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
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

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

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

Answer:
A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. Normal distributions are symmetrical, but not all symmetrical distributions are normal.

11. How do you handle missing data? What imputation techniques do you recommend?
Answer:
Missing data means, where the data values are missing for some of the attributes. For this few steps has followed for data repair. The steps are.
1. Implementing the given/required dataset.
2. Marking the missing data in the dataset.
3. Identifying the problems cause through missing data.
4. Removing the missing data from dataset
5. Imputation process for missing data
6. Going for suitable algorithm which support for missing data

Imputation techniques are used to impute the missing values to the data. The techniques list
1. Deductive imputation
2. Mean/Mode/Median imputation
3. Regression imputation
4. Stochastic Regression imputation
5. Multiply-Stochastic Regression imputation

Mean/Mode/Median imputation is best one for imputation purpose

12. What is A/B testing?
Answer:
A/B testing, is also refers as split testing, which promotes to a randomized experiment technique, in which two or more versions of a variables like web paging, page element etc... are shows to various segments of website viewers at the same time duration to Spotify which version leaves the maximum influence/impact and drive business metrics.

13. Is mean imputation of missing data acceptable practice?
Answer:
Yes., because, imputing the mean reserves in the mean of the observed data. For this, whenever the data are missing fully at random, the estimate of the mean remains same/unchanged. That's approach thing. On this factor, most researchers are like to adopt to this relationship among variables.
14. What is linear regression in statistics?

Answer:
In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). ... Linear regression has many practical uses.

15. What are the various branches of statistics?

Answers:

Statistics analysis has key role in the field of data research. It brings the path in the collection, analysis and presentation of data. In statistics analysis two branches are main.

They are.

1. Descriptive Statistics
2. Inferential Statistics

Descriptive Statistics - describes the important characteristics/properties of the data using the measures the central tendency like mean/median/mode and the measures of dispersion like range, Standard deviation, variance and so on. Data can be summarized and represented in an accurate way using charts, tables and graphs.

Inferential Statistics - about using data from sample and then making inferences about the larger population from which the sample is drawn. The goal of the inferential statistics is to draw conclusions from a sample and generalize them to the population. It determines the probability of the characteristics of the sample using probability theory. The most common methodologies used are hypothesis tests, Analysis of variance etc.