

STATISTICS WORKSHEET-1

- 1.** True
- 2.** Central Limit Theorem
- 3.** Modeling contingency tables
- 4.** All of the mentioned
- 5.** Poisson
- 6.** False
- 7.** Hypothesis
- 8.** 0
- 9.** Outliers cannot conform to the regression relationship

Q 10. What do you understand by the term Normal Distribution?

Ans: A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

Height is one simple example of something that follows a normal distribution pattern: Most people are of average height, the numbers of people that are taller and shorter than average are fairly equal and a very small (and still roughly equivalent) number of people are either extremely tall or extremely short.

A graphical representation of a normal distribution is sometimes called a bell curve because of its flared shape. The precise shape can vary according to the distribution of the population but the peak is always in the middle and the curve is always symmetrical. In a normal distribution, the mean, mode and median are all the same.

Normal distribution curves are sometimes designed with a histogram inside the curve. The graphs are commonly used in mathematics, statistics and corporate data analytics.

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

Ans While handling missing data there are two primary methods used to solve the error: imputation or the removal of data.

The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.

The other option is to remove data. When dealing with data that is missing at random, related data can be deleted to reduce bias. Removing data may not be the best option if there are not enough observations to result in a reliable analysis. In some situations, observation of specific events or factors may be required.

Imputation techniques:

Complete Case Analysis(CCA):- This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.

Arbitrary Value Imputation.

Frequent Category Imputation.

Q.12 What is A/B testing?

Ans: A/B testing is also known as split testing or bucket testing. It is a method of comparing two versions of a webpage or app against each other to determine which one performs better.

Q.13. Is mean imputation of missing data acceptable practice?

Ans: Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations. That is why mean imputation is not considered as a good practice.

Q.14 What is linear regression in statistics?

Ans. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Q.15 What are the various branches of statistics?

Ans: There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.