

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?
a) Modeling bounded count data
4. Point out the correct statement.
a) All of the mentioned
5. _____ random variables are used to model rates.
a) Poisson
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) False
7. 1. Which of the following testing is concerned with making decisions using data?
a) Hypothesis
8. 4. 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?
a) 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?

10. The normal distribution model is important in statistics and is key to the Central Limit Theorem CLT. This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled provided it has finite variance.

11. Understanding the nature of missing data is critical in determining what treatments can be applied to overcome the lack of data. Data can be missing in the following ways:

Missing Completely At Random When missing values are randomly distributed across all observations, then we consider the data to be missing completely at random. A quick check for this is to compare two parts of data – one with missing observations and the other without missing observations. On a t-test, if we do not find any difference in means between the two samples of data, we can assume the data to be MCAR.

Missing At Random The key difference between MCAR and MAR is that under MAR the data is not missing randomly across all observations, but is missing randomly only within sub-samples of data. For example, if high school GPA data is missing randomly across all schools in a district, that data will be considered MCAR. However, if data is randomly missing for students in specific schools of the district, then the data is MAR.

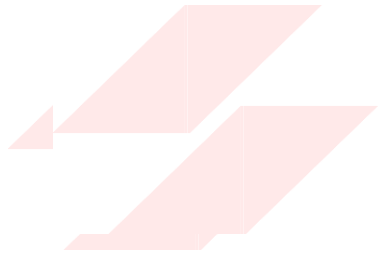
Not Missing At Random When the missing data has a structure to it, we cannot treat it as missing at random. In the above example, if the data was missing for all students from specific schools, then the data cannot be treated as MAR.

12. A/B testing is a experimentation process where two or more variants A and B are compared, in order to determine which variable is more effective.

13. In a single imputation method the missing data are filled by some means and the resulting completed data set is used for inference. Mean imputation is one such method in which the mean of the observed values for each variable is computed and the missing values for that variable are imputed by this mean.

14. Linear regression analysis is used to predict the value of a variable based on the value of another variable.

15. There are three real branches of statistics data collection descriptive statistics and inferential statistics



FLIP ROBO

