STATISTICS WORKSHEET – 1

1. Question -1 - Answer - **A**
2. Question -2 - Answer - **A**
3. Question -3 - Answer - **B**
4. Question -4 - Answer - **D**
5. Question -5 - Answer **- C**
6. Question -6 - Answer - **B**
7. Question -7 - Answer - **B**
8. Question -8 - Answer - **A**
9. Question -9 - Answer – **C**

10. The Normal distribution, also known as standard normal distribution, is the probability distribution that plots all if its values in s symmetrical fashion and most of the results are situated around the probability's mean. Values are equally likely to plot either above or below the mean.

The normal distribution is a probability distribution that 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.

Nowadays, it is more common to show up as a model for the "lifespan" of a product, like a light hub, or the outcome of standardized tests, like IQ. Biological measurements like height and weight.

11. Missing data can saved from mishandling in the following ways:

Missing data reduces the statistical power of the analysis, which can distort the validity of the results,

One way of handling missing values is the deletion of the rows or columns having null values.

If any columns have more than half of the values as null then you can drop the entire column.

In the same way, rows can also be dropped if having one or more columns values as null.

These are the following Imputation techniques:

1. **Complete Case Analysis(CCA):-**

This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.

This method is also popularly known as “Listwise deletion”.

2. **Arbitrary Value Imputation**

This is an important technique used in Imputation as it can handle both the Numerical and Categorical variables.

This technique states that we group the missing values in a column and assign them to a new value that is far away from the range of that column.

3. **Frequent Category Imputation**

This technique says to replace the missing value with the variable with the highest frequency or in simple words replacing the values with the Mode of that column.

12. A/B Testing :

A/B testing is a method of comparing two versions of a WebPages or app against each other to determine which one performs better. B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

A/B testing isn't a new idea with the advent of digital marketing. At one time, direct mail was the master of "splitting" or "bucketing" offers to see which one worked best. Digital capabilities build on the same idea but enable more specific, reliable, and faster test results.

For example, you might send two versions of an email to your customer list and figure out which one generated more sales. Then you can just send out the winning version next time. Or you might test two versions of ad copy and see which one converts visitors more often.

13.. The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation.

Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score.

If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias.

As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. Linear regression is basic and commonly used types of predictive analysis. The overall idea of regression is to examine two things :

1. Does a set of predictor variables do a good job in predicting an outcome?
2. Which variables in particular are significant predictors of the outcome variables and in what way do they indicated by the magnitude and sign of the beta estimates.

Three major uses for regression analysis are:

1. Determining the strength of predictors.
2. Forecasting an effect.
3. Trend forecasting.

Linear Regression models the relationships between at least on explanatory variables and outcome variables. These variables are known as the independent and dependent variables. When there is one independent variables, the procedures are known as simple linear regression.

15. There are three real branches of statistics :

1. Data Collection
2. Inferential Statistics.

1. DATA COLLECTION :

In this type of statistics, the data is summarized through the given observations. The summarization is one from a sample of population using parameters such as the mean.

Descriptive statistics is a way to organize, represent and describe a collection of data using tables, graphs and summary measures. For example the collection of people in a city using the internet or using television.

Descriptive statistics are also categorized into four different categories:

- Measure of frequency
- Measure of dispersion
- Measure of position

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: Suppose we want to have an idea about the percentage of the illiterate population of our country. We take a sample from the population and find the proportion of illiterate individuals in the sample. With the help of probability, this sample proportion enables us to make some inferences about the population proportion. This study belongs to inferential statistics.