STATISTICS WORKSHEET-1

ANSWERS:

1. (a) True
2. (a) Central Limit Theorem
3. (b) Modelling 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. A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It can also be skewed to the left or skewed to the right.
11. Missing data can be dealt with in two ways: 1.) Either by deleting the rows (or columns) with missing data. 2.) By filling the missing data using various Imputation techniques.
    Recommended Imputation methods:

    - Statistical techniques like using mean of data is commonly used (median and mode also used wherever suitable).
    - Simple Imputer
    - KNNImputer
    - Iterative Imputer

12. A/B testing—also called split testing or bucket testing. It is randomized experiment where data is split in 2 and compare results to make better decisions.
13. Mean Imputation is incredibly simple but should only be used as a last resort as it has 2 major drawbacks- 1) Doesn't preserve the relationships among variables, 2) Leads to an underestimate of standard errors.
14. In Statistics, linear regression is a linear approach to modelling the relationship b/w one or more dependent variables with one independent variable.
15. There are 2 main Branches of Statistics:
    - Descriptive Statistics: Deals with collection, measurement and presentation of data. Measure of Central Tendency and Measure of Dispersion come under this branch.
    - Inferential Statistics: This branch deals with using the analysed statistical data (produced from descriptive stats) for making/inferring conclusions for the entire population.