

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

Ans: 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

Ans: a)

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

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

Ans: b)

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)

5. _____ random variables are used to model rates.

- a) Empirical

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

Ans: c)

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

- a) True
- b) False

Ans: a)

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

Ans: 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

Ans: 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

Ans: c)

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?

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

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?

14. What is linear regression in statistics?

15. What are the various branches of statistics?

Ans.10: whenever samples of any variables are distributed with normal range and the mean, median and mode will be the same with no outlier in the dataset. Data will be distributed in such a way where First quartile will have 25% of overall data, in second quartile will have 75% and in 3rd quartile will have 100% of data distribution.

Ans. 11: I would recommend to apply encoder method to handle missing data based on % of missing data. If very less amount like less than 5% of data are missing, then we can delete. Other methods are like using mean, mode and median to handle missing data.

Ans.12: It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. Will keep A scenario remain unchanged and make the changes in scenario B and then compare it which is performing better.

Ans.13: I feel its not a good idea always trusts on mean imputation. It totally depends on % of missing data and also depends on kind of data. Sometimes we miss the feature correlations.

Ans.14: The relationship between variables by fitting a line to the observed data. Linear regression models use a straight line. Regression allows you to estimate how a dependent variable changes as the independent variables change.

Ans.15:

Statistic:

Two types: 1- Descriptive Statistic, 2- Inferential Statistic

a) Descriptive Statistic:

Three types: Univariate, Bivariate, Multivariate

Univariate:

1. Frequency: - Freq. Table Histogram
2. Central Tendency: - Mean, Mode & Median
3. Dispersion: - Range (Min-Max), IQR, Std Deviation and Variance

Bivariate:

Covariance, Correlation

Multivariate:

Covariance Matrix, Correlation Matrix

b) Inferential Statistic:

1. Model Testing
2. Hypothesis Testing