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:
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:
a) Sums of normally distributed random variables are again normally distributed even if the variables are dependent.

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

Answer:
a) Poisson

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

   Answer:
   a) True

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

   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:

   b) Outliers can be the result of spurious or real processes

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

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

Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. The normal distribution appears as a "bell curve" when graphed.

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

Handling missing data:

Deletion

Remove rows or columns that contain missing values from your analysis. This is also known as "listwise deletion" or "complete case analysis". It's the most common method because it's easy to implement and is the default option in most statistical packages.

Imputation

Replace missing values with estimated values based on other information. Some imputation techniques. Replaces missing data with the most frequently occurring value for that variable.

Recommended imputation technique:

- Next or Previous Value
- K Nearest Neighbors
- Maximum or Minimum Value
- Missing Value Prediction
- Most Frequent Value
- (Rounded) Mean or Moving Average or Median Value

12. What is A/B testing?

A method for comparing two versions of a product or website to determine which one performs better. A/B testing can help businesses get better ROI from existing traffic and increase conversions without spending more money on acquiring new traffic.

13. Is mean imputation of missing data acceptable practice?

No, mean imputation is not considered an acceptable practice for missing data, because of the following reasons.

- Biased estimates: Mean imputation can lead to biased estimates of variances and covariances.
- Underestimates: Mean imputation can underestimate standard errors, which can lead to Type I errors.
- Reduces variance: Mean imputation can reduce the variance of a variable, which can affect covariances with other variables.
- Doesn't preserve relationships: Mean imputation doesn't preserve relationships among variables.
- Alternatives: There are many alternatives to mean imputation that provide more accurate estimates and standard errors.

14. What is linear regression in statistics?

In statistics, linear regression is a statistical model that estimates the linear relationship between a scalar response (dependent variable) and one or more explanatory variables (regressor or independent variable).

15. What are the various branches of statistics?

The two main branches of statistics are descriptive statistics and inferential statistics: