in a variety of varieties. A "One-Way" and a "Two-Way" are the two most popular. The Formula for ANOVA is:

$$F = \text{MST} / \text{MSE}$$

Where,

F=ANOVA coefficient

MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error

15. Ans –

When examining three or more variables, an ANOVA is useful. It is comparable to other 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.

We can determine whether or not there are appreciable variations in the means of the independent variables using the one-way ANOVA. We can start to identify which independent variable is related to the dependent variable and get a better understanding of what is causing that behavior when we recognize how each independent variable's mean differs from the others.