STATISTICS WORKSHEET – 1

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

Ans. 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?

Ans. Central Limit Theorem

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

Ans. Modeling bounded count data

4. Point out the correct statement.

Ans. All of the mentioned

5. _____ random variables are used to model rates

Ans. Poisson distribution

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

Ans. False

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

Ans. Hypothesis

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

Ans. 0

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

Ans. Outliers cannot confirm to the regression relationship

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

Ans. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmatric about the mean, showing that data near the mean are more frequent in occurance than data far from the mean.

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

Ans. “Mean” will replace missing values using the mean in each column. It is preferred if data is numeric and not skewed. “Median” will replace missing values using the median in each column if data is numeric and skewed.

12. What is A/B testing?

Ans. A/B testing is a user experienced research methodology. A/B tests consist of a randomized experiment with two variants, A and B . it includes application of stastical hypothesis testing are used in the field of statistics.

13. Is mean imputation of missing data acceptable practice?

Ans. Mean imputation is typically considered terrible practices since it ignores feature correlation.

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. Tha 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.

15. What are the various branches of statistics?

Ans. The two major areas of statistics are descriptive and inferential statistics.