

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

Correct Answer = 'A'

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

Correct Answer = 'A'

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

Correct Answer = 'A'

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

Correct Answer = 'B'

5. _____ random variables are used to model rates.

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

Correct Answer = 'C'

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

a) True
b) False

Correct Answer = 'B'

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

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

Correct Answer = 'B'

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

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

Correct Answer = 'A'

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

Correct Answer = 'D'

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 also called the Gaussian Distribution. In this, the mean, median and mode all line up in such a way that the center of distribution is the mean & it creates a bell shape diagram.

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

The best ways to handle missing data using 'Deletion method to eliminate missing data'.

Imputation techniques are: -

- Imputation with a constant value
- Imputation using the statistics (mean, median & mode)

12. What is A/B testing?

A/B testing is also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is not considered in practices since it ignores features of correlation. Also, it decreases the variance of our data while increasing bias. As a result of the reduced variance, the method is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Linear regression is used to predict the value of a variable based on the value of another variable.

15. What are the various branches of statistics?

There are two main branches of statistics are 'Descriptive Statistics' and 'Inferential Statistics'.

