STATISTICS WORKSHEET-3

1) b) Total Variation = Residual Variation + Regression Variation
2) c) binomial
3) a) 2
4) a) Type-I 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 . 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 is also known as standard score** gives us an idea of how far a data point is from the mean. It indicates how many standard deviations an element is from the mean. Hence, Z-Score is measured in terms of standard deviation from the mean. 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, we need to know the population mean (μ) and also the population standard deviation (σ)

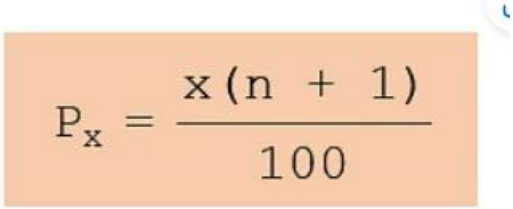
*A z-score can be calculated using the following formula.*

$$z = (X - \mu) / \sigma$$

***12)*** 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. The t-test is a parametric test of difference, meaning that it makes the same assumptions about the data as other parametric tests. The t-test assumes your data:

1. are independent
2. are (approximately) normally distributed.
3. have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

13)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 that fall below a given value. The general rule is that if a value is in the kth percentile, it is greater than K per cent of the total values.




$P_x$ = The value at which x percentage of data lie below that value

n = Total number of observations

14) 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.

ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests.

15) ANOVA is a powerful statistical tool that can help data scientists improve their data analysis and more accurately identify the factors that are influencing the data and improve their models. ANOVA can also be used to partition the total variance into between-groups and within-groups variance, which can help data scientists better understand the data.

ANOVA can help to identify the sources of variation in a data set. This can help to improve the accuracy of data predictions and analyses. Additionally, ANOVA can help to identify relationships between different variables in a data set. This information can be used to improve data models and predictions