Statistics Worksheet

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 + Regression Variation

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: 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

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?

Bayes Theorem is basic framework for critical thinking and gives information about the probability of different outcomes from each possible cause. This technique is widely used in the area of pattern recognition. Let us describe the setting for a classification problem and then briefly outline the procedure.

11. What is z-score?

A Z-score/standard score, represents the number of standard deviations, a data point is away from the mean of the group. It is one of the methods of standardising values.

$$Z = \frac{x - \mu}{\sigma}$$

12. What is t-test?

It is statistical test used to compare two features i.e., t test tells you how significant the differences between group means are. When having less than 30 records and S.D. is unknown.

$$t = \frac{\bar{x} - \mu}{\sqrt{n}}$$

13. What is percentile?

Group/Feature divided into 100 parts. First, Second and third part is 25^{th}, 50^{th} and 75^{th} percent value. Stands for first, second and third quartile.

14. What is ANOVA?

It is statistical method used to calculate, difference between two or more features. This test results, if there are any statistical differences between the means of three or more independent groups.

15. How can ANOVA help?

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.