

## 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: (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

Answer: (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

Answer: (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

Answer: (c) The square of a standard normal random variable follows what is called chi-squared distribution

5. \_\_\_\_\_ random variables are used to model rates.

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

Answer: (c) Poisson

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

- a) True
- b) False

Answer: (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

Answer: (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

Answer: (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

Answer: (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?

Answer: The term Normal Distribution relates to a statistical concept which represents a probability distribution of variables. It is also known as the Gaussian distribution. It can be determined by a bell-shaped curve in statistical reports.

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

Answer: Basically we can handle missing data with two ways i.e. Imputation or the removal of data. Considering imputation method as the most useful one when the percentage of missing data is low; however, when the percentage of missing data is high, we can use the removal of data method.

I would highly recommend Mean/Median/Mode imputation. In this technique, missing values are replaced with mean, median, or mode values of the variable.

12. What is A/B testing?

Answer: A/B testing is also called as split testing it used to randomize two or more variables i.e. webpage or page element. Its spiltting the users in two groups, one group exposed to vertion A for control and the other group exposed to version B for variation. A/B testing being used to optimize design,emails, advertisements and and user interfaces.

13.Is mean imputation of missing data acceptable practice?

Answer: Yes it is acceptable, Mean imputation is a widely known method for handling missing data and mostly used in practice. It is easy to implement and have ability to retain the original size of data. However, mean imputation also has some drawbacks. It assumes that the missing values are missing randomly or do not have any systematic pattern.

14.What is linear regression in statistics?

Answer: Linear regression is a model used to understand and predict the difference between response variable and predictor variable. It indicate the relationship between two variables. The role of linear regression is to find out the proper fitting line that reduce the difference between predicted values and observed values.

15.What are the various branches of statistics?

Answer: There are multiple brances for statistics however , there are two main branches for statistics Descriptive Statistics and Inferential Statistics.

Descriptive Statistics: Descriptive statistics involve summarizing and describing data using measures such as mean, median, mode, and graphical representations.

Inferential Statistics: Inferential statistics involve drawing conclusions and making predictions about a population based on a sample. This branch includes hypothesis testing, confidence intervals, and regression analysis.

---