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 + 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 the 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-1 error
b) Type-2 error
c) Standard error
d) Sampling error

Answer: a) Type-1 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 the sample size is:
   a) Decrease
   b) Increase
   c) Both of them
   d) None

Answer: a) 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: c) 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 centered 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:

In probability theory and statistics **Bayes' theorem**, describes the probability of an event based on prior knowledge of conditions that might be related to the event.
For example, if the risk of developing health problems is known to increase with age, Bayes' theorem allows the risk to an individual of a known age to be assessed more accurately (by conditioning it on their age) than simply assuming that the individual is typical of the population as a whole.

11. What is a z-score?

Answer:

In statistics, the **standard score** is the number of standard deviations by which the value of a raw score (i.e., an observed value or data point) is above or below the mean value of what is being observed or measured. Raw scores above the mean have positive standard scores, while those below the mean have negative standard scores.
Standard scores are most commonly called ***z*-scores**; the two terms may be used interchangeably, as they are in this article. Other equivalent terms in use include **z-values, normal scores**, **standardized variables**
Computing a z-score requires knowledge of the mean and standard deviation of the complete population to which a data point belongs

12. What is a t-test?

Answer:

The ***t*-test** is any statistical hypothesis test in which the test statistic follows a Student's *t*-distribution under the null hypothesis.

A *t*-test is the most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. The scaling term is a Nuisance parameter. When the scaling term is estimated based on the data, and the variances of groups compared are equal, then this test statistic (under certain conditions) follows a Student's *t* distribution.

The *t*-test can be used, for example, to determine if the means of two sets of data are significantly different from each other.

13. What is a percentile?

Answer:

In statistics, a **k-th percentile** (**percentile score** or **centile**) is a score below which a given percentage k of scores in its frequency distribution falls (exclusive definition) or a score at or below which a given percentage falls (inclusive definition).
For example, the 50th percentile (the median) is the score below which (exclusive) or at or below which (inclusive) 50% of the scores in the distribution may be found.
Percentiles are expressed in the same unit of measurement as the input scores.
The 25th percentile is also known as the first quartile (Q_1), the 50th percentile as the median or second quartile (Q_2), and the 75th percentile as the third quartile (Q_3).

14. What is ANOVA?

Answer:

Analysis of variance (**ANOVA**) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyse the differences among means.

ANOVA is based on the law of total variance, where the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

In its simplest form, ANOVA provides a statistical test of whether two or more population means are equal, and therefore generalizes the *t*-test beyond two means.

In other words, the ANOVA is used to test the difference between two or more means.

15. How can ANOVA help?

Answer:

An ANOVA test allows an analyst to compare two or more groups at the same time to establish if there is any relationship between the groups. The result of the test, F-ratio/statistics, allows an analysis of multiple groups of data to determine the variability within samples as well as between samples.

Example

The One-way ANOVA can help you examine if there is any difference between the means of your independent variables. You can understand how the mean differs from one independent variable to another. As a result, you can understand which independent variable has a relationship with your dependent variable and what is motivating that behaviour.