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

Answer: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

Answer: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

Answer:b)Modeling bounded count 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

Answer:d)All of the mentioned.

5. _____ random variables are used to model rates.

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

Answer:C)Poisson

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

a) True b) False

Answer: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

Answer: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

Answer:a)0

9. Which of the following statement is incorrect with respect to outliers?

a) Outliers can have varying degrees of influ

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) Outliers cannot conform to the regression relationship

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

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

Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.It is also called as Gaussian distribution.When represented in the graphical form it takes the shape of a bell,hence called as Bell curve.In normal distribution man=median=node=0,standard deviation=1 and it has zero skewness.Central limit theorem can actually help in understanding the Normal distribution.

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

Missing data can be handled by using the imputation techniques. Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data/information of the dataset.If the column in which the data is missing is integer then the missing data will be filled with the mean or mode of that column.If the column in which the data is missing is categorical then he mode will be filled for the missing data.

Numerical Variables: Mean/median imputation, Arbitrary value, End of tail, Mode imputation.

Categorical Variable: Frequent category imputation, adding a missing category

12. What is A/B testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

A/B testing has the below steps.

1.Make a Hypothesis: A hypothesis is a tentative insight into the natural world; a concept that is not yet verified but if true would explain certain facts or phenomena.

It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

we have to make two hypotheses i Null hypothesis and alternative hypothesis.

2.Create Control Group and Test Group: Once we are ready with our null and alternative hypothesis, the next step is to decide the group of customers that will participate in the test. Here we have two groups – The Control group, and the Test (variant) group.

3. Conduct the A/B Test and Collect the Data:

13. Is mean imputation of missing data acceptable practice?

Mean imputation does not preserve the relationships among variables so it is not a good solution.In this technique large part of data will be lost hence it is not a good solution.

14. What is linear regression in statistics?

Linear regression is a predictive analysis. These regressions are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Types of Linear regression:

Simple linear regression

Multiple linear regression

Logistic regression

Multinomial regression

15. What are the various branches of statistics?

The four branches of statistics are :

Mathematical statistics:It helps in forming the experimental and statistical distribution.

Statistical methods or functions: It helps in collection, tabulation and interpretation of the data .It helps in analysing the data and returns insights from data.

Descriptive statistics:It helps in summarizing and organizing any data set characteristics.It also helps in the representation of the data in both classification and disgramatic .

Inferential statistics:It helps in finding the conclusion regarding the population after analysis on the sample drawn from it.