STATISTICS WORKSHEET-1
Ques 1) - A (True)
Ques 2) - A (Central Limit Theorem)Ques 3) - B (Modeling bounded count data)
Ques 4) - D (All of the mentioned)
Ques 5) - C (Poisson)
Ques 6) - A (True)
Ques 7) - B (Hypothesis)Ques 8) - A (0)
Ques 9)- C (Outliers cannot conform to the regression relationship)
Ques 10)In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, it follows a bell shape, with most values clustering around a central region and the values decreasing as they go further away from the center.
Ques 11) We can deal with the missing data by Imputation which includes the substitution of an estimated value for the missing values and analyzing the entire data set, as if the imputed values were close to the observed values. We recommend using the mean/mode method as an imputation techniques.
Ques 12)A/B testing refers to the experiments where two or more variations of the same webpage are compared against each other by displaying them to real-time visitors to determine which one performs better for a given goal.
Ques 13) It is acceptable when the missing value proportion is not large enough.
Ques 14) Linear Regression: Linear regression fits a straight line and is commonly used in predictive analysis. Linear regression showcases the relationships between one explanatory variable and an outcome variable.
Question 15) The two types of statistics are: -Inferential Statistics -Descriptive Statistics

MACHINE LEARNING

Ques1) b) 4

Ques 2) c) 2 and 4

Quest 3) - d) formulating the clustering problem

ques 4) - a) Euclidean distance

Ques 5)- c) Agglomerative clustering

Ques 6) - d) All answers are correct

Ques 7) - a) Divide the data points into groups

Ques 8) - b) Unsupervised learning

Ques 9) - a) K- Means clustering

Ques 10) - a) K-means clustering algorithm

Ques 11) - d) All of the above

Ques 12 - d) Categorical data

Ques 13)
The hierarchical cluster analysis follows three basic steps:
1) calculation the distance
2) linking the cluster
3) choosing a solution by selecting the right number of clusters.

Ques 14)
In order to measure the quality of a clustering, we can use the average silhouette coefficient value of all objects.

Ques 15) Cluster analysis is a technque of data analysis that understand the naturally occurring groups within a data set known as clusters.
The type of clusters are Hierarchical Clustering and Partitional Clustering