Rumnaz Binte Islam Sarah

Statistics Worksheet 1

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)

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

Ans. b(False)

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 cannot conform to the regression relationship)

10. What do you understand by the term normal distribution?

Ans. Normal distribution is a representation of frequencies of a dataset in the form of a bell curve. The mean is 0 and standard deviation is 1 for the data. Cumulative distribution function is used to find the standardized random variables for normal distribution.

11. How do you handle missing data?

Ans. There are different ways of handling missing data.
When analyzing a big dataset, we can completely remove the null data points. However, if the dataset is small, it is possible to use the available variable information to train the model and find the missing values. Another way is to first check the missing values, maybe by using a heat map. We can then check the mean, median and percentiles using a boxplot. This will give a

clear picture of how the data is distributed and then we can replace a realistic value with Nan. The new data can be used to train the model.

12. What is A/B Testing?

Ans. A/B Testing is used to understand user preferences based on the number of times or the number of users chooses or clicks on one product (A) compared to another (B). So, it tells us about the user behavior about a product or even how it is being presented to them. This kind of testing is also known as bucket testing or split-run testing.

13. Is mean imputation of missing data acceptable practice?

Ans. No, it is not.

14. What is linear regression in statistics?

Ans. Linear regression in statistics is used to predict values of variables using values of variables that are present. The linear regression formula helps us to find the best fit line for paired data to estimate the dependent variables from the independent variables.

15. What are the various branches of statistics?

Ans. The various branches of statistics are:

1. Mathematical or Theoretical Statistics
2. Statistical Methods or Functions
3. Descriptive Statistics
4. Inferential Statistics.