Statistics

1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

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

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

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

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

a) True

b) False

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

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

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

10. What do you understand by the term Normal Distribution?

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution

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

Missing values can be handled by two categories.

•Deletion • Imputation

Deletion is the prevalent methods for dealing with missing data list-wise deletion method is commonly used. Int list-wise deletion we Can remove a record or observation in the data set easily

Eg:

Pandas have dropna function df. dropna (axis, inplace: True)

Handling missing data imputation:

In imputation method the missing data is filled with a substitute value. There are various methods for replacing the missing values with a specific value

Some of Methods are

MAR

MCAR

fillna method Can be used

df. fillna (inplace: True)

I would Recommend to Use bitray Value imputation as it can handle both Numerical and Categorical variables The missing values are group in a column and assign them to a new value not in the range of that column

12. What is A/B testing?

A/B testing, 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?

Yes mean imputation is acceptable for missing data. It is the good choice for series which randomly fluctuate. Around a certain value / level

14. What is linear regression in statistics?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable

15. What are the various branches of statistics?

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics