

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: a) Modeling event/time 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 of 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: The normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely.

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

Ans: Simple imputation

Knn imputation

Iterative imputation

12. What is A/B testing?

Ans: A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. ... A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants are more effective.

13. Is mean imputation of missing data acceptable practice?

Ans: Yes, it can be accepted if NaN values are lesser and it is a continuous data or else there will be a data leakage.

14. What is linear regression in statistics?

Ans: Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = c + b \cdot x$, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

15. What are the various branches of statistics?

Ans:

Branches of statistics

1. Descriptive statistics.

A descriptive statistic (in the count noun sense) is a summary statistic that quantitatively describes or summarizes features from a collection of information, while descriptive statistics (in the mass noun sense) is the process of using and analyzing those statistics.

2. Inferential statistics

While descriptive statistics summarize the characteristics of a data set, inferential statistics help you come to conclusions and make predictions based on your data.

When you have collected data from a sample, you can use inferential statistics to understand the larger population from which the sample is taken.