Statistics Worksheet 1

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

Answer: True.

QUES 2: 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?

Answer: Central Limit Theorem

QUES 3: Which of the following is incorrect with respect to use of Poisson distribution?

Answer: B: Modeling bounded count data.

QUES 4: Point out the correct statement.

Answer: D): All of the mentioned.

QUES 5: _____ random variables are used to model rates.

Answer: c) Poisson.

QUES 6: Usually replacing the standard error by its estimated value does change the CLT.

Answer: False.

QUES 7: Which of the following testing is concerned with making decisions using data?

Answer: B) Hypothesis.

QUES 8: Normalized data are centered at _____and have units equal to standard deviations of the original data.

Answer: A) Zero.

QUES 9: Which of the following statement is incorrect with respect to outliers?

ANSWER: C) Outliers cannot conform to the regression relationship.

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

Answer: Normal Distribution is a continuous probability distribution wherein values lie in a symmetrical fashion mostly situated around the mean. It is also known as Gaussian distribution or probability bell curve.

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

Answer: Two primary methods to solve missing data is:

1) Removing the data.
2) Imputation.

Imputation is the process of substituting an estimate for missing values and analyzing the entire data set as if the imputed values were the true observed values. Various Imputation methods are:

A) Mean Imputation. B) Substitution c) Hot Deck Imputation. D) Cold Deck Imputation. E) Regression Imputation.

QUES 12. What is A/B testing?

Answer: A/B testing refers to the experiments where two or more variations of the same webpage are compared against each other by displaying them to real-time visitors to determine which one performs better for a given goal.

QUES 13: Is mean imputation of missing data acceptable practice?

ANSWER: Mean Imputation is typically considered terrible practice since it ignores feature correlation.

Mean Imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

QUES 14: What is linear regression in statistics?

Answer: 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.

QUES 15: What are the various branches of statistics?

Answer: There are three real branches of statistics: Data Collection, Descriptive Statistics and Inferential Statistics.

Data Collection: Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

Descriptive Statistics: Descriptive statistics describe, show, and summarize the basic features of a dataset found in a given study, presented in a summary that describes the data sample and its measurements. It helps analysts to understand the data better.

Inferential Statistics: Inferential statistics use measurements from the sample of subjects in the experiment to compare the treatment groups and make generalizations about the larger population of subjects.