

STATISTICS WORKSHEET-1

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.

- a) True
- b) False

Answer:

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

Answer:

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

Answer:

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

Answer:

Sorry sir, I am confused.

5. _____ random variables are used to model rates.

- a) Empirical
- b) Binomial
- c) Poisson
- d) All of the mentioned

Answer:

c) Poisson

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

- a) True
- b) False

Answer:

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

Answer:

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

Answer:

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

Answer:

d) None of the mentioned

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?

Answer:

The Normal Distribution, or Gaussian distribution, describes a bell-shaped curve where data is symmetrically distributed around the mean. It's used for modeling continuous variables, making predictions, and assumption testing, with properties like mean, variance, and standard deviation guiding statistical inference and algorithm performance.

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

Answer:

To handle missing data I would prefer to use techniques such as mean/mode imputation, median imputation or predictive modeling (e.g., k-NN, regression). For complex cases, I would consider advanced methods like multiple imputation or using algorithms robust to missing data. The choice depends on data patterns and the impact on analysis accuracy.

12. What is A/B testing?

Answer:

A/B testing is a method to compare two versions (A and B) of a variable to determine which performs better. By randomly assigning users to each version and analyzing their behavior, we can

identify which version leads to better outcomes, such as higher conversion rates or user engagement.

13. Is mean imputation of missing data acceptable practice?

Answer:

Mean imputation is a simple method for handling missing data by replacing missing values with the mean of observed values. While easy to implement, it can reduce variability and introduce bias, potentially skewing results. It is acceptable for small datasets or when other methods are impractical, but not always ideal.

14. What is linear regression in statistics?

Answer:

Linear regression in statistics models the relationship between a dependent variable and one or more independent variables by fitting a linear equation. It estimates how changes in independent variables impact the dependent variable, helping to predict outcomes and understand relationships in data.

15. What are the various branches of statistics?

Answer:

The main branches of statistics are:

- 1. *Descriptive Statistics:* Summarizes and describes data.**
- 2. *Inferential Statistics:* Makes predictions or inferences about a population based on sample data.**
- 3. *Predictive Analytics:* Uses data to make predictions about future events.**
- 4. *Bayesian Statistics:* Applies Bayes' theorem to update probabilities.**