STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.

Ans: True

1. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans: Central Limit Theorem

1. Which of the following is incorrect with respect to use of Poisson distribution?

Ans: Modeling bounded count data

1. Point out the correct statement.

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

1.  random variables are used to model rates.

Ans: Poisson

1. Usually replacing the standard error by its estimated value does change the CLT.

Ans: False

1. Which of the following testing is concerned with making decisions using data?

Ans: Hypothesis

1. Normalized data are centered at and have units equal to standard deviations of the original data.

Ans: 0

1. Which of the following statement is incorrect with respect to outliers?

Ans: Outliers cannot conform to the regression relationship

1. What do you understand by the term Normal Distribution?

Ans: A normal distribution is similar as bell curve. For example in office if we do annual appraisal for employee and we score them according to their performance. So majority of employee will score average (C), smaller number of employee will score a B and D. An even very few will score an A. Half of data will fall right of mean; and half will half on left and whwill create proper bell curve.

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

Ans: We can try different models Random Forest, Mean-Media-Mode, Linear Regression and KNN(K Nearest Neighbour).

1. What is A/B testing?

Ans: A/B testing is a basic randomized control experiment. It is a way to compare the two version of a variable to find out best one. A/B testing is one of the most prominent and widely used statistical tool.

1. Is mean imputation of missing data acceptable practice?

Ans: Yes! Outliers data points will have a significant impact on the mean and hence, in such cases, it is not recommended to use the mean for replacing the missing values.

1. What is linear regression in statistics?

Ans: Linear regression is predictive analysis which is most commonly used. These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. Naming the Variables.  There are many names for a regression’s dependent variable.  It may be called an outcome variable, criterion variable, endogenous variable, or regress. The independent variables can be called exogenous variables, predictor variables, or regressors.

1. What are the various branches of statistics?

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