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.
   a) True
   b) False

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

Answer : - a) Central Limit Theorem

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

Answer : - b) Modeling bounded count data

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

Answer : - c) The square of a standard normal random variable follows what is called chi-squared distribution

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

Answer : - c) Poisson

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

Answer : - 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

Answer : - b) Hypothesis

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

Answer : - 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 : - 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?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10. Normal distribution has a bell-shaped curve and describes the probability distribution of random variables that is symmetric around the mean of data. It has two parameters, mean and standard deviation, as mean represents the centre of distribution and standard deviation shows the spread of data distribution.

11. if we have missing data then which denotes nan in our data, if we less no of missing data then we can remove it by dropna function and if we have more missing data then we use mean, median or regression imputation and if we have missing data in string data then we can use mode method also , the imputation we applied as per nature of missing data.

12. A/B testing is a user experience research methodology that tests two or more versions of a variable to determine which version is more effective. It eliminates guesswork and enables experience optimizers to make data-backed decisions.

13. yes, Mean imputation should be commonly used to handle missing data but it has some drawback also as it does not added with relationship among the variable so it is used only when the missing data is missing completely at random.

14. Linear regression is a statistical method for show the connection between one or more independent variables and a dependent variable. It is used to analyze and predict future of a dependent variable based on the values of one or more independent variables. Basically it can use predict future of dependent variable with using different independent variables.

15. There are two main branches of statistics are descriptive statistics and inferential statistics,
In descriptive statistics it describe data using mean, median, mode, for centrality of data, range , standard devation for spread of data.
In inferential statistics we using hypothesis testing.