

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
 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
 3. Which of the following is incorrect with respect to use of Poisson distribution?
B) Modeling bounded count data
 4. Point out the correct statement.
D) All of the mentioned
 5. _____ random variables are used to model rates.
C) Poisson
 6. Usually replacing the standard error by its estimated value does change the CLT.
a) True
 7. Which of the following testing is concerned with making decisions using data?
B) Hypothesis
 8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.
a) 0
 9. Which of the following statement is incorrect with respect to outliers?
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?

The normal distribution, also known as the Gaussian or standard normal distribution, is the probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated around the probability's mean. Values are equally likely to plot either above or below the mean.

The normal distribution is a probability distribution that describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques. In graph form normal distribution will appear as a bell curve.

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

Complete Case Analysis(CCA), This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.

12. What is A/B testing?

A/B also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

13. Is mean imputation of missing data acceptable practice?

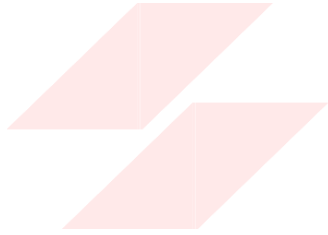
True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean **remains unbiased**. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Linear regression quantifies the relationship between one or more predictor variables and one outcome variable. Linear regression is commonly used for predictive analysis and modeling.

15. What are the various branches of statistics?

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.



FLIP ROBO