WORK SHEET_1 STATISTICS

Answers from Q1 to Q9

1. Bernoulli random variables take (only) the values 1 and 0. a) True b) False

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. It is a bell shaped curve. In normal distribution, the total area under curve is equal to 1. The data is distributed symmetrical. The Mean = Median = Mode. The mean is 0 and the standard deviation is 1.

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

Ans. The most common strategy to handle missing data is Imputation. Imputation is a process of substituting an estimate for the missing values. There are various techniques to handle missing data. Mean and Median are the simplest imputation techniques to replace the missing values. If not many rows contain missing data then we can drop the rows.

12. What is A/B testing?
Ans. A/B testing is a split test. It is a way to compare the two tests and find out which performs the best.

13. Is mean imputation of missing data acceptable practice?
Ans. Mean imputation of missing data is the simplest imputation but it is not acceptable due to the following reasons:

i. This does not preserve the relationship among variables
ii. Mean imputation leads to the underestimation of the standard errors

14. What is linear regression in statistics?

Ans. Linear regression is a basic and commonly used type of predictive analysis. The simplest form of regression equation is y = mx + c, where y is estimated dependent value, m is coefficient, x is data or feature and C is the constant/intercept.

15. What are the various branches of statistics?

Ans. Two main branches of statistics are Descriptive statistics and Interferntial statistics.
1. Descriptive statistics deals with presentation and collection of data. It can be categorised into
i. Measure of tendency (Mean, Median and Mode)
ii. Measure of Dispersion(Variance and Standard Deviation)

2. Interferntial Statistics deals with large amount of data from a sample to predict the behaviour of the given population. It can be categorised into
i. Normal Distribution
ii. Z- Distribution (ANOVA, Chi – Square Test, Ttest)