STATISTICS WORKSHEET-1

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. What do you understand by the term Normal Distribution?

Ans: - Normal Distribtion is ditribution of independent and random variables which occurs more frequent near the mean and lesser at far from the mean. It is also called gaussian distribution. And because of their specific data spread nature, it looks like bell. So, it's also known as bell-shaped distribution.

11. How do you handle missing data? What imputation techniques do you recommend?

Ans: - I handle missing data according to data types. If it is continuous types of data, I may prefer mean or median method. Which I can do with –
df['column_name'].fillna(df[column_name].mean()).
If the data is categorical, then I may delete them. Otherwise, I can use imputation techniques –
Simple imputer, itirative imputer or knn imputer.

12. What is A⁄B testing?

Ans: - A process of comparing two elements to find which one performs better is known as A/B testing.

13. Is mean imputation of missing data acceptable practice?

Ans: - No, it cannot be applied to all types of nulls, It can be accepted until it doesn't impact the output of the final model or score.

14. What is linear regression in statistics?

Ans: - Regression in statistics is the process of predicting a label or dependent variable based on the features (independent variables) at hand. Regression is used for time series modelling and finding the casual effect relationship between the variables and forcasting.

15. What are the various branches of statistics?

Ans: - Mainly there are two branches of statistics, first one is Descriptive statistics and the second is Inferential statistics.
Former only make statements about the set of data from which they were calculated, it never goes beyond the data it has. And the later can work with the sample data and infer the result to describe entire population.