Statistics Worksheet:

Q.1 True

Q.2 Central Limit Theorem

Q.3 Modeling bounded count data

Q.4 All of the mentioned

Q.5 Poisson

Q.6 False

Q.7 Hypothesis

Q.8 0

Q.9 Outliers cannot conform to the regression relationship

Q.10 Normal Distribution: This means that approximately all of the data is normally distributed of independently and randomly generated variables. It has a common Bell curve structure. It represents two things the mean and standard deviation. A small standard deviation produces a steeper curve, while a large standard deviation produces a flatter curve.

Q.11 There are 3 common ways to handle missing data:

a. Mean or Median Imputation

b. MICE

c. Random Forest

Q.12 A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

Q.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. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does. 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.

Q.14 Linear Regression: It's a kind of analysis that shows relationship between two variables. One of which is called dependent variable and other is independent variable. Example: In the tips data, the amount of tips was dependent on the total bill. The higher the amount of bill the higher was the tip given.

Q.15 Branches of Statistics: There are 3 branches of statistics

a. Data collection: In this phase the actual data is collected. There are 2 terms mentioned in this, one is Population which is the entire set of data and the second is Sample which is a random portion of the total data.

b. Descriptive statistics: This phase deals with presenting the data we have. It can be done either by showing graphs, bars charts etc, or by showing average or numerical values.

c. Inferential statistics: Here we make the final conclusions on the data and presentation of data.