

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. Normal distribution (or bell curve distribution) is a distribution that is symmetric about the mean, where essentially data near mean are far more frequent than data far from mean (when symmetric). 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 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 Stats, Linear Regression is a linear approach for modelling relationship between 1 or more dependent variables with 1 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.