Name: NALLA NANDAKISHORE

Batch: DS2302

Institute: DataTrained

Internship: Statistics Worksheet Solutions – **Assignment - I**

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.

Ans: 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?

Ans: a) Central Limit Theorem

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 d) All of the mentioned

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

Ans: b) False

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

Ans: b) Hypothesis

8. 4. 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 cannot conform to the regression relationship

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: In a normal distribution, data is symmetrically distributed with no skew. When plotted on a

graph, the data follows a bell shape, with most values clustering around a central region and

tapering off as they go further away from the center

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

Ans: One way of handling missing values is the deletion of the rows or columns having null values. Mean imputation Substitution, Hot deck imputation, Regression imputation

12. What is A/B testing?

Ans: A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?

Ans: It's not appropriate

14. What is linear regression in statistics?

Ans: Regression analysis helps in determining the cause and effect relationship between variables

15. What are the various branches of statistics?

Ans: Descriptive statistics and inferential statistics.