

Statistics Worksheet-1

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

- a) True
- b) False

Ans-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
- b) Central Mean Theorem
- c) Centroid Limit Theorem
- d) All of the mentioned

Ans-A) Central Limit Theorem

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

- a) Modeling event/time data
- b) Modeling bounded count data
- c) Modeling contingency tables
- d) All of the mentioned

Ans-B) Modeling bound data

4. Point out the correct statement.

- a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
- b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
- c) The square of a standard normal random variable follows what is called chi-squared distribution
- d) All of the mentioned

Ans-D) All the mentioned

5. _____ random variables are used to model rates.

- a) Empirical

- b) Binomial**
- c) Poisson**
- d) All of the mentioned**

Ans-C) Poisson

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

- a) True**
- b) False**

Ans-B) False

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

- a) Probability**
- b) Hypothesis**
- c) Causal**
- d) None of the mentioned**

Ans -B) Hypothesis

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

- a) 0**
- b) 5**
- c) 1**
- d) 10**

Ans-A) 0

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

- a) Outliers can have varying degrees of influence**
- b) Outliers can be the result of spurious or real processes**
- c) Outliers cannot conform to the regression relationship**
- d) None of the mentioned**

Ans-C) Outliers cannot conform to the regression relationship.

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

Ans- **Normal Distribution** :- The normal distribution is the most widely known and used of all distribution . Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems.

Many things actually are normally distributed, or very close. For example, height and intelligence are approximately normally distributed measurement errors also often have normal distribution.

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

Ans- Missing data can be dealt with in a variety of ways. The most common reaction is to ignore it. The application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

Another common strategy is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

Single or Multiple Imputation

Single and multiple imputation are the two forms of imputation. When we say imputation, it usually is mean single.

The term "single" refers to the fact that we only use one of the seven methods to estimate the missing number.

It generates a sample with the same number of observations as the complete data set.

When listwise deletion eliminates a considerable amount of the data set, single imputation appears to be a good option.

Unless the data is Missing Completely at Random, certain imputation processes, such as means, correlations, and regression coefficients, result in skewed parameter estimations.

The level of the bias is determined by a number of factors, including the imputation technique, the missing data mechanism, the fraction of missing data, and the information in the data set.

Furthermore, standard errors are underestimated by all single imputation approaches.

Because the imputed observations are estimates, their values have a random error associated with them. However, your programme is unaware of this when you enter that estimate as a data point. As a result, it ignores the additional source of error, resulting in too-small standard errors and p-values. In multiple imputation, numerous estimates are generated, two of the approaches : hot deck and stochastic regression—work as the imputation method. The multiple estimates varied significantly because these two approaches contain a random component.

12. What is A/B testing?

Ans - A/B Testing (also known as Split testing) defines a way to compare two versions of an application or a web page that enables us to determine, which one performs better. A/B Testing is one ways, where one can modify an application or a web page to create a new version and then comparing both these versions to find the conversion rate. This also lets us know, which is the better performer of the two.

13. Is mean imputation of missing data acceptable practice?

Ans - It is a solution which preserves the mean of the preserve data. If the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

Ans— It is a statistical method that allows to summarize and study relationship between the two variables. One variable is considered to be an explanatory or independent variable, and the other is considered to be a dependent variable.

15. What are the various branches of statistics?

Ans- Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are two main branches of statistics

- Inferential Statistic.
- Descriptive Statistic.

Inferential Statistics:

Inferential statistics used to make inference and describe about the population. These stats are more useful when it is not easy or possible to examine each member of the population.

Descriptive Statistics:

Descriptive statistics are used to get a brief summary of data. You can have the summary of data in numerical or graphical form.