

Velivela lakshmikanth Chowdary ; Ds2302

STATISTICS WORKSHEET-1

please note the answers are marked in red

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

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?

Ans: The mean, median, and mode are equal. The middle point of a normal distribution is the point with the maximum frequency, which means that it owns the most explanations of the variable

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

Ans: One way of handling missing values is the deletion of the rows or columns having null values. If any columns have more than half of the values as null then you can drop the entire column. In the same way, rows can also be dropped if having one or more columns values as null

Some of imputation techniques are

Mean imputation

Substitution, Hot deck imputation, Cold deck imputation, Regression imputation

12. What is A/B testing?

Ans

A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs. variant B and concluding which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?

Ans no it's a least considered method

14. What is linear regression in statistics?

Ans :it is used to explain the relationship between the data and describe it ,it's a kind of predictive analysis

15. What are the various branches of statistics?

Ans; descriptive statistics and inferential statistics.



FLIP ROBO