

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
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10. Normal Distribution

Normal distribution is also called bell curve it is the probability distribution which is symmetric and the mean, median and mode will be the same and it is also used to check for symmetry of data example appraisal calculations

11. Missing values can be handled as per the situation, if the data are missing from a huge set of data set then the entire event can be erased however now every data is important so we would approximate the numerical missing value using averaging techniques like mean/mode/median imputation, arbitrary value imputation. Categorical variable Frequency category imputation, adding missing category . If both then we could use complete case analysis, adding a missing indicator, random sample imputation.

12. It is also known as split testing and is an experiment on two variants to see which variant performs better based on given metric.

13. Yes it is an acceptable practice and is used for numerical missing data

14. Linear regression is used as a predicting technique used for features that are numerical value and is a supervised machine learning technique. It is the process of finding the best fit to the data points available on the plot. We use linear regression when we want to find out the relationship between two variables.

15. The branches of statistics are data collection, descriptive statistics and inferential statistics.