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:- 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

Ans:- d) All of the mentioned

----

5. _____ random variables are used to model rates.

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

Ans:- c) Poisson

----

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

 a) True b) False

Ans:- b) False

----

7. 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. 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:- 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?

Ans :- 1) A normal distribution is an arrangement of a data set .

2) A graphical representation of a normal distribution is sometimes called a bell curve because of its flared shape. The precise shape can vary according to the distribution of the population but the peak is always in the middle and the curve is always symmetrical. In a normal distribution,

the mean, mode and median are all the same.

In skewed right mean>median>mode

In skewed left mean<median<mode

----

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

Ans:- a)Missing data occurs when no data is stored for the variable as an observation.

b) In a result, no information is provided for one or more items or for a whole unit.

--→ **Multiple imputation** is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values.

----

12. What is A/B testing?

Ans:- A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

----

13. Is mean imputation of missing data acceptable practice?

Ans:- **a)** In statistics, imputation is the process of replacing missing data with substituted values. When substituting for a data point, it is known as "unit imputation"; when substituting for a component of a data point, it is known as "item imputation".

b) Mean imputation is the practice of replacing null values in a data set with the mean of the data. Mean imputation is generally bad practice because it doesn't take into account feature correlation.

----

14. What is linear regression in statistics?

Ans:- 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). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

----

15. What are the various branches of statistics?

Ans:- **Descriptive Statistics**:- a) Descriptive statistics is the first part of statistics that deals with the collection of data.

b) Here you can describe the data.

c) Example:- Height of a person, how student in a class.

Inferential Statistics: - a) The inference statistics are techniques that enable statisticians to use the information.

b) From huge data we collect few data (random data) and can find the average or give conclusion.

c) Example: - Exit Poll, Average marks of many batches.