G H Raisoni College of Engineering and Management, Pune.

F.Y B.TECH (Engineering)
FIRST TERM 2020-21
CAE-II (2020 Pattern)

Department - Information Technology (IT)
Term / Section - Term I
Date of Examination - 05/04/2021
Subject Name/Code - Foundation of Data Analysis (UCOL102).
Roll NO - C70 Name - Swayam Terode.

CO2 1. Analyze the given data 6,2,3,1. Find out the Standard deviation of given data set.

Answer: Given data set is
6, 2, 3, 1.

$$\text{Mean} = \frac{6+2+3+1}{4} = \frac{12}{4} = 3$$

$$\text{Variance} = \frac{\sum_{i=1}^{n} (x - x_i)^2}{n}$$

$$= \frac{(6-3)^2 + (2-3)^2 + (3-3)^2 + (1-3)^2}{4} = \frac{3^2 + (-1)^2 + 0^2 + (-2)^2}{4}$$

$$= \frac{9+1+0+4}{4} = \frac{14}{4} = 3.5$$

Standard deviation = $\sqrt{\text{variance}} = \sqrt{3.5} = 1.87082$

2. Distinguish Normal Distribution and Standard Normal Distribution.

Answer: A normal distribution is distributed by two parameters the mean and the variance.

Now the standard normal distribution is a specific distribution with mean 0 and variance 1.

If x has the normal distribution with mean m and variance s^2 then if we define.

$z = \frac{x-m}{s}$ then z has the standard normal distribution.

So for any specific normal distribution we can calculate probabilities of the form $P[a<x<b]$ from the table.

This is because we write $x = sz+m$.

So the standard normal plays a special role with respect to the general family of normal distribution.

The standard normal distribution is just a normal distribution scaled/standardized by the z-formula.

CO2 3. Explain Inferential Statistics in detail.

Answer: Inferential Statistics use a random sample of data taken from a population to describe and make inferences about the population.

This is done by taking a random sample of individuals within the population of interest, and taking measurements.

From these measurement, various parameters can be estimated about the overall population. Because inferential statistics does not sample everyone in a population, the results will always contain some level of uncertainty. Uncertainty can be reduced by increasing the size of your sample.

Methods of Inferential Statistics:

1. Confidence intervals
2. Hypothesis Testing

CO3 4. Demonstrate the concept of Simple Linear regression and multiple linear regression.

Answer: Simple Linear Regression:

Simple Linear Regression is used to find out the best relationship between a single input variable and output variable provided that both variables are continuous in nature.

Steps to Implement Simple Linear Regression:

1. Analyze data
2. Get sample data for model building
3. Then design a model that explains the data
4. And use the same developed model on the whole population to make predictions.

The Equation is $y = \beta 0 + \beta 1 * x$

Multiple Linear Regression (MLR).

It is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables.

Formula:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + -- + \beta_p x_{ip} + \epsilon .$$

where y_i = dependent variables.
x_i = explanatory variables.
β_0 = y-intercept (constant term).
β_p = slope coefficient for each explanatory variable.
ϵ = Residuals.

CO3 5. Elaborate the terms 1.) Correlation 2.) Regression

Answer: 1.) Correlation :

Correlation is a statistic that measures the degree to which two variables move in relation to each other.

The fit of the data can be visually represented in a scatterplot. Using scatterplot, we can generally assess the relationship between the variables and determine whether they are correlated or not.

2.) Regression :

Regression analysis is statistical method that helps us to analyse and understand the relationship between two or more variables of interest.

The process that is adapted to perform regression analysis helps to understand which factors are important, which factors can be ignored and how they are influencing each other.

Regression is of two type:

1.) Linear Regression
2.) Multiple Linear Regression.