STATISTICS WORKSHEET-1

1)A

2)A

3)B

4)D

5)C

6)B

7)B

8)A

9)C

10)Normal Distribution also known as Gaussian Distribution, is the most significant continuous probability Distribution It refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.

11) Missing data are data points that are absent for a specific variable in a dataset. They can be represented in various ways, such as blank cells, null values, or special symbols like “NA” or “unknown.” These missing data points pose a significant challenge in data analysis and can lead to inaccurate or biased results.
To handle missing data we can use any of the below imputation method :-

A)Mean Imputation

b) Median Imputation

c)Mode Imputation

d)Maximum Likelihood Imputation

E)Multiple Imputation

f)Regression Imputation

g)Pattern based Imputation

Imputation depends on what type of data are missing and accordingly we need to decide which one has to be followed.

Type of missing data	Imputation method
Missing Completely At Random	Mean, Median, Mode, or any other imputation method
Missing At Random	Multiple imputation, Regression imputation
Missing Not At Random	Pattern Substitution, Maximum Likelihood estimation

12) A/B testing is a powerful experimental method used to evaluate the effectiveness of changes to a system. It involves dividing users into two or more groups, exposing each group to different variations of a product, feature, or intervention, and then measuring the impact of these variations on user behaviour or outcomes. It helps data-driven organizations make informed decisions about product improvements, user experience enhancements, and overall business strategies based on empirical evidence rather than intuition or assumptions.

13) While mean imputation is a quick and easy method, it is generally recommended to use it with caution. For a more accurate and robust approach, especially with larger amounts of missing data or when the missingness might be related to other variables, exploring more advanced imputation techniques is advisable.

14) Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15) The two main branches of statistics are descriptive statistics and inferential statistics.