**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

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

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

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:-C

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

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans:-C

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

a) True

b) False

Ans:-B

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

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

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

**Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

Ans:- The normal distribution is also known as a Gaussian distribution or probability bell curve.it is symmetric about mean and indicates that values near the mean occur more frequently rhan the values that are father away from the mean.

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

Ans:- Handling missing data is a crucial aspect of data preprocessing. The choice of imputation techniques depends on the nature of the data and the reasons for missing values. Here are some common imputation techniques:

**Type of Imputation:-**

**1.**Mean or median imputation

**2.**Mode imputation

**3.**Forward Fill or Backward Fill

**4.**Linear Regression imputation

**5.**K-nearest neighbors (KNN) imputation

**6.**Multiple imputation

**7.**Data Augmentation

12. What is A/B testing?

Ans:- A/B testing, also known as split testing, is a statistical method used in marketing, product development, and other fields to compare two versions of a variable to determine which performs better. The "A" and "B" in A/B testing refer to two different versions of something, whether it's a webpage, email campaign, mobile app, or other elements. The goal of A/B testing is to identify changes that increase or maximize a specific outcome, often referred to as a key performance indicator (KPI).

Here's a basic overview of the A/B testing process:

**1. Selection of Variations:**

Choose the element you want to the test(e.g., a webpage, email design, ad copy).

Create Two or more Version of that elements each with a distinct change (A, B, C, etc.).

**2. Random Assignment:**

**Randomly assign user or participants to each version.this help ensure that the group are**

comparable, and any differences in outcomes are likely due to the changes being tested.

**3. Measurement of Key Metrics:**

**Define key matrics or KPIs that you want to measure** (e.g., click-through rates, conversion rates, revenue).

Collect data on how each version performs in terms of these metrics.

**4. Statistical Analysis:**

**Use statistical methods to analyze the data determine if the difference of the observed are statistical significants**

13. Is mean imputation of missing data acceptable practice?

Ans:- Mean imputation is a common method for handling missing data, where missing values are replaced with the mean of the observed values for that variable. While mean imputation is a simple and quick way to address missing data, there are both advantages and disadvantages to consider:

14. What is linear regression in statistics?

Ans:- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The goal of linear regression is to find the best-fitting straight line (or hyperplane, in the case of multiple independent variables) that minimizes the sum of the squared differences between the observed and predicted values of the dependent variable.

1. **Simple Linear Regression:** Involves one independent variable.
2. **Multiple Linear Regression:** Involves two or more independent variables.

15. What are the various branches of statistics?

Ans:- Statistics is a broad field that encompasses various branches, each focusing on different aspects of data analysis, interpretation, and application. Here are some key branches of statistics:

**Types of Branches of Statistics**

1. Descriptive Statistics:
2. Inferential Statistics:
3. **Biostatistics:**
4. **Econometrics:**
5. **Psychological Statistics:**
6. **Social Statistics:**
7. **Business Statistics:**
8. **Environmental Statistics:**
9. **Engineering Statistics:**
10. **Spatial Statistics:**
11. **Bayesian Statistics:**
12. Machine Learning and Data Mining: