STATISTICS WORKSHEET

1 – [A]

2 – [A]

3 – [B]

4 – [D]

5 – [C]

6 – [B]

7 – [B]

8 – [A]

9 – [C]

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

Answers: Normal Distribution is a probability distribution that is symmetric about the mean, showing that data near the mean more frequently in the occurrence than data far from the mean.
It is easy to work with mathematically, the methods developed using normal theory work quite well even when distribution is not normal.

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

Answer: Missing data define as values is not available, to handle the missing data we used the imputation techniques, most common techniques are: -
   a) Random Forest
   b) K-NN (K Nearest Neighbour)
   c) Mean and Medium
A common technique is to use the mean or medium of the non- missing observations. This can be useful in cases where the number of missing observations is low.
However, for large number of missing values, using mean or median can result in loss of variation in data and it better to use k-NN (K Nearest Neighbour) techniques.

12. What is A/B testing?
Answer: A/B testing is a basic randomized control experiment, it is a way to compare the two versions of variable to find out which performs better in a controlled environment.
A/B testing is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

13. Is mean imputation of missing data acceptable practice?

Answers: Mean reduces a variance of the data, a smaller variance leads to the narrower confidence interval in the probability distribution.
Mean imputation of missing data acceptable practice only where the number of missing observations is low.

14. What is linear regression in statistics?

Answer: Linear regression model the relationship between two variables by fitting a linear equation to observed data. One variable considered to be an explanatory variable, other is considered to be a dependent variable.
A Linear Regression equation is: $Y = a + bX$
Where **X** is explanatory variable
and, **Y** is dependent variable
the slope of the line is **b**, and **a** is the intercept (the value of y when x=0)

15.What are the various branches of statistics?

Answer: There are two branches of statistics: -

1. **Descriptive statistics:** Descriptive statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.
   For example- Industrial statistics, population statistics, trade statistics, etc. Businessmen make use of descriptive statistics in presenting their annual reports, final accounts, and bank statements.
2. **Inferential statistics:** Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates. For example - We take a sample from the population and find the proportion of illiterate in individuals in the sample.