

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

Answer = 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

Answer = 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

Answer = 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

Answer = d) All of the mentioned

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

Answer = c) Poisson

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

Answer = 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

Answer = b) Hypothesis

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

Answer = 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

Answer = c) Outliers cannot conform to the regression relationship

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?

Answer = It is an arrangement of data set in which most values cluster in middle of the range and rest taper off symmetrically toward either extreme

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

Answer = Use regression analysis to systematically eliminate data, use data imputation technique, use deletion method to eliminate missing data

Imputation Technique are frequent category imputation, complete case analysis, arbitrary value imputation

12. What is A/B testing?

Answer = It is a method of comparing two version of web page or app against each other to determine which one performs better

13. Is mean imputation of missing data acceptable practice?

Answer = No

14. What is linear regression in statistics?

Answer = Linear regression is approach for modelling the relationship between a scalar response and one or more explanatory variables

15. What are the various branches of statistics?

Answer = Data collection, descriptive statistics and inferential statistics