U3A Maths

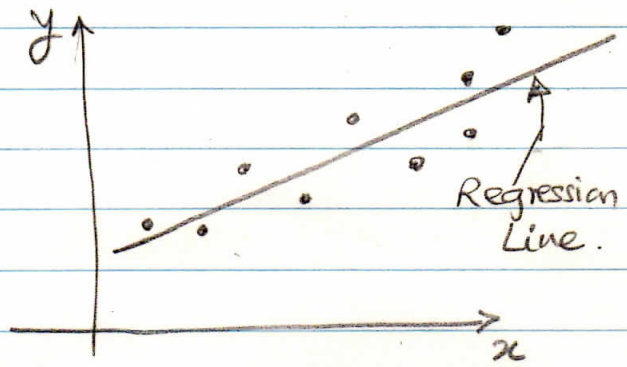
9-Dec-2022 Term 4/Week 9

Linear Regression

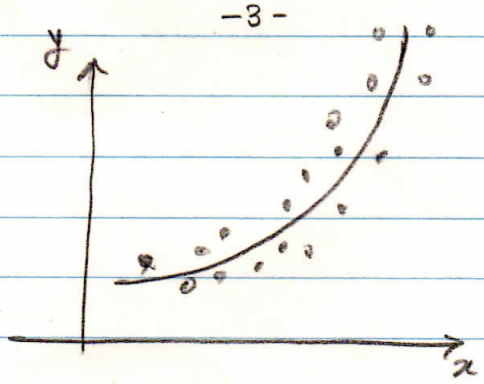
- Regression analysis is used to predict the value of a variable (dependent variable) based on the value of another variable (independent variable).

- Linear Regression uses a method to fit a straight line through a given values of observations.

- A graphical illustration of a linear regression line is as shown below:



- In practice, it is possible to have a set of observations, which may need non-linear (curved line) regression, as shown on the next page.



- However, for most problems (applications) in statistics, linear regression is found to be adequate. Hence, non-linear regression is rarely used.

- Linear regression is extensively used in the following statistical applications.

  - Economics & Finance (mainly for forecasting) (Ex: sales, profit, shares etc)
  - Weather forecasting
  - Biological analysis (Ex: relation between age, height and weight)
  - Behavioural Analysis & prediction.

The following example illustrates the regression line equation. It is useful, even though obtaining statistical regression line equation is more complicated.

Ex.1

A plumber charges $120 for 1 hour and $200 for 3 hours work. How much will the charge for 3 hours work.

Let

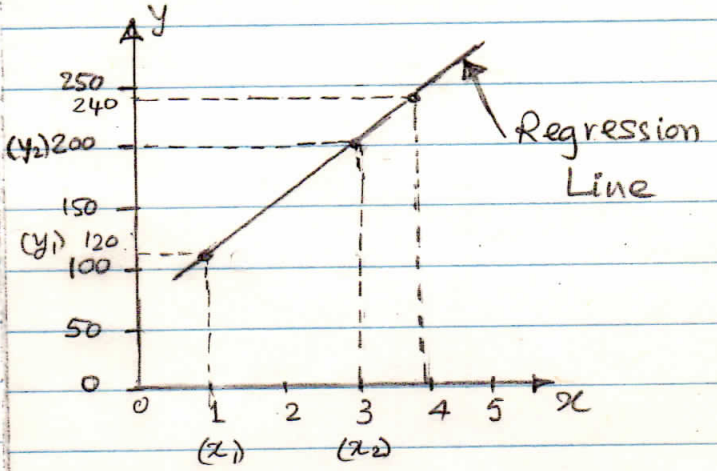
x ⇒ No. of hours (Independent Variable)

y ⇒ Amount (Dependent Variable)

Given data is:

for $x_1 = 1$; $y_1 = 120$

$x_2 = 3$; $y_2 = 200$



A graphical representation of the given data and the regression line are as shown above.

From mathematics, we know that equation of the line is "$y = mx + c$"; however, in statistics the following form of the equation is used:

$$y = a + bx$$

We have to find the values of 'a' and 'b' using the given data values.

∴ we have:

Eqn(1): $y_1 = a + bx_1 \Rightarrow 120 = a + b$

Eqn(2): $y_2 = a + bx_2 \Rightarrow 200 = a + 3b$

We need to solve the above simultaneous Eqns.

Eqn(2) − Eqn(1) ⇒ $2b = 80$

$\therefore b = 40$

Using Eqn(1) ⇒ $a = 120 - b$

$= 120 - 40$

$= 80$

Hence, the regression line is

$$y = 80 + 40x$$

Hence, for $x = 4$

$y = 80 + 40 \times 4$

$= 240$

∴ It costs $240 for 4 hrs.