STATISTICS WORKSHEET-1

1. 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. <u>NORMAL</u> <u>DISTRIBUTION</u>

   A large number are variables are either nearly or exactly represented by the normal distribution. It is also called Gaussian Distribution. Sometimes it is called bell curve.

12. What is A/B testing?

   A/B testing is a popular way to test your products and is gaining steam in the data science field.

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

For instance, let's say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

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

   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. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Answer:

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more

independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

15.What are the various branches in statistics?

Answer:

1) Descriptive Statistics

Descriptive statistics is the first part of statistics that deals with the collection of data. People think it is too easy, but it is not that easy. The statisticians need to be aware of the design and experiments. They also need to select the correct focus group and keep away from biases. On the contrary, Descriptive statistics are used to do various kinds of analysis on different studies.

2) Inferential Statistics

Inference statistics are techniques that enable statisticians to use the information collected from the sample to conclude, bring decisions, or predict a defined population.

Inference statistics often speak in terms of probability by using descriptive statistics. Besides, a statistician uses these techniques for data analysis, drafting, and making conclusions from limited information. That is obtained by taking samples and testing how reliable they are.