STATISTICS WORKSHEET – 1

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C

Ans 10. NORMAL DISTRIBUTION-

(a) A normal distribution refers to a probability distribution where the values of a random variable are distributed in a symmetrical manner.
(b) The values are distributed equally from the left and the right side of the central side.
(c) And therefore, it forms a bell-shaped curve.
(d) A normal distribution is also known as Gaussian Distribution.
(e) The mean and the standard deviation are the two key parameters in normal distribution.
(f) The mean and standard deviation of the function are 0&1 respectively.
(g) This method supplies an important role in asset calculation and risk management decisions.
(h) The mean, median and mode of the data in this are equal.

Thus, normal distribution method is one of the most popular functions because of its efficiency close to the results and it can be universally applied in every field like data science, finance, economics, business, health etc.

Ans 11. The missing data is a data that is not captured for the observation in the question. Missing data can affect all kinds of tasks for data scientist, or from economic analysis to the clinical trials. Any analyst does not want to produce biased estimates that can surely results into invalid outcomes. Therefore, imputation technique is recommended to handle the missing data.

Imputation-

It is the process of replacing missing values with some estimates based on some assumption and on the available data. It can help in improving the accuracy and efficiency of the processed analysis. By altering the distribution and varying the data it can be done for the process of missingness.

Imputation Techniques-

Mean, Median and Mode are the most commonly used methods of imputing values when it's the time to deal with missing data.

In the cases, where there is number of missing observations, the scientists and analysts can calculate the mean or median or mode of the existing observations and then insert or input the observed data in the place of the missing observations.

Ans 12. A/B testing-

A/B testing is also known by the name Split testing or Bucket testing. It is a methodology to compare two versions or two sides of a web page or app against to each other to calculate which one is performing better.

(a) It is an essential testing method where two or more variants of an app page are shown to users and statistical approach of analysis is used to calculate what performs better for a given conversion goal.
(b) A/B testing makes teams and companies to do the careful changes to their user experiences while collecting the data on the effect it makes.
(c) By taking one change at a time helps that what changes had an effect on user behaviour and what changes did not.
(d) The following are the steps of A/B testing-
   > Collect the data
   > Identify the goals
   > Generate the test hypothesis
   > Creating different variations
   > Run the experiment
   > Wait for the results
   > Analyse the result

Therefore, this method of introducing the changes to a user also allows the experience to be optimized for a desired output and therefore can play an important role in a marketing campaign to make it more effective.

Ans 13. Imputation is the procedure to replace missing values with some estimates based on the assumptions to improve the efficiency and accuracy of the processed analysis.

Mean Imputation

Mean Imputation is also known as mean substitution. It is the method to replace a missing observation with the mean of the non-missing observations for that particular processed analysis. It is a popular solution for missing data because it is easy. But it cannot help in finding the relationships with strong parameter estimates.

But there are many reasons not to use mean imputation-

(a) Mean imputation does not preserve the relationship with variables
Imputing the mean conserves, the mean of the observed data. Thus, if the data is missing completely at random, the estimate of the mean remains unbiased. But the analysts and researchers are more interested in the relationship with variables and therefore mean imputation is not a good solution.

(b) Mean imputation conducts to an underestimate of the standard errors
Any statistical method which uses the imputed data will have a low amount of error. That is, you will get the same mean from mean-imputed data that you have gotten without any imputation method. Ultimately, because of these standard errors that are too low, you are making type errors which is not good.

Ans 14. Linear Regression in Statistics

Linear Regression in Statistics is used to find the relationship between two variables in an observed data. In Linear regression, there are two types of variables i.e. independent variable and Dependent variable. It is commonly used at the time of predictive analysis. It is a type of data analysis technology which predicts the value of unknown data by using another known data value.

Example of linear regression

The weight of a person is related to its height. It shows that a linear relationship presents between the height and weight of the person. As we increase the height, the weight will also increase. And therefore, it shows that there is some critical relationship between the two variables.

In these types of cases, we use scatter plot to simplify the relationship between the two variables.

Linear regression equation is-

$Y = a + bx$

x= independent variable on x-axis

Y= dependent variable on y-axis

b= slope

a= intercept

Ans 15. Statistics

Statistics is all related to the interpretation of data. It is a branch of mathematics that deals with data. The study of the collection, the analysis, interpretation and the organising of the data is known as Statistics. It is a mathematical discipline to collect, organise and summarize the data.

For example: To examine the mean of the marks obtained by each number of students in the class whose strength is 40. The average value is the statistics of the marks obtained.

There are two main branches of Statistics:

1) Descriptive Statistics
2) Inferential Statistics

    Descriptive Statistics:
    It concerns with the presentation and collection of data. The descriptive statistics part is usually the very positive part of a statistical analysis. For a successful analysis, the statistician needs to be attentive of designing experiments, to choose the right focus group and to avoid the biases that are easy to crawl into the experiments.
    Distinct areas of study need distinct kinds of analysis, using the descriptive statistics.
    For example- A physicist studying the cyclone needs the average quantities that change over small interval of time. The nature of this analysis required physical quantities to be averaged from a host of the data collection through the experiment.

Inferential Statistics:

It involves the effective and right conclusions from the statistical analysis that has been performed using the Descriptive Statistics.

For example: The predictions of the future, the generalization about a population comes under the Inferential Statistics by studying a smaller sample.

By drawing, and designing the right experiment, the analyst is able to draw the relevant conclusions of the study. Both the Descriptive Statistics and Inferential Statistics cannot exist without each other. Good and effective methodologies need to be followed in both of these steps. Therefore, both the branches are equally important for an analyst/researcher.