STATISTICS WORKSHEET-3

Answers:

1. Which of the following is the correct formula for total variation?
b) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
c) binomial

3. How many outcomes are possible with Bernoulli trial?
a) 2

4. If Ho is true and we reject it is called
a) Type-I error

5. Level of significance is also called:
b) Size of the test

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

7. Which of the following testing is concerned with making decisions using data?
b) Hypothesis

8. What is the purpose of multiple testing in statistical inference?
d) All of the mentioned (Minimize errors, Minimize false positives, Minimize false negatives)

9. Normalized data are centred at and have units equal to standard deviations of the original data
a) 0

10. What Is Bayes' Theorem?
Answer: 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: $P(A|B) = P(A \cap B) / P(B) = P(A) \cdot P(B|A) / P(B)$
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 and B occurring

It provides a way to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis and the observed data itself

11. What is z-score?
A measure of how many standard deviations below or above the population mean a raw score is called z score.

- Positive if the value lies above the mean
- Negative if it lies below the mean
- 0 z-score indicates that the data point's score is identical to the mean score.
- Z-score of 1.0 indicates a value that is one standard deviation from the mean.
- Z-score of 2.0 indicates a value that is two standard deviations from the mean.

Z-score is measured in terms of standard deviations from the mean.

12. What is t-test?

A t test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

- It is one of the inferential statistic test used as hypothesis testing tool
- The t-value measures the size of the difference relative to the variation in your sample data.
- T is simply the calculated difference represented in units of standard error.
- The greater the magnitude of T, the greater the evidence against the null hypothesis.

13. What is percentile?

A percentile is a comparison measure in statistics between a particular score and the scores of the rest of a group

Or A percentile is a comparison measure in statistics between a rank of a particular observation and rank of the rest of observations of group

For example, if your score on a test is on the 95th percentile only 5% of the scores were higher than yours

14. What is ANOVA?

Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups.

Tells you if there are any statistical differences between the means of three or more independent groups.

Formula: F = MSB/MSE

F - ANOVA coefficient

MSB - Mean sum of squares between the groups

MSE - Mean squares of errors

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.

The one-way ANOVA helps to know whether or not there are significant differences between the means of your independent variables

To understand which of them has a strong connection to your dependent variable and understand what is driving that behaviour.