STATISTIICS WORKSHEET-1

1: True

2: Central Limit Theorem

3: Modeling bounded count data

4: All of the mentioned

5: Poisson

6: False

7: Hypothesis

8: 0

9: Outliers cannot conform to the regression relationship

10: In Normal distribution, data distributed with no skewness and graph of normal distribution follow bell shape curve. Most of values around central region of curve. The mean, median and mode of normal distribution is exactly same. The distribution is like half of the value fall below the mean and half above the mean.

11: missing data can effect the model prediction and cause bais in the estimation of parameters.

Data set often contain missing data. There Is no single universally acceptable method to handle missing data. It is on data scientist whether drop or impute missing data. Most of time i use imputation technique to fill NaN rather than deleting Entire row or column.

Following are some imputation technique

- SimpleImputer: In this technique missing value can be be imputed with a mean and median method. It use mean for continues and median for categorical data.
- Replace with men or mode it is common method of imputing missing value. if data contain outliers this method may lead to error.
- Forward fill / Backword fill
- KNNImputer: in this technique filling missing values using the k-nearest neighbors

13: mean imputation of missing data is not acceptable practice. Because mean reduce varience of data and ignore feature correlations. Mean imputation doesn't take into account fact of relation between feature to feature. This leads to creating bias model.

14: Linear regression is a basic and commonly used predictive analysis. This regression estimate are use to explain relationship between dependent and independent variable.it is simplest form of regression equation formula $Y = a + bX$

15: There are Tow main branches in statistics

- Descriptive Statistics: this offers method to summarise data by transforming raw observation into meaningful information that is easy to interpret and share. Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures

Inferential Statistics: this offers methods to study experiments done on small samples of data and chalk out the inferences to the entire population. Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population