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

Ans : d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical b) Binomial c) Poisson d) All of the mentioned

Ans : c) Poisson

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

a) True b) False

Ans : 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

Ans : b) Hypothesis

8.Normalized data are centered at_____and have units equal to standard deviations of the original data. a) 0 b) 5 c) 1 d) 10

Ans : 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

c) Outliers cannot conform to the regression relationship

10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

In probability theory and statistics, the Normal Distribution**, also called the** Gaussian Distribution**, is the most significant continuous probability distribution. Sometimes it is also called a bell curve.**

11. How do you handle missing data? What imputation techniques do you recommend?
    Ans :
    Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

12. What is A/B testing?
    A/B testing (sometimes called split testing) is comparing two versions of a web page to see which one performs better. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "twosample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?
    Ans :
    The process of replacing null values in a data collection with the data's mean is known as mean imputation.
    Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

14. What is linear regression in statistics?
    Ans :
    Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?
    Ans :
    There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.