**STATISTICS WORKSHEET-1**

**Ans-1**. True

**Ans-2.** Central limit Theorem

**Ans-3.** Modeling bounded count data

**Ans-4**. All of the above

**Ans-5**.Poisson

**Ans-6.** False

**Ans-7**.Hypothesis

**Ans-8.** 0

**Ans-9.** Outliers cannot conform to the regression relationship

**Ans-10**

**Normal Distribution**- Normal Distribution or Gaussian Distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus a bell shaped curve is formed.

**Ans-11.**

There are two primary method of handling the missing data which are listed below:

1. Imputation
2. Mean, Mode and Median
3. Time Series Specific Method
4. LOCF-Last Observation Carried forward

NOCB-Next Observation Carried Backward

1. Linear Interpolation
2. Deletion
3. List wise and Pairwise
4. Dropping Variable
5. Removing Data

Few Techniques used In Imputation are:

**For Numerical Variable -**

* Mean Mode and Median
* Arbitary Value Imputation
* End of tail Imputation
* Mode imputation

**For Categorical Variable –**

* Frequency category Imputation
* Adding a “Missing” Indicator

**For Both Numerical and Categorical Variable-**

* Complete Case Analysis
* Adding a Missing Indicator
* Random Sample Imputation

**Ans-12 :**

**A/b testing** is a type of experiment in which one can split their web traffic or user base into two groups and show two different versions of a web page , App ,Email and so on with the goal of comparing the result to find the more successful version.

With a **A/B test** , One element is changed between the original and the test version to see if this modification has any impact on user behavior or conversion rates.

**A/B testing** is a form of statistical hypothesis testing or a significance test.

**Ans- 13**

The process of replacing null values in a data collection with the data’s mean is known as **Mean Imputation.**

Mean Imputation is typically considered terrible practice since it ignores feature correlation.

Second, Mean Imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

**Ans-14**

**Linear Regression** is a basic and commonly used type of Predictive analysis .

The Simplest form of the Regression Equation with one Dependent and one Independent Variable is defined by the formula **y = c+b\*x** where y= estimated dependent variable score , c = constant , b= Regression coefficient and x= score on the independent variable

**Three major uses for regression analysis are :**

1. Determining the strength of Predictors
2. Forecasting an effect
3. Trend forcasting

**Ans-15:**  
Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are**two main branches** of statistics  
- Inferential Statistic.  
- Descriptive Statistic.

**Inferential Statistics:**  
Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

**Descriptive Statistics:**  
Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.