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: Option 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: Option C- binomial

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

Answer: Option A - 2

FLIP ROBO

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: Option 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: Option A – Power of the test

6. The chance of rejecting a true hypothesis decreases when sample size is:
   a) Decrease
   b) Increase
   c) Both of them
   d) None

Answer: Option 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: Option 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: Option 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: Option A - 0

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

10. What Is Bayes' Theorem?

Answer: In statistics and probability theory, the Bayes' theorem (also known as the Bayes' rule) 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.

The theorem is named after English statistician, Thomas Bayes, who discovered the formula in 1763. It is considered the foundation of the special statistical inference approach called the Bayes' inference.

Besides statistics, the Bayes' theorem is also used in various disciplines, with medicine and pharmacology as the most notable examples. In addition, the theorem is commonly employed in different fields of finance. Some of the applications include but are not limited to, modeling the risk of lending money to borrowers or forecasting the probability of the success of an investment.

11. What is z-score?

Answer: A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The z-score is positive if the value lies above the mean and negative if it lies below the mean. It is also known as a standard score, because it allows comparison of scores on different kinds of variables by standardizing the distribution. It allows researchers to calculate the probability of a score occurring within a standard normal distribution. Enables us to compare two scores that are from different samples (which may have different means and standard deviations). The formula for calculating a z-score is is $z = (x-u)/\sigma$, where x is the raw score, μ is the population mean, and σ is the population standard deviation.

12. What is t-test?

Answer: 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. A test can only be used when comparing the means of two groups. A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. The t-test is one of many tests used for the purpose of hypothesis testing in statistics. Calculating a t-test requires three key data values. They include the difference between the mean values from each data set (called the mean difference), the standard deviation of each group, and the number of data values of each group.

13. What is percentile?

Answer: In statistics, percentiles are used to understand and interpret data. The nth percentile of a set of data is the value at which n percent of the data is below it. In everyday life, percentiles are used to understand values such as test scores, health indicators, and other measurements.

For example, an 18-year-old male who is six and a half feet tall is in the 99th percentile for his height. This means that of all the 18-year-old males, 99 percent have a height that is equal to or less than six and a half feet. An 18-year-old male who is only five and a half feet tall, on the other hand, is in the 16th percentile for his height, meaning only 16 percent of males his age are the same height or shorter.

Percentiles are used to understand and interpret data. They indicate the values below which a certain percentage of the data in a data set is found.

Percentiles can be calculated using the formula n = (P/100) x N, where P = percentile, N = number of values in a data set (sorted from smallest to largest), and n = ordinal rank of a given value.

Percentiles are frequently used to understand test scores and biometric measurements.

14. What is ANOVA?

Answer: 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. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

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?

Answer: An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help to figure out if we need to reject the null hypothesis or accept the alternate hypothesis.
Basically, our testing groups to see if there’s a difference between them.