Multiple Linear Regression.

Problem: Predict the Value of y given x_1 and x_2

Solution: Here we have 2 independent Variable and 1 dependent variable.

Given:

y	x_1	x_2
-3.7	3	8
3.5	4	5
2.5	5	7
11.5	6	3
5.7	2	1

Multiple Regression Equation with 2 independent Variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

β_0: Y-intercept; β_1, β_2: slopes; ε: Random Error.

The above equation can be rewritten as

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$b_1 = \frac{\Sigma x_2^2 \, \Sigma x_1 y - \Sigma x_1 x_2 \, \Sigma x_2 y}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$b_2 = \frac{\Sigma x_1^2 \, \Sigma x_2 y - \Sigma x_1 x_2 \, \Sigma x_1 y}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

Matrix form Multiple Linear Regression for 2 independent Variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & & \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\sum x_1^2 = \sum X_1 X_1 - \frac{\sum X_1 \sum X_1}{N}$$ N - no. of record

$$\sum x_2^2 = \sum X_2 X_2 - \frac{\sum X_2 \sum X_2}{N}$$

$$\sum x_1 y = \sum X_1 Y - \frac{\sum X_1 \sum Y}{N}$$

$$\sum x_2 y = \sum X_2 Y - \frac{\sum X_2 \sum Y}{N}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{N}$$

	Y	X_1	X_2	X_1X_1	X_2X_2	X_1X_2	X_1Y	X_2Y
	-3.7	3	8	9	64	24	-11.1	-29.6
	3.5	4	5	16	25	20	14	17.5
	2.5	5	7	25	49	35	12.5	17.5
	11.5	6	3	36	9	18	69	34.5
	5.7	2	1	4	1	2	11.4	5.7
$\sum$	19.5	20	24	90	148	99	95.8	45.6

$$\sum x_1^2 = \sum X_1 X_1 - \frac{\sum X_1 \sum X_1}{N} = 90 - \frac{20 \times 20}{5} = 10$$

$$\sum x_2^2 = \sum X_2 X_2 - \frac{\sum X_2 \sum X_2}{N} = 148 - \frac{24 \times 24}{5} = 32.8$$

$$\sum x_1 y = \sum X_1 Y - \frac{\sum X_1 \sum Y}{N} = 95.8 - \frac{20 \times 19.5}{5} = 17.8$$

$$\sum x_2 y = \sum X_2 Y - \frac{\sum X_2 \sum Y}{N} = 45.6 - \frac{24 \times 19.5}{5} = -48$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{N} = 99 - \frac{20 \times 24}{5} = 3$$

$$b_1 = \frac{\Sigma x_2^2 \, \Sigma x_1 y - \Sigma x_1 x_2 \, \Sigma x_2 y}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$= \frac{32.8 * 17.8 - 3 * (-48)}{10 * 32.8 - 3 * 3} = 2.28$$

$$b_2 = \frac{\Sigma x_1^2 \, \Sigma x_2 y - \Sigma x_1 x_2 \, \Sigma x_1 y}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$= \frac{10 * (-48) - 3 * (17.8)}{10 * 32.8 - 3 * 3} = -1.67$$

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 =$$

$$= \frac{19.5}{5} - 2.28 * \frac{20}{5} - (-1.67) * \frac{24}{5}$$

$$= 2.796.$$

Final Regression equation or Model is:

$$\hat{y} = 2.796 + 2.28 x_1 - 1.67 x_2.$$

For given $x_1 = 3$ and $x_2 = 2$

Find y.

$$\hat{y} = 2.796 + 2.28 * 3 - 1.67 * 2$$

$$= 6.296.$$

Error calculation: $R^2 = \frac{(\Sigma \hat{y} - \bar{y})^2}{(y - \bar{y})^2}$

$\bar{y}$ - mean of y

y - actual value

$\hat{y}$ - Predicted value.