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 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]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?

Answer: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?

Answer:One way of handling missing values is the deletion of the rows or columns having null values.If any columns have more than half of the values as null then you can drop the entire column.In the same way,rows can also be dropped if having 1 or more column values as null. IMPUTATION TECHNIQUES are deleting the entire row,replacing with mean,replacing with an arbitrary value,replacing with the mode, replacing with the median,replacing with the previous calue.

12. 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 2 groups,and show 2 different version of a web page app,email,and so on,with the goal of comparing the results to find the more succesful version.A/B testing is essentially an experiment where 2 or more variants of a page are shown to users at random,and statistical analysis is used to determine which variation perform better for a given conversion goal.

13. Is mean imputation of missing data acceptable practice?

Answer:The process of replacing null values in a data collection with the data's mean is known as mean imputation.Mean imputation is typically considered terrible practice since it ignores feature correlation.Consider the following scenario:we have a table with age and fitness score,and a 8year old has a missing fitness score.If we average the fitness score of a people between the ages of 15 to 80,the 8 year old will appear to have a significantly greater fitness level than he actually does.Second,mean inputation 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.

14. 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.The variable you are using to predict the other variable's value is called independent variable.EX. The weight of the person is linearly related to their height so this shows a linear relationship between the height and weight of the person.According to this,as we increase the height,the weight of the person will also increase.

15. What are the various branches of statistics?

Answer:Statistics is a study of presentation,analysis,collection,interpretation and organization of data There are 2 main branches of statistics they are inferential and descriptive statistic

1] Inferential Statistics: It is 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.

2] Descriptive Statistics:It is use to get a brief summary of data.You can have the summary of data in numerical or graphycal form