STATISTICS WORKSHEET-3

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial
3. a) 2
4. b) Type-II error
5. b) Size of the test
6. b) Increase
7. b) Hypothesis
8. d) All of the mentioned
9. a) 0
10. 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". For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability
11. 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. A standard normal distribution (SND) is a normally shaped distribution with a mean of 0 and a standard deviation (SD)
12. The t test tells you how significant the differences between group means are. It lets you know if those differences in means could have happened by chance. The t test is usually used when data sets follow a normal distribution but you don't know the population variance. For example, you might flip a coin 1,000 times and find the number of heads follows a normal distribution for all trials. So you can calculate the sample variance from this data, but the population variance is unknown. Or, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). So while the control group may show an average life expectancy of +5 years, the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population. In addition, a t test uses a t-statistic and compares this to t-distribution values to determine if the results are statistically significant.
13. 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. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.
14. The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. For example, if a score is at the 86th percentile, where 86 is the percentile rank, it is equal to the value below which 86% of the observations may be found. In contrast, if it is in the 86th percentile, the score is at or below the value of which 86% of the observations may be found. Every score is in the 100th percentile. The

25th percentile is also known as the first quartile (Q1), the 50th percentile as the median or second quartile (Q2), and the 75th percentile as the third quartile (Q3). In general, percentiles and quartiles are specific types of quantiles.

15. Analysis of variance ANOVA is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples. We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not.

16. 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 behaviour. One-way ANOVA can help you know whether or not there are significant differences between the groups of your independent variables (such as USA vs Canada vs Mexico when testing a Location variable). You may want to test multiple independent variables (such as Location, employment status or education). When you understand how the groups within the independent variable differ (such as USA vs Canada vs Mexico, not location, employment status, or education), you can begin to understand which of them has a connection to your dependent variable (NPS score).