STATISTICS WORKSHEET-1

Q1-Ans- A

Q2-Ans-A

Q3-Ans- B

Q4-Ans-C

Q5-Ans-C

Q6-Ans-B

Q7-Ans-B

Q8-Ans-A

Q9-Ans-C

Q10-Ans-In normally distributed data, there is a constant proportion of data points lying under the curve between the mean and a specific number of standard deviations from the mean.

normal distribution have the following characteristic properties:-

They're all symmetric bell curves. The Gaussian distribution cannot model skewed distributions.

The mean, median, and mode are all equal.

Half of the population is less than the mean and half is greater than the mean.

The Empirical Rule allows you to determine the proportion of values that fall within certain distances from the mean.

Q11-Ans- ways to handle missing values in the dataset:

1. Deleting Rows with missing values
2. Impute missing values for continuous variable
3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values

6. Prediction of missing values

7. Imputation using Deep Learning Library — Data wing

Techniques used In Imputation…

Numerical Variables :-

- Mean/median imputation
- Arbitrary Value Imputation
- End of tail Imputation
- Mode Imputation

Categorical Variable :-

- Frequent category Imputation
- Adding “Missing” category

Both :-

- Complete Case Analysis
- Adding a “Missing” Indicator
- Random Sample Imputation

Q12-Ans- A/B testing — also called split testing or bucket testing—compares the performance of two versions of content to see which one appeals more to visitors/viewers. It tests a control (A) version against a variant (B) version to measure which one is most successful based on your key metrics. As a digital marketing practitioner doing either B2B marketing or B2C marketing, your options for conducting A/B tests include:

- Website A/B testing (copy, images, colors designs, calls to action), which splits traffic between two versions—A and B. You monitor visitor actions to identify which version yields the highest number of 1) conversions or 2) visitors who performed the desired action.
- Email marketing A/B testing (subject line, images, calls to action), which splits recipients into two segments to determine which version generates a higher open rate.
- Content selected by editors or content selected by an algorithm based on user behavior to see which one results in more engagement.

Regardless of the focus, A/B testing helps you determine how to provide the best customer experience (CX).

In addition to A/B tests, there are also A/B/N tests, where the "N" stands for "unknown". An A/B/N test is a type with more than two variations.

Q13-Ans- 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

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.

Q14-Ans- Linear regression is a basic and commonly used type 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 (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable. Naming the Variables. There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regress and. The independent variables can be called exogenous variables, predictor variables, or repressors.

Three major uses for regression analysis are (1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.

Q15-Ans- Statistics may be divided into two main branches:
(1) Descriptive Statistics (2) Inferential 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: 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.