

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

Ans: a) 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

Ans: a) Central Limit Theorem

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

Ans: b) 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

Ans: d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans: c) Poisson

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

a) True

b) False

Ans: b) False

7. Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

Ans: b) Hypothesis

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

- a) 0
- b) 5
- c) 1
- d) 10

Ans: a) 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

Ans: c) Outliers cannot conform to the regression relationship

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: A normal distribution is a bell-shaped frequency distribution curve. It shows that the data that tend to gather around the mean have more frequent occurrences than the data that are far from the mean. The further away from the mean a data point is, the less likely it is to occur.

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

Ans: We can handle missing data by:

- Drop the missing values
- Replacing the missing values
- Leave it as missing data

There are different imputation techniques to handle missing data depending on time series problem or general problem are:

- Linear Interpolation
- Make NA as level, Multiple Imputation, Logistic regression
- Mean, Median, Mode, Random Sample Imputation
- Mean, Median, Mode, Random Linear Regression

12. What is A/B testing?

Ans: It is a basic randomized control experiment. Basically, a method of comparing two versions of a variable (Here it is A and B) in a controlled environment to see which performs better.

For example: A company sells caps, then it will make two versions of cap.

A = red colored caps

B = blue colored caps

Then on the basis of response from the sample who purchased A and B respectively, we try to find which version is performing better.

It is a hypothetical testing process for making decisions based on sample statistics to estimate population parameters. The population refers to the total number of customers who have purchased your goods, whereas the sample refers to the number of customers who have taken part in the test.

13. Is mean imputation of missing data acceptable practice?

Ans: No, because we lose the essence of the relationship between variables and it may bias the standard error by decreasing the variance of data.

14. What is linear regression in statistics?

Ans: Linear regression is a type of regression used for predictive analysis. In statistics, it is a linear approach used for modelling the relationship between a dependent variable and one or more independent/explanatory variable.

The linear regression line has an equation of the form $Y = A + BX$, where X is the independent variable and Y is the dependent variable.

15. What are the various branches of statistics?

Ans: There are two branches:

1. Descriptive statistics
2. Inferential statistics

