

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

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= 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= All of the mentioned

5. _____ random variables are used to model rates.

- a) Empirical
- b) Binomial
- c) Poisson
- d) All of the mentioned

Answer= Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

- a) True
- b) False

Answer= 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

Answer= 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

Answer=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= Outliers cannot conform to the regression relationship

WORKSHEET

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?

Answer= normal distribution is a type of continuous probability distribution in which most of the are in middle curve and rest data are in the last down curve, the centre of the curve goes they tell u the average measurement. how wide the curve should be they tell u the standard deviation of the measurement and this is super useful for statistics. the middle of the range is known as mean of the distribution, in normal distribution id data are in under the range of +3 to -3 so that data is normal distribution if data are out of +3 or -3 so that a type of data is outliers

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

Answer= one way of handling missing value is In any data if one or two values are missing so you can drop it they don't effect the data if more than two or more values are null in column so drop is not correct in place of drop we can replace this null value with mean or median of the column if the data of that column in int or numbers or if the data of that column in string value so we can change it into mode value of that column, so that technique we can use to handle missing data

12. What is A/B testing?

Answer= AB testing is a type of experiment this is help to split your data or web traffic into two groups, and show two different version of web page, app, email with the goal of comparing the result to find the more successful version, AB testing is a form of statistical hypothesis testing or a significance test

13. Is mean imputation of missing data acceptable practice?

Answer= NO

14. What is linear regression in statistics?

Answer= Linear regression uses one independent variable to explain the outcome of the dependent variable, basically linear regression is used to predict income, education level, demand, sales and interest rates, linear regression is used for that type of data in which the TARGET COLUMN or the Y COLUMN in continuous form in that place we can use linear regression

15. What are the various branches of statistics?/

Answer= the various branches of statistics are:-

Descriptive statistic AND Inferential statistic AND Data Collection

In descriptive statistic are two branches

Central tendency OR Dispersion of data

In Inferential Statistic

ZScores, hypothesis testing , Ttest , chisquare test , regtest , Anova