FLIP ROBO

ASSIGNMENT – 39

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

Ans-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

Ans- 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

Ans- a) Modeling event/time 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

ans- c) The square of a standard normal random variable follows what is called chi-squared distribution

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

ans- c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

ans- a) True

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) Hypothesis

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) 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

ans- d) None of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans – A normal distribution of data always tends to be symmetric to the mean and the distribution curve tends to be a bell shaped curve .

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

Ans- Missing data can be handled by either by dropping the specific row or by replacing it with the mean or median . although dropping is not recommended as it cause loss of data .

As per my view I would recommend simple imputer instead of using get dummies . as in simple imputer we can use the mean to replace Null values which will cause minimal deviation to the data set but get dummies will create high number of new columns to represent every variable and adding those many column can have a high changing effect on the data sets .

12. What is A/B testing?

Ans- A/B testing is a method of getting 2 or more random contents at the same time to determine which of them performs well and creates a better impact which would help the organization to grow better .

Example - getting 2 different UI of an online shopping app used and checked by a regular customer to get better insight of how each one of them is performing when it comes to the ease of shopping on that particular app .

13. Is mean imputation of missing data acceptable practice?

Ans – Mean imputation is kind of good solution as it remains unbiased towards any specific variable in the computation yet in case if he data is skewed , or the data is having high outliers or the number of missing values are high in a specific data set/variable then imputation of mean can have a great impact on the output and that will affect the efficiency of the model .
Note- As being unknown to the industry standard , Above Answer is given based on personal understand .

14. What is linear regression in statistics?

Ans – in Linear regression the model learns the data and analyses the intercept and coefficient of the data . based on the learned aspects of the data set it tries to predict the outcome of a new data that we assign to the model .

15. What are the various branches of statistics?

Ans- Descriptive and Inferential statics are the two branches of statistics .

In Descriptive statistics gives the insight about the data where as in inferential we make assumption with taking the insight from the sample of a population into consideration .