

STATISTICS WORKSHEET-1

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

Ans: 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?

Ans: Central Limit Theorem

3) Which of the following is incorrect with respect to use of Poisson distribution?

Ans: Modeling bounded count data

4) Point out the correct statement

Ans: The square of a standard normal random variable follows what is called chi-squared distribution

5) _____ random variables are used to model rates.

Ans: Poisson

6) 10. Usually replacing the standard error by its estimated value does change the CLT.

Ans: False

7) 1. Which of the following testing is concerned with making decisions using data?

Ans: Hypothesis

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

Ans: 0

9) Which of the following statement is incorrect with respect to outliers?

Ans: Outliers cannot conform to the regression relationship

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

Ans: A random variable is said to be normally distributed if it follows this rule: the lesser its value differs from the mean value of the random variable, the higher is its probability. The mean hence has the highest probability. As a value moves away from the mean, its probability diminishes. Say a population of people is such that most 20 year old women are 5 feet in height. The average height of a 20 year old female is 5 feet. This means very few girls of 20 years are 6 feet or 4 feet in height. But plenty of girls will be 5 feet 1 inch or 4 feet 9 inches. We say that the height of 20 year old females in this population is normally distributed.

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

Ans) When dealing with missing data, there are a few common approaches. One option is to simply remove the observations that have missing values. Another approach is to fill in the missing values with estimated values based on the available data. For example, you can replace missing values with the mean or median of the other values in that variable. The specific technique used depends on the nature of the data and the goals of the analysis. It's important to choose an approach that minimizes bias and accurately represents the missing values.

12) What is A/B testing?

Ans: A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13) Is mean imputation of missing data acceptable practice?

Ans: The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score.

14) What is linear regression in statistics?

Ans: A linear regression model in context of machine learning/statistics is basically a linear approach for modelling the relationships between the dependent variable (known as the result) and your independent variable(s) (known as 'features').

15) What are the various branches of statistics?

Ans: The two main branches of statistics are descriptive and inferential statistics. They are both used as part of most statistical analyses. Descriptive statistics are first used to organise and present collected data into a coherent form that can be further analysed through inferential statistics.