

STATISTICS WORKSHEET-1 Solution:

- 1) a - True .
- 2) a - Central Limit Theorem
- 3) b - Modeling bounded count data
- 4) d - All of the mentioned
- 5) c - Poisson
- 6) b - False
- 7) b - Hypothesis
- 8) a - 0
- 9) c - Outliers cannot conform to the regression relationship
- 10) Distribution is always normal irrespective of sample size. When mean = median= mode and distribution is symmetrical then it is normal distribution.
- 11) Missing data sets can be filled by taking mean of that particular column if data set is continuous and if data set is categorical then mode should be taken of that column to fill the missing data. There are several imputation techniques which are as follows:
 - 1) **SIMPLE Imputer** : Used to fill data by mean and mode as per the data type.
 - 2) **KNN Imputer** : It will try to find the relation with other columns and impute the data according to the relation with other columns.
 - 3) **ITERATIVE Imputer** : It takes NaN column as label and will treat one column as a feature to fill missing data.
- 12) A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.
- 13) Yes mean imputation of missing data acceptable practice.
- 14) A Linear Regression is a process for predicting a label (dependent variable) based on the feature (independent variable) at hand. For that required continuous/discrete independent variables and dependent variable should be continuous.
- 15) There are two types of statistics :
 - 1) Descriptive Statistics : Able to describe the data through mean, median and mode.
 - 2) Inferential Statistics : A huge data set is taken as a population and with sample the data set is described.