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.

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

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. The random variable is said to be normally distributed, when it forms bell shaped curve where the mean is zero and standard deviation is one.It is symmetrical in shape which has zero skewness and kurtosis is equal to 3.

Properties of Normal Distribution.

(i) Mean-Median=Mode
(ii) 68.25% data falls in one std from mean,95% falls from second std. From mean and 99% data falls from third std from mean.
(iii) Standard deviation is 1.

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

Ans. Checking missing value :-The first step in handling missing values is to look at the data carefully and find out all the missing values.

i. Delete the record missing value.
ii. Create the separate model to handle missing value.
iii. Statistical method through mean, median and mode.
iv. Various imputation techniques are:
v. Simple Imputer from sklearn library
vi. Fillna from pandas library

12. What is A/B testing?

Ans. A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better.

13. Is mean imputation of missing data acceptable practice?

Ans. Mean imputation of missing data can be used in practice based on scenario, if numerical data with minimal range is given, otherwise mean imputation hamper the correlation of features.

14. What is linear regression in statistics?

Ans. Linear regression analysis used to predict the value of variable based on value of another variable. The main objective is to predict output variable. The equation is given below: y=mx+c

15. What are the various branches of statistics?

Ans. There are two main branches of statistics

1. Inferential Statistic.

2.Descriptive Statistic.