STATISTICS

ASSIGNMENT ; 1 part 2

Ques1 Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Answer…. TRUE

Ques2 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

Ques3 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

Ques4 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… D ALL ARE MENTIONED

Ques5 _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Answer… C POISSON

Ques6… 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Answer….False

Ques7… 1. 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

Ques8… 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

Answer.. A 0

Ques9…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… C

Ques10… What do you understand by the term Normal Distribution?

Answer… The normal distribution is a form presenting data by arranging the probability distribution of each value in the data.Most values remain around the mean value making the arrangement symmetric.

We use various functions in numpy library to mathematically calculate the values for a normal distribution…!!

Ques11… How do you handle missing data? What imputation techniques do you recommend?

Answer… We can **use pandas DataFrame method . dropna() to remove rows or records which contain atleast one missing value…**

The techniques are as follows….

- Missing data can be dealt with in a variety of ways. ...
- Another common strategy among those who pay attention is imputation. ...
- Mean imputation. ...
- Substitution. ...
- Hot deck imputation. ...
- Cold deck imputation. ...

Ques12….What is A/B testing?

Answer….A/B testing is a type of experiment in which you split your web traffic or user base into two groups, and show two different versions of a web page, app, email, and so on, with the goal of comparing the results to find the more successful version

Ques13… Is mean imputation of missing data acceptable practice?

Answer….Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Ques14…. What is linear regression in statistics?

Answer… 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.

Ques15….What are the various branches of statistics??

Answer….**data collection**

descriptive statistics

inferential statistics…

Inferential Statistics:

Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

Descriptive Statistics:

Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphycal form…