**STATISTICS WORKSHEET-1**

Answers: -

1. A) True
2. A) Central Limit Theorem
3. B) Modeling bounded count data
4. C) The square of a standard normal random variable follows what is called chi-squared distribution
5. C) Poisson
6. A) True
7. B) Hypothesis Testing
8. A) 0
9. C) Outliers cannot conform to the regression relationship is incorrect.

10. Normal distribution is a way that data is often spread out, where most of the data is around the average value and less data is farther away from the average. It's shown as a bell shape curve and is usually described by two numbers: the average and the amount of spread. It's also known as Gaussian distribution or bell curve.
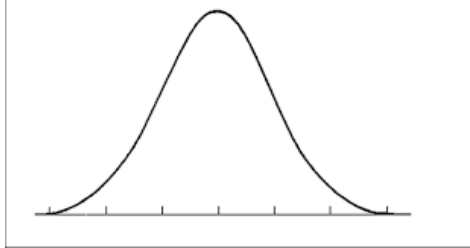
11. Handling missing data is important because it can affect the accuracy of the analysis. There are several ways to handle missing data:

    i.   Deleting the rows with missing data: This method is known as listwise deletion. It's simple but it can lead to a loss of information.
    ii.   Imputation: This method involves replacing the missing data with some estimated values. There are several imputation techniques such as mean imputation, median imputation, and multiple imputation.

           Recommended imputation techniques:

           a. Mean imputation: It's replacing the missing value with the mean value of the variable for all non-missing observations. It's simple but can be problematic if missing data is not random.
           b. Median imputation: It's replacing the missing value with the median value of the variable for all non-missing observations.
           c. Multiple imputation: Multiple imputation is a technique used to handle missing data where multiple sets of data are created with estimated values for the missing data, and each set is analysed separately. The results are then combined to give a more accurate overall result.

12. A/B testing is a way of comparing two versions of something, like a website or a product, to see which one performs better. It involves showing one version to one group of people and a different version to another group, and then comparing the results to see which version is more effective.

It is commonly used to make decisions on website design, product development and marketing strategy.

13. Linear regression is a way to predict a variable using a straight-line formula. It helps to find the best fit line between one or more independent variables and a dependent variable. It is used to make predictions about a variable that can take any value based on one or more input variable.
It is used for simple and multiple linear regression (more than one variable) to minimize the difference between predicted values and actual values.

Linear regression is represented by an equation of the form **y = a + bx**.

y = the dependent variable

x = the independent variable

a = the y-intercept

b = the slope of the line.

14. Following are the few of many branches of statistics:

i. **Descriptive Statistics:** This branch deals with the collection, organization, and summarization of data.
ii. **Inferential Statistics:** This branch deals with making predictions or inferences about a population based on a sample of data.
iii. **Probability:** This branch deals with the study of random events and the likelihood of their occurrence.
iv. **Statistical Inference:** This branch deals with the process of drawing conclusions about a population based on sample data.
v. **Time Series Analysis:** This branch deals with the analysis of data collected over time, such as stock prices or weather patterns.
vi. **Multivariate Statistics:** This branch deals with analysing more than one variable at a time.