STATISTICS WORKSHEET-3

Q1 TO 9 OBJECTIVE ANSWERS

1. A) Total Variation = Residual Variation – Regression Variation
2. C) binomial
3. A) 2
4. A) Type-I error
5. A) Power of test
6. B) Increases
7. B) Hypothesis
8. A) Minimize errors
9. A) 0

Q10 TO 15 SUBJECTIVE ANSWERS

10. Bayes' Theorem: It refers to a mathematical formula that describes how to update a belief, given some evidence and for calculating conditional probability and statistics. It is the most important rule in data science. It determines the probability of an event with uncertain knowledge. In probability theory, a means for revising predictions in light of relevant evidence, also known as conditional probability or inverse probability. It is a deceptively simple formula used to calculate conditional probability. For two events A and B Bayes' theorem allows to figure out the probability that event A happened, given that test B was positive[p(A|B)] from the probability that test B happened, given that event A happened[p(B|A)]. It can be a little confusing around as technically working backwards, have to switch tests and events around, which can get confusing; an example should clarify what it mean by "switch the tests and events around". It provides a way to more

calculative towards the probability of a hypothesis based on its prior probability.

11. Z-score: It is also known as standard score that gives an idea of how far a data point is from the mean. The average value of the population to which the unstandardized value belongs. It refers as a measure of the number of standard deviations by which a score is below or above the mean of a distribution. The distribution on the right is a standard normal distribution with a standard score of z = -0.60 indicated. Z-scores measures the distance of any data point from the mean in units of standard deviations and are useful because they allow us to compare the relative positions of data values in different samples. It allows to standardize two or more normal distributions, or more appropriately, to put them on the same scale. For example a standard deviation of 2 indicates the value is 2 standard deviations away from the mean. In order to use a z-score, there used the population mean and also the population standard deviation.

12. T-test: It refers to a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis testing, with null hypothesis through an inferential statistic. The calculation of a confidence interval for a sample mean. It may work on a single group differs from a known value means a one – sample t-test. There are three types of T-test are: One-sample t-test, Two-sample t-test, Paired t-test. While t-tests are relatively robust to deviations from assumptions, t-tests do assume that: the data are continuous, the sample data have been randomly sampled from a population, there is homogeneity of variance, the distribution is approximately normal. For two-sample t-tests, have independent samples. If the samples are not independent, then a paired t-test may be appropriate.

13. Percentile: The sequence to compute, which must be between 0 and 100 inclusive. One of the 100 equal parts that a set of people or things is divided into. A percentile is denoted by the letter P. Suppose we have 99 percentiles denoted by the symbols P1, P2,...,P99. In statistics, a percentile is a term that describes how a score compares to other scores from the same set. While there is no universal definition of percentile, it is commonly expressed as

the percentage of values in a set of data scores. It tells about the computational and inferential thinking of the data.

14. ANOVA: The full form of ANOVA is Analysis Of Variance. A collection of statistical models and their associated estimation procedures used to analyze the difference among means. The statistical technique that is used to check if the means of two or more groups are significantly different from each other. It checks the impact of one or more factors by comparing the means of different samples. The use of ANOVA to prove/ disprove if all the medication treatments were equally effective or not. Another measure to compare the samples is called a t-test.

15. ANOVA test is a type of statistical test used to determine if there is a statistically significant difference between two or more categorically groups by testing for differences of means using variance. There are some assumptions while doing ANOVA test are the same as the general: it can only be conducted if there is no relationship between the subjects in each sample, the different groups must have equal sample sizes, if the dependent variable is normally distributed, then only ANOVA be conducted, population variances must be equal; homogeneity of variance. There are two types: “One-Way” and a “Two-Way”. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent groups. The two-way ANOVA is used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables on a continuous outcome variable.

THANK YOU