

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 bounded count 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

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

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?

The Normal distribution is a bell-shaped curve and majority data falls near to mean and spread of the data is decided by standard deviation.

Normal distribution mean is 0 and standard deviation is 1.

It is a symmetric, the left side and right side of the mean are equal.

In normal distribution, approximately 68% of data falls under standard deviation of 1, 95% of data falls under standard deviation of 2 and 99.7% of data falls under standard deviation of 3.

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

Handling the missing data is one of the important aspects of data preprocessing in Machine learning and statistical analysis.

Imputing missing values depends on the nature of the data and the reason for missingness.

One of the simplest methods to replace the values with Mean/median/Mode of the variable/column/feature. It is not suitable if the data has skewness.

13. Is mean imputation of missing data acceptable practice?

Yes, it is a simple and quick solution to handling the missing values. It depends on the nature of the variables/data.

14. What is linear regression in statistics?

The Linear regression is a statistical method to find the relationship between one dependent variable and one or more independent variables.

15. What are the various branches of statistics?

Statistics is broad field and it has own set of implementations.

Descriptive statistics: Used for summarize, organize, and present data. it measures the mean, median, mode, range, variance, and standard deviation

Inferential Statistics: Involves making inferences and predictions about a population based on a sample of data.

Probability: Helps to find likelihood of events occurring