WORKSHEET

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?
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?
a) Minimize errors
b) Minimize false positives
c) Minimize false negatives
d) All of the mentioned

WORKSHEET

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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Bayes' Theorem is a mathematical formula used to determine the conditional probability of events. Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.

$$P(A|B) = P(B|A)P(A)/P(B)$$

- P(AIB) – the probability of event A occurring, given event B has occurred

- P(BIA) – the probability of event B occurring, given event A has occurred
- P(A) – the probability of event A
- P(B) – the probability of event B

11. What is z-score?
A z-score (also called a standard score) gives you an idea of how far from the mean a data point is. It's a measure of how many standard deviations below or above the population mean a raw score is.

z = (x – μ) / σ

μ - mean

σ - standard deviation

12. What is t-test?

A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related. T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s^2(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

13. What is percentile?
A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall.

14. What is ANOVA?
An Analysis of Variance(ANOVA) test is a type of statistical test used to determine if there is a statistically significant difference between two or more categorical groups by testing for differences of means using variance.

There are different types of ANOVA tests. The two most common are a "One-Way" and a "Two-Way."

One-Way ANOVA compares the means of two or more independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different. One-Way ANOVA is a parametric test.

The One-Way ANOVA is often used to analyze data from the following types of studies:

- Field studies
- Experiments

- Quasi-experiments

The One-Way ANOVA is commonly used to test the following:

- Statistical differences among the means of two or more groups
- Statistical differences among the means of two or more interventions
- Statistical differences among the means of two or more change scores

A two-way ANOVA tests the effect of two independent variables on a dependent variable.A two-way ANOVA test analyzes the effect of the independent variables on the expected outcome along with their relationship to the outcome itself. Random factors would be considered to have no statistical influence on a data set, while systematic factors would be considered to have statistical significance.

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. It is employed with subjects, test groups, between groups and within groups.