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: The Normal Distribution, also called the Gaussian Distribution, is the most significant continuous probability distribution. Sometimes it is also called a bell curve. A large number of random variables are either nearly or exactly represented by the normal distribution.The Normal Distribution is defined by the probability density function for a continuous random variable in a system.

The essential characteristics of a normal distribution are:

It is symmetric, unimodal and asymptotic.The values of mean, median, and mode are all equal.

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

ANS: 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.

There are different type of imputation: Mean imputation, subsitution method , hot deck imputation ,cold deck imputation , regression imputation , single imputation and mutiple imputation.

Multiple imputation is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values.

12. What is A/B testing?

Ans:A/B testing, also known as split testing, A/B testing allows any organization to be more data-driven and strategic about their digital communications. It removes the guesswork from decision making and lets the data decide the path forward. split testing helps facilitate the conversation to focus more on the data, rather than opinion or emotion.

13. Is mean imputation of missing data acceptable practice?

Ans : Yes,It is acceptable when the missing value proportion is not large enough. But, when the missing values are larger enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been. Almost all of the methods are superior to mean imputation like single and multiple imputation , regression imputation etc

14.What is linear regression in statistics?

Ans:Linear regression strives to show the relationship between two variables by applying a linear equation to observed data. One variable is supposed to be an independent variable, and the other is to be a dependent variable. For example, the weight of the person is linearly related to his height. Hence this shows a linear relationship between the height and weight of the person. As the height is increased, the weight of the person also gets increased.

15.What are the various branches of statistics?

Ans: The two main branches of statistics are:

Descriptive Statistics: deals with the presentation and collection of data. This is usually the first part of a statistical analysis.

Inferential Statistics: drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics. It makes the predictions and inferences.