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

   CORRECT ANSWER: a

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

   CORRECT ANSWER: a

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

   CORRECT ANSWER: a

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

   CORRECT ANSWER: d

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

   CORRECT ANSWER: c

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

   CORRECT ANSWER: a

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

CORRECT ANSWER: b

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

CORRECT ANSWER: a

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

CORRECT ANSWER: a

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: Normal distribution is:

- Generally all the variables governed by nature follow Normal distribution
- Errors in the Regression model are assumed to follow Normal distribution
- Normal distribution is symmetric and Mesokurtic.
- If the sample size is large sample mean taken from any distribution follows Normal distribution.

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

Ans: Generally due to unawareness of importance of the data, or due to man power or due to less time availability data are not completely filled, and some fields in data set remain unfilled.

- If the feature having missing data has qualitative data then it is filled by modal value which is the most repeated value.
- If the feature having missing data has quantitative data, then it is filled by arithmetic mean. Depending on the situation we can assign class-wise arithmetic mean.
- Sometimes if situation is that the replacement is not meaningful then one can ignore such records.

Imputation techniques recommended is: 1) One hot encoder 2) Ordinal Encoder

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?

Ans: When missing value occurs in Feature variable having continuous or quantitative data then mean imputation is meaningful.

14. What is linear regression in statistics?

- After drawing the scatter diagram if one finds the linear relationship between feature variable and response(Label) variable then one can set the linear model (power of all variables is 1). Such a model is called as linear regression. $Y=b0+b1X1+b2X2+\ldots+bkXk+\varepsilon$
- Using the linear regression equation one cam estimate or predict the response (Label) variable.

15. What are the various branches of statistics?

Ans: Various branches of Statistics are

- Descriptive statistics
- Inferential Statistics
- Bayesian Statistics
- Asymptotic inference


FLIP ROBO