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

2. Collection of exchangeable binary outcomes for the same covariate data is called_____outcomes.
   a) random
   b) direct
   c) binomial
   d) none of the mentioned

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

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

5. Level of significance is also called:
   a) Power of the test
   b) Size of the test
   c) Level of confidence
   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

7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned

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

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

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

10. What Is Bayes' Theorem?

Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of "causes".

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

Bayes Theorem Statement

Let E_1, E_2,…, E_n be a set of events associated with a sample space S, where all the events E_1, E_2,…, E_n have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,

$$P(Ei|A) = P(Ei)P(A|Ei)\sum_{k=1}^{n}P(Ek)P(A|Ek)$$

for any k = 1, 2, 3, …., n

11. What is z-score?

A z score is simply defined as the number of standard deviation from the mean. The z-score can be calculated by subtracting mean by test value and dividing it by standard value. Where x is the test value, μ is the mean and σ is the standard value.

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

Where x is the test value, μ is the mean and σ is the standard value.

12. What is t-test?

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.

- 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.

13. What is percentile?

A percentile (or a centile) is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall.
The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests.

14. What is ANOVA?

ANOVA stands for Analysis of variance. Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests.

- 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.
- If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

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.
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 behavior.