**STATISTICS WORKSHEET-1**

**---------------Solutions---------------**

1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Solution 1- 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

Solution 2- 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

Solution 3- 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

Solution 4- 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

Solution 5- c) Poisson

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

a) True

b) False

Solution 6- 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

Solution 7- 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

Solution 8- 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

Solution 9- c) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

Solution 10- The normal distribution, also known as the Gaussian distribution or bell curve, is a symmetric probability distribution. It is characterized by its shape, which is a symmetric, bell-shaped curve centered around its mean. The normal distribution is fully described by its mean and standard deviation.

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

Solution 11- Handling missing data involves techniques such as deletion (removing incomplete entries), imputation (replacing missing values), and special handling methods (creating separate categories or using domain knowledge). The choice of method depends on the dataset and the analysis goals. Common imputation methods include mean, median, mode imputation, regression imputation, and K-nearest neighbours imputation.

12. What is A/B testing?

Solution 12- A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Solution 13-Mean Imputation can be acceptable for simple cases, but may bias results or underestimate variance.

14. What is linear regression in statistics?

Solution 14- Linear regression is finding the best line between two variables. It tells you how one variable (X) affects another (Y) like predicting sales from ads.

15. What are the various branches of statistics?

Solution 15- The various branches of statistics are:-

Descriptive: Summarize data (mean, median, etc.) like drawing a map of your dataset.

Inferential: Draw conclusions about whole groups based on samples, like predicting future sales from past data.