

STATISTICS WORKSHEET-1

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

Ans – True (Bernouli random variable take only the value 0 and 1)

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?

Ans - Central Limit thermo.

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

Ans- (b) Modeling bounded count data.

4- Point out the correct statement

Ans - d) All of the mentioned.

5- _____ random variables are used to model rates.

Ans - c) Poisson

6- Usually replacing the standard error by its estimated value does change the CLT

Ans- (b) Usually replacing the standard error by its estimated value doesn't change the CLT

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

Ans- b) Hypothesis

8- Normalized data are centered at _____ and have units equal to standard deviations of the original data

Ans- (A) 0

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

Ans – (c) Outliers can conform to the regression relationship.

10- What do you understand by the term Normal Distribution

Ans- Normal Distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

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

Ans- Data Dropping.

Mean/Median Imputation.

Random Sample Imputation.

Multiple Imputation

The Best imputation Techniques is **Model-based Imputation: This involves creating a statistical model to predict the missing values based on other features in the data. This can be a powerful technique**

12- What is A/B testing.

Ans- A/B testing (also known as split testing or bucket testing) is a methodology for comparing two versions of a webpage or app against each other to determine which one performs better.

13- Is mean imputation of missing data acceptable practice

Ans- Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, 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.

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.

15- What are the various branches of statistics

Ans- The two main branches of statistics are descriptive statistics and inferential statistics.