STATISTICS ASSIGNMENT- 1

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. Normal Distribution: - The normal distribution, also known as the Gaussian or standard normal distribution. It is the probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.
    It plots all of its values in a symmetric fashion, and most of the results are situated around probability's mean. Values are equally likely to plot either above or below the mean.
    It describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques.
11. The best way to handle missing data is:
    1. Use deletion methods to eliminate missing data. It works for certain datasets where participants have missing fields.
    2. Use regression analysis to systematically eliminate data.
    3. Data scientists can use data imputation techniques.

       The imputation techniques which we recommend are:
       1. Complete Case Analysis (CCA)
       2. Arbitrary Value Imputation.
       3. Frequent Category Imputation.

12. A/B Testing is basically statistical hypothesis testing, or statistical inference.
    It is an analytical method for making decisions that estimates population parameters based on sample statistics.
    An A/B test is a process where 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. Yes, mean imputation of missing data is acceptable practice because imputing the mean preserves the mean of the observed data. Thus, if data are missing completely at random, the estimate of the mean remains unbiased.
14. Linear regression analysis is used to predict the value of a variable on the value of another variable. The variable you want to predict is called the dependent variable

where as the variable you are using to predict the other variable's value is called the independent variable.

15. There are two branches of statistics:
    1. Descriptive Statistics: - It is used to present the data in an understandable way, so that a meaningful description can be made. This branch of statistics deals with methods of collection of data, its presentation and organization in various forms, such as distribution tables, graphs, diagrams and finding measures of central tendency and measures of dispersed or spread which are used in the description of data.
       Managers, CEOs etc. make use of descriptive statistics in presenting their annual reports, financial accounts and bank statements.

    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.
       It is used to make comparisons or predictions about a larger group known as population using information gathered about a small part of that population called a sample.