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
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
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
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
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   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
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
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

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---

Normal distribution, also known as the Gaussian distribution. The mean, median, and mode are equal. It 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. In graph form, normal distribution will appear as a bell curve.

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

Answer---

I can handle missing data by using

1) Deleting Rows with missing values.
2) Impute missing values for continuous variable.
3) Impute missing values for categorical variable.
4) Using Algorithms that support missing values.
5) Prediction of missing values.

The following are techniques of imputation:

1) Mean imputation. Simply calculate the mean of the observed values for that variable column for all individuals who are non-missing. ...
2) Substitution of the mean
3) Regression imputation.
4) Stochastic regression imputation.
5) Interpolation and extrapolation.

12. What is A/B testing?

Answer---

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. It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Answer---

Yes, 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. But it leads to an underestimate of the standard deviation and distorts relationships between variables by “pulling” estimates of the correlation toward zero.

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 one wants to predict is called the dependent variable. The variable one is using to predict the other variable's value is called the independent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.

15. What are the various branches of statistics?
Answer---

There are three real branches of statistics:

Data collection,
Descriptive statistics and
Inferential statistics.