

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 distribution

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

- a) True
- b) False

Ans = 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

Ans = b) 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

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 = Normal distribution is similar to a bell graph which denotes that most of the data will remain close to the mean. The data is normally distributed around the mean.

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

Ans = We can either delete the missing values or impute with some technical guess.

Would recommend replacing the null values with mean or median.

12. What is A/B testing?

A/B testing is an initial hypothesis approach where two or more conditions are tested with users interest. It helps in eliminating the less interesting feature/data.

13. Is mean imputation of missing data acceptable practice?

Ans = If the missing data is in huge proportion, the mean might get effected and might result in an inappropriate model.

14. What is linear regression in statistics?

Ans = A linear regression model is used to find the relationship between a dependent and independent variable by a straight line graph.

15. What are the various branches of statistics?

Ans = Descriptive statistics (describes the properties of sample data) and Inferential statistic (uses the properties of the data to drive hypothesis and results)