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

ANSWERS

1.(a)True

2.(b)Central limit theorem

3.(c)Modelling bounded count data

4.(d) All of the mentioned

5.(c)Poisson

6.(a) True

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

Q10 and 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?

ANSWERS

7.(b)Hypothesis

8.(a)0

9.(c)Outliners cannot confirm to regression relationship

10. It is a probability distribution that is symmetric about mean it is a continous probability.

11. A common teqnique is to use the mean or median of the non-missing observations which is use to handle the missing data.

Imputation teqniques:- CCA(complete case analysis), Arbitary Value Imputation, Frequent Category Imputation

12. A/B testing in its simplest sense is an experiment on two variants to see which performs better based on a given metric.

13. True ,imputing the mean of observed data , So if data are missing complpetely at random , estimate of mean remain same

14. Linear Regression is a linear approach for modeling the relationship between a scaler response and one or more explanatory variables

15. Data collection statistics,Descriptive statistics,Inferential statistics