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

Ans = 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

Ans = c) binomial

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

Ans = a) 2

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

Ans = b) Type-II 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

Ans = b) Size 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

Ans = 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

Ans = 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

Ans = 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

Ans = a) 0

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

10. What Is Bayes' Theorem?

Ans = Bayes Theorem governs the likelihood that one event is based on the occurrence of some other events. It depends upon the concepts of conditional probability. This theorem gives us the probability of some events depending on some conditions related to the event. We know that the likelihood of heart disease increases with increasing age. So, if we know someone's age and know the chances of getting heart disease, we can easily find the opportunity of the person having a heart disease. Let us take another example. We have been given a bag of green, white and red balls. It's been said that we have to pick a green ball only after picking a blue ball. It's been sure that we can only pick green balls after picking a blue ball. This makes the case of conditional probability.

P(A) = Probability of occurrence of event A

P(B) = Probability of occurrence of event B

P(A|B) = Probability of occurrence of event A given B

P(B|A) = Probability of occurrence of event B given A

11. What is z-score?

Ans = A standard Normal Table (also called the unit normal table or z-score table is a mathematical table for the values of indicating the values of the cumulative distribution function of the normal distribution Z-score also known as the standard score, indicates how many standard deviations an entity is, from the mean. Since probability tables cannot be printed for every normal distribution, as there is an infinite variety of normal distribution, it is common practice to convert a normal to a standard normal and then use the z-score table to find probabilities.

It is a way to compare the results from a test to a “normal” population.

- if X is a random variable from a normal distribution with mean and standard deviation its Z-score may be calculated by subtracting mean from X and dividing the whole by standard deviation. A z-score of less than 0 represents an element less than the mean. A z-score greater than 0 represents an element greater than the mean. A z-score equal to 0 represents an element equal to the mean. A z-score equal to 1 represents an element, which is 1 standard deviation greater than the mean; a z-score equal to 2 signifies 2 standard deviations greater than the mean; etc.

12. What is t-test?

Ans = The t-test is any statistical hypothesis test in which the test statistic follows a Student’s t-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other, and is 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. In the statistics t-test is a very important hypothesis test. In this test, statistician follows a student t-distribution. It is done under the null hypothesis. It can be used to determine whether two sets of data are significantly different from each other or not. And also it is most commonly applied when the test statistic would follow a normal distribution. In this article, we will discuss the t-test formula with an example. Let us learn the concept this is called t-test

13. What is percentile?

Ans = The percentile score and the percentile rank are related terms. The percentile rank of a score is the percentage of scores in its distribution that are less than it, an exclusive definition, and one that can be expressed with a single, simple formula. Percentile scores and percentile ranks are often used in the reporting of test score from norm reference test but, as just noted, they are not the same. For percentile rank, a score is given and a percentage is computed. Percentile ranks are exclusive. If the percentile rank for a specified score is 90%, then 90% of the scores were lower. In contrast, for percentiles a percentage is given and a corresponding score is determined, which can be either exclusive or inclusive. The score for a specified percentage indicates a score below which (exclusive definition) or at or below which (inclusive definition) other scores in the distribution fall.

14. What is ANOVA?

Ans = The analysis of variance can be used to describe otherwise complex relations among variables. A dog show provides an example. A dog show is not a random sampling of the breed: it is typically limited to dogs that are adult, pure-bred, and exemplary. A histogram of dog weights from a show might plausibly be rather complex, like the yellow-orange distribution shown in the illustrations. Suppose we wanted to predict the weight of a dog based on a certain set of characteristics of each dog. One way to do that is to the distribution of weights by dividing the dog population into groups based on those characteristics. A successful grouping will split dogs such that (a) each group has a low variance of dog weights (meaning the group is relatively

homogeneous) and (b) the mean of each group is distinct (if two groups have the same mean, then it isn't reasonable to conclude that the groups are, in fact, separate in any meaningful way). ANOVA is based on the low of total variance where the observe 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 mean 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?

Ans = ANOVA stands for Analysis of Variance. One-Way Analysis of Variance tells you if there are any statistical differences between the means of three or more independent groups. 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.