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

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

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

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

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

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

Ans. **normal distribution**, also called **Gaussian distribution**, the most common distribution function for independent, randomly generated variables. Its familiar bell-shaped curve is ubiquitous in statistical reports, from survey analysis and quality control to resource allocation.

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

Ans. 1 Mean or Median Imputation. When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. ...

2. Multivariate Imputation by Chained Equations (MICE) MICE assumes that the missing data are Missing at Random (MAR).

3. Random Forest.

3. What is A/B testing?

Ans. A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

4.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. That's a good thing.Plus, by imputing the mean, you are able to keep your sample size up to the full sample size. That's good too.

5.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).

5.What are the various branches of statistics?

Ans: There are three real branches of statistics: **data collection, descriptive statistics and inferential statistics**.

FLIP ROBO