

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

Q10 and 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 distribution, 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 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?

Ans:- Best techniques to handle missing data

1. Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.
2. Use regression analysis to systematically eliminate data.
3. Data scientists can use data imputation techniques

12. What is A/B testing?

Ans:-

An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

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. since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

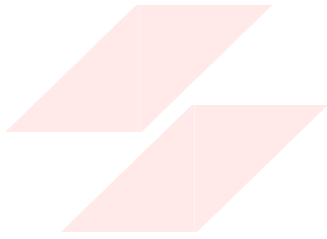
Ans:-

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). Linear regression has many practical uses.

15. What are the various branches of statistics?

Ans:-

The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypotheses and draw conclusions.



FLIP ROBO