WORKSHEET

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

a) True

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?

d) All of the mentioned

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

b) Modeling bounded count data

4. Point out the correct statement.

d) All of the mentioned

5. _____ random variables are used to model rate

c) Poisson

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

b) False

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

b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0

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

c) Outliers cannot conform to the regression relationship

WORKSHEET

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans. As with any probability distribution, the normal distribution describes how the values of a variable are distributed. It is the most important probability distribution in statistics because it accurately describes the distribution of values for many natural phenomena.

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

Ans. For handling missing data we can do 2 things. Imputation or the removal of data. The imputation method develops reasonable guesses for missing data. Its most useful when the percentage of missing data is low.

MEAN, MEDIAN AND MODE

This is one of the most common methods of imputing values when dealing with missing data.

12. What is A/B testing?

Ans. A/B testing is also known as split testing, refers to a randomized experimentation process wherein two or more version of variable are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business matrics.

13. Is mean imputation of missing data acceptable practice?

Ans. NO, mean Imputation of missing data is not acceptable practice.

14. What is linear regression in statistics?

Ans. Linear regression attempts to model the relationship between two variable by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to bea dependent variable.

15. What are the various branches of statistics?

Ans. 1 Data Collection, 2. Descriptive statistics, 3. Inferential statistics.