

Statistics Worksheet

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

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. It is a symmetrical, but not all symmetrical distributions are normal. In reality, most pricing distributions are not perfectly normal. In reality, most pricing distributions are not perfectly normal.

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

The most important missing data imputation techniques for handling missing data during prediction time are reduced feature models, distribution-based imputation, prediction value imputation.

12. What is A/B testing?

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

Statistics Worksheet

13. Is mean imputation of missing data acceptable practice?

It is acceptable when the missing value proportion is not large enough.

But, when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been.

14. What is linear regression in statistics?

The dependent and independent variables show a linear relationship between the slope and the intercept.

Its independent variable is not random and value of the residual (error) is zero.

The value of the residual (error) is constant across all observations and value of the residual (error) is not correlated across all observations.

15. What are the various branches of statistics?

There are two main branches of Statistics

1. Descriptive Statistics

Descriptive statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.

2. Inferential Statistics

Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.