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

4.

Ans: b) Modeling bounded count data

5. 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

FLIP ROBO

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

Ans: c)Poisson

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

Ans: a)True

8. 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

9. 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

Ans: a) 0

10. 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

Ans: c) Outliers cannot conform to the regression relationship

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans: Normal distribution is a distribution in which the distribution of probability is symmetric on both sides of the mean. The area under the curve equals one and the tail is asymptomatic that is it never touches the horizon.

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

Ans: If more than 50% of the data s are missing its better to drop them. Fillna can be used to fill the missing values. Techniques recommended are Simple Imputer.

13. What is A/B testing?

Ans: A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample 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.

FLIP ROBO

14. Is mean imputation of missing data acceptable practice?

Ans: Mean imputation is a method in which the missing value on a certain variable is replaced by the mean of the available cases. This method maintains the sample size and is easy to use, but the variability in the data is reduced, so the standard deviations and the variance estimates tend to be underestimated. The magnitude of the covariances and correlation also decreases by restricting the variability and this method often causes biased estimates, irrespective of the underlying missing data mechanism.

15. What is linear regression in statistics?

Ans: It is a technique in which the dependent variable is continuous in nature. The relationship between the dependent variable and independent variables is assumed to be linear in nature.

Assumptions of linear regression:
There must be a linear relation between independent and dependent variables.
There should not be any outliers present.
No heteroscedasticity
Sample observations should be independent.
Error terms should be normally distributed with mean 0 and constant variance.
Absence of multicollinearity and auto-correlation.

16. What are the various branches of statistics?

Ans: Branch of Statistics
Descriptive Statistics
Inferential / inductive Statistics.

1.Descriptive Statistics:

A branch of statistics in which we analyze and interpret the results of collected and arranged data is Descriptive Statistics. In this branch of Statistics some basic calculations are made on the data to observe the characteristics of data and no conclusion is drawn from the calculations.

2 Inferential I inductive Statistics:

A branch of statistics in which we analyze and interpret the results of collected and arranged data and also draw conclusions about the population from the interpreted results of collected data is called Inferential / inductive Statistics. Inferential Statistics is used in making decisions and predictions at some future time. In this branch of Statistics some basic calculations are made on the data to draw conclusions about the population from the sample data.

FLIP ROBO