

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 ANSWER
 - 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 ANSWER
 - 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 ANSWER
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 ANSWER
 - d) All of the mentioned
5. _____ random variables are used to model rates.
 - a) Empirical
 - b) Binomial
 - c) Poisson ANSWER
 - d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
 - a) True
 - b) False ANSWER
7. 1. Which of the following testing is concerned with making decisions using data?
 - a) Probability
 - b) Hypothesis ANSWER
 - 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 ANSWER
 - 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 ANSWER
 - 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? Normal distribution, also known as Gaussian distribution, is a continuous probability distribution that is widely used in statistical inference and modeling. It is a bell-shaped distribution that is symmetric around the mean, and is characterized by its mean and standard deviation
11. How do you handle missing data? What imputation techniques do you recommend? Handling missing data is a common problem in data analysis and can have a significant impact on the results and conclusions drawn from the data. There are several approaches to handling missing data, including deletion, imputation, and modeling-based methods.
12. What is A/B testing? A/B testing, also known as split testing, is a method of comparing two versions of a product or marketing campaign to determine which one performs better. The A/B test involves randomly assigning users to one of two groups, A and B, and presenting each group with a different version of the product or campaign.
13. Is mean imputation of missing data acceptable practice? Mean imputation is a simple method for handling missing data that involves replacing missing values with the mean of the observed values for that variable. While mean imputation is widely used due to its simplicity and ease of implementation, it has several limitations and may not always be an acceptable practice.
14. What is linear regression in statistics? Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It is a parametric approach that assumes that the relationship between the variables can be approximated by a linear function.
15. What are the various branches of statistics? Statistics is a broad field that encompasses many different branches, each with its own focus and applications. Here are some of the main branches of statistics: Descriptive statistics: Inferential statistics: Probability theory: Mathematical statistics: Bayesian statistics: Bio statistics: Econometrics: Data mining: Statistical computing: