

STATISTICS WORKSHEET

Q1 Bernoulli random variables take (only) the values 1 and 0

Ans.: A) True

Q2. 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?

Ans. A) Central Limit Theorem

Q3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans. B) Modeling bounded count data

Q4. Point out the correct statement

Ans. D) All of the mentioned

Q5. _____ random variables are used to model rates

Ans. C) Poisson

Q6. Usually replacing the standard error by its estimated value does change the CLT.

Ans. B) False

Q7. Which of the following testing is concerned with making decisions using data?

Ans. B) Hypothesis

Q8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Ans. A) 0

Q9. Which of the following statement is incorrect with respect to outliers?

Ans. C) Outliers cannot conform to the regression relationship

Q10. What do you understand by the term Normal Distribution?

Ans. Normal Distribution is a statistical probability function that describes how data values are distributed. After plotting the value, a bell-shaped curve graph is formed. The peak point is the data set's mean, and half of the data set's values are on the left side of the mean, while the other half is on the right side, as seen in the graph.

Q11. How do you handle missing data? What imputation techniques do you recommend?

Ans. Missing data can be addressed in a number of ways. The most usual reaction is to disregard it or delete the row but sometimes it will have negative impact on the output.

Another option is imputation.

Mean Imputation

Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean, but it also has a number of disadvantages.

Substitute

Assume the value of a new person who isn't part of the sample. To put it another way, find a new topic and focus on it instead.

Q12. What is A/B testing?

Ans. An AB testing is a type of statistical hypothesis testing in which a hypothesis is formed about the relationship between two data sets, and then the two data sets are compared to see if there is a statistically significant relationship.

Q13. Is mean imputation of missing data acceptable practice?

Ans. Mean imputation is the process of replacing null values in a data set with the data's mean.

Q14. What is linear regression in statistics?

Ans. Linear regression is a statistical analysis for predicting the value of one variable based on the value of another. The dependent variable is the variable you want to predict. The independent variable is the variable you're utilising to predict the value of the other variable.

Q15. What are the various branches of statistics

Ans. There are two branches of statistics.

1. Descriptive Statistics
2. Inferential Statistics

Descriptive Statistics: Descriptive statistics are used to describe or summarize data in ways that are meaningful and useful.

Inferential Statistics: inferential statistics allows you to make predictions from data.