

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. Option (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. Option (c)

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

Ans. Option (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. Option (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. Option (b)

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

Ans. Option (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. Option (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. Option (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

Ans. Option (a)

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

10. What Is Bayes' Theorem?

Ans: 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.

Formula for Bayes' Theorem

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

where:

$P(A)$ = The probability of A occurring

$P(B)$ = The probability of B occurring

$P(A|B)$ = The probability of A given B

$P(B|A)$ = The probability of B given A

$P(A \cap B)$ = The probability of both A

Examples of Bayes' Theorem:

Below are two examples of Bayes' Theorem in which the first example shows how the formula can be derived in a stock investing example using Amazon.com Inc. (AMZN). The second example applies Bayes' theorem to pharmaceutical drug testing.

11. What is z-score?

Ans: A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

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

12. What is t-test?

Ans: The t-test is a test that is mainly used to compare the mean of two groups of samples. It is meant for evaluating whether the means of the two sets of data are statistically significantly different from each other. There are many types of t-test.

There are many types of t-test, some of these are:

- The one-sample t-test, which is used to compare the mean of a population with a theoretical value.
- The unpaired two-sample t-test, which is used to compare the mean of two independent given samples.
- The paired t-test, which is used to compare the means between two groups of samples that are related.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\Delta}}$$

where

$$s_{\Delta} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where,

\bar{x}

= Mean of first set of values

\bar{x}_2

= Mean of second set of values

S_1

= Standard deviation of first set of values

S_2

= Standard deviation of second set of values

n_1

= Total number of values in first set

n_2

= Total number of values in second set.

13. What is percentile?

Ans: In statistics, a percentile is a term that describes how a score compares to other scores from the same set. While there is no universal definition of percentile, it is commonly expressed as the percentage of values in a set of data scores that fall below a given value.

$$P_x = \frac{x(n + 1)}{100}$$

P_x = The value at which x percentage of data lie below that value

n = Total number of observations

14. What is ANOVA?

Ans: 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.

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

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

15. How can ANOVA help?

Ans: ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer types 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.

The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as the first example: age, sex, income). When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behaviour.

ANOVA tells you if the dependent variable changes according to the level of the independent variable. For example:

- Your independent variable is social media use, and you assign groups to low, medium, and high levels of social media use to find out if there is a difference in hours of sleep per night.
- Your independent variable is brand of soda, and you collect data on Coke, Pepsi, Sprite, and Fanta to find out if there is a difference in the price per 100ml.
- Your independent variable is type of fertilizer, and you treat crop fields with mixtures 1, 2 and 3 to find out if there is a difference in crop yield.