**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

**Answer-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

**Answer-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

**Answer-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

**Answer-D**

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

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

**Answer-C**

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

a) True

b) False

**Answer-A**

**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

**Answer-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

**Answer-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

**Answer-C**

**WORKSHEET**

**Q10and 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 normal distribution is often called the bell curve because the graph of its probability density looks like a bell. It is also known as called Gaussian distribution, after the German mathematician Carl Gauss who first described it.

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

**Answer-** Seven Ways to Handle Missing Values in Machine Learning:

1. Deleting Rows with missing values
2. Impute missing values for continuous variable
3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library

**12. What is A/B testing?**

**Answer-** A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better.

**13. Is mean imputation of missing data acceptable practice?**

**Answer-** True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing.

**14. What is linear regression in statistics?**

**Answer-** Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables. Linear regression looks at various data points and plots a trend line.

**15. What are the various branches of statistics?**

**Answer-** Statistics has two main branches namely Descriptive statistics and Inferential or predictive statistics.

1. **Descriptive statistics:** This is a branch of statistics which deals with methods of collection of data, its presentation and organization in various forms, such as distribution tables, graphs (e.g., ogive, Lorenz curves, etc.), diagrams (e.g., pie charts) and finding measures of central tendency and measures of dispersion or spread which are used in the description of data.
2. **Inferential or predictive statistics:** This is a branch of statistics which deals with techniques used for analysis of data, making estimates that lead to predictions and drawing conclusions or inferences from limited information taken on sample basis and testing the reliability of the estimates or predictions.