Statistics_work_sheet_1

1.

Ans:- True

2.

Ans:- Central Limit Theorem

3.

Ans:- Modeling contingency tables

4.

Ans:- The square of a standard normal random variable follows what is called chi-Squared distribution

5.

Ans:- All of the mentioned(Empirical, Binomial, Poisson Distribution)

6.

Ans:-False

7.

Ans:-Hypothesis

8.

Ans:-0

9.

Ans:- Outliers cannot conform to the regression relationship

10. what does you understand by the term Normal Distribution?

Ans:- 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 has zero skew and a kurtosis of 3.Normal distributions are symmetrical, but not all symmetrical distributions are normal.

11. how do you handle missing data? What imputation technique do you recommend?

Ans:1. We can handle the missing data by deleting the rows with missing data

2. Replacing the missing value with mean / mode:

   Columns in the dataset which are having numeric continuous values can be replaces with mean, median or mode of remaining values in the column. This method can prevent the loss of data compared to earlier method

3.Imputation method for the categorical column:

When missing values is from categorical columns(string) then the missing values can be replacing with the most frequent category. If the number of missing values is very large then it can be replaced with a new category.

4.using algorithms:

1.Knn, Naive bayes can also support missing values when making a prediction.

2. Random Forest that works well on non-linear and categorical data.

12. What is A/B testing?

Ans:- A/B testing is basically statistical hypothesis testing or statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13.Is mean imputation of missing data acceptable practice?

Ans:- True, imputing the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimate of the mean remains unbiased

14.What is linear Regression in statistics?

Ans:- In statistics ,Linear regression is a linear approach for modelling the relation between a scalar response and one or more explanatory variables also know as dependent and independent variables..

15.What are the various branches of statistics?

Ans:- statistics has 3 real branches

1.Data collection

2.Discriptive statistics

3.inferencetial statistics