STATISTICS WORKSHEET-1

Answers Highlighted in Yellow
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.
a) True
b) False

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?
a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling event/time data
b) Modeling bounded count data
c) Modeling contingency tables
d) All of the mentioned

4. Point out the correct statement.
a) The exponent of a normally distributed random variables follows what is called the log-normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

7. 1. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

8. 4. Normalized data are centered at_____ and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

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

a) Outliers can have varying degrees of influence

b) Outliers can be the result of spurious or real processes

c) Outliers cannot conform to the regression relationship

d) None of the mentioned

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 Data is said to be normally distributed when the mean, median and mode are Equal. The data is symmetrically distributed along the center. It is normally denoted by a bell-shaped curve.

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

Ans: The Different techniques used to handle missing data are as follows:

1. Deleting the rows and columns from which the data is missing.
2. Replacing the missing data by mean or median
3. Using Regression Analysis
4. Imputation Techniques: Average Imputation

12. What is A/B testing?

Ans: A/B testing is basically a randomized control experiment in which we compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Ans: Mean Imputation is acceptable when data is missing completely at random, hence it is unbiased, but when relationship among variables is important, mean imputation is not a good solution as it will lead to variation in data

14. What is linear regression in statistics?

Ans: In statistics, linear regression is a linear approach for modelling the relationship between dependent and independent variables.

15. What are the various branches of statistics?

Ans: The various branched of Statistics are

1. Descriptive
2. Inferential
3. Decision Theory