

## WORKSHEET

## 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

WORKSHEET

**10 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 referred to as the Gaussian distribution, is a fundamental concept in statistics and probability theory. It describes a specific type of continuous probability distribution where the majority of the data points settle towards the centre of the bell-shaped curve. This distribution is highly symmetrical and is defined by the mean and the standard deviation

11. How do you handle missing data?

There are different ways of handling missing value

- 1) Dropping
- 2) Imputation

What imputation techniques do you recommend?

My recommendation will be based on the type of dataset. Some of the imputation techniques include

- 1) Mean, Median, or Mode imputation
- 2) K-Nearest
- 3) Interpolation and extrapolation

12. What is A/B testing?

A/B testing, also known as split testing, is a controlled experiment conducted to compare two versions of a webpage, app, marketing campaign, or any other digital asset. The goal of A/B testing is to determine which version (A or B) performs better

in terms of a desired outcome, such as click-through rates, conversion rates, user engagement, or any other relevant metric.

13. Is mean imputation of missing data acceptable practice?

Yes

14. What is linear regression in statistics?

Linear regression is a statistical method used to model the relationship between a dependent variable (also known as the response variable) and one or more independent variables (also known as predictor variables or features). The goal of linear regression is to find the best-fitting linear relationship that can predict the value of the dependent variable based on the values of the independent variables.

15. What are the various branches of statistics?

There are two main types of statistics

- 1) Descriptive
- 2) Inferential