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
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
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. 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
5. _____ random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   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
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
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

FLIP ROBO

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?

FLIP ROBO

15. What are the various branches of statistics?

ANSWERS

Ans1- A) true

Ans2- A) central limit theorem

Ans3- C) modeling contingency tables

Ans4- D) all of the mentioned

Ans5- C) poission

Ans6- A) True

Ans7- B) hypothesis

Ans8- A)0

Ans9- C) outliers cannot conform to the regression relationship

Ans10- Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

Ans11- Types of missing data are:

1.missing completely at random (MCAR)
2. missing at random (MAR)
3.not missing at random (NMAR)

Common methods to handle missing data are:

1.mean or median imputation
2.multivariate imputation by chained equations (MICE)
3.random forest

Ans12- A/B testing is a user experience research methodology that consists of randomized experiment with 2 variants a and b. Comparing 2 versions of webpage or app to determine which one performs better.

Ans13- Imputing the mean preserves the mean of the observed data so if the data are missing completely at random the estimate of the mean remains unbiased but since most research studies are interested in the relationship among variables , mean imputation is not a good solution

Ans14- In statistics , linear regression is a linear approach for modelling the relationship between the scalar response and one or more explanatory variables.

Ans15- The 2 main branches of statistics are :

1. Descriptive statistics
2. Inferential statistics