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

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

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: B

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: D

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

Ans: C

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

Ans: B

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

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

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: C

WORKSHEET

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: Normal Distrubution,also known as the Gaussian Distribution,is a probability distribution that is symmetric about the mean,showing that data near the mean are more frequent in occurance than data far from the mean.in graph form,normal distribution will appear as a bell curve.the graph of the normal distribution is characterised by two parameters i.e., mean or average.A small standard deviation (compared with the mean)produces a steel graph,where as standard deviation produces a flat graph.

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

Ans: "Mean" will replace missing values using the mean in each other column.it is preferred if data is numeric and not skewed."Median will replace missing values using the median in each other column.It is preferred if data is numeric and skewed.Best techniques to handle missing data,use deletion methods to eliminate missing data.Use regression analysis to systematically eliminate data.
Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data/information of the dataset.

12. What is A/B testing?

Ans: Optimise your app experience through experimentation to engage users.Run app experiments to test ideas,&see impact on key metrics.Easy to set up & scale.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.

13. Is mean imputation of missing data acceptable practice?

Ans: True, imputing the mean preserves the mean of the observed data.so if the data are missing completely at random,the estimate of the mean remains unbiased.The process of replacing null values in a data collection with the data's mean is known as Mean Imputation.

14. What is linear regression in statistics?

Ans: Linear Regression analysis is used to predict the value of a variable basedon the value of another variable.The variable you want to predict is called the dependent variable.The variable you are using is called the Independent Variable.Linear Regression is the most basic and commonly used predictive analysis.Regression estimates are used to describe dataand to explain the relationship.

15. What are the various branches of statistics?

Ans: Statistics:

Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are two main branches of statictics

1.Inferential Statistic.
2.Descriptive Statistic.

Inferential Statistics:

Inferential statistics used to make inferenceand describe about the population.These stats are more useful when its not esy 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 numericalor graphical form.