STATISTICS WORKSHEET – 1

ANSWERSHEET

1. TRUE (This theorem is result of binary outcome)
2. Central Limit theorem
3. Modeling bounded count data
4. D
5. Poisson
6. (b) Estimated value doesn't change the CLT.
7. (b) Hypothesis
8. (a) 0
9. (C) Outliers can conform to the regression relationship.

10. **Normal Distribution** :

- ◊ It is also known as bell curve due to its presentation.
- ◊ The value of Mean, Median and Mode are exactly same in Normal Distribution.
- ◊ If you check any graph of normal distribution then you may observe that the distribution are symmetric about its mean.
- ◊ Normal distribution is a special normal distribution with the value of mean of 0 and a standard deviation of 1.

11. Data can have missing values for several reasons such as observations that were not recorded and data corruption. We can handle missing values by few methods.

- ◊ We can delete the records of missing value. (Not Applicable when we have huge Dataset.)
- ◊ Create a separate model to handle the missing value.

- ◊ **By using statistical Method: Mean {When data is distributed symmetrically}, Median {When data is distributed Asymmetrically}.**
- ◊ **Model Based Technique:**
  - **KNN**
  - **EM (Expectation maximization)**
  - **ML (Maximum Likelihood Function)**
  - **Regression Model.**

12 A/B testing is a method of comparing two versions of a design or webpage or app against each other to determine which one performs better.

13. Yes, Mean imputation of missing data is acceptable practice if data is distributed symmetrically.

14. Linear regression analysis is used to predict the value of a dependent variable on the value of independent variable using a straight line.

Linear regression are of two types.

1. **<u>Simple Linear Equation</u>: when there is only one independent variable are used to predict the value of only dependent variable. It can be shown as below.**

   Y = C+MX. Where y is independent variable (output), X is dependent variable (Input), C is intercept and M is slope of the linear equation.

2. **Multiple Linear Equation: when there is multiple independent variable are used to predict the value of only dependent variable. In Multiple Linear Equation Regression coefficient value is more important as it affects the dependent variable (Output).**

15. Branches of statistics:

1. **Descriptive Statistics:**
2. **Inferential Statistics:**

<u>Descriptive Statistics</u>: Deals with collection, presentations and Analysis of Data. There are two methods we can use in descriptive statistics.

a. **Major of central tendency (Mean, Median & Mode)**
b. **Major of Dispersion (Range, Std Deviation & Variance).**

<u>Inferential Statistics:</u> Deals on Samples from Study Populations which is used to determine whether those populations are truly different.

a. **Hypothesis Test**
b. **Regression Analysis**