STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

Q1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Answer: Option A (True)

Q2. 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

Answer: Option (A) Central Limit Theorem

Q3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modelling event/time data

b) Modelling bounded count data

c) Modelling contingency tables

d) All of the mentioned

Answer: Option B (Modelling bounded count data)

Q4. 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

Answer: Option D (All of them mentioned)

Q5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Answer: Option C (Poisson)

Q6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Answer: Option (False)

Q7. 1. Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

Answer: Option B (Hypothesis)

8. 4. Normalized data are centred at_____and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

Answer: Option A (0)

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

Answer: Option 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?

Answer: **Normal distribution, it is also called as Gaussian distribution, is probability that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. It forms a bell shape curve in a graph when constructed.**

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

Answer: **Missing values can be handled by deleting the rows or columns having null values. If columns have more than half of the rows as null then they can be replaced accordingly by replacing it with mean, mode, median.**

Techniques:

1. Multiple imputations

2.Pairwise deletion

3.Mean substitution

12. What is A/B testing?

Answer: **A/B testing which is also called 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?

Answer: **Yes True, imputing the mean preserves the mean of the observed data.**

14. What is linear regression in statistics?

Answer: **The process of finding a best straight line (as by least squares) that best approximates a set of points.**

15. What are the various branches of statistics?

Answer: **The two main branches of statistics are descriptive statistics and inferential statistics.**