*Department of Computer Science*
*University of Karachi*
*Operation Research*
*Supervisor: Dr. Shaista Rais*

# <u>Forecasting Model</u>

Multiple linear regression is a method we can use to quantify the relationship between two or more predictor variables and a response variable.

## Question:

Suppose we have the following dataset with one response variable $y$ and two predictor variables $X_1$ and $X_2$:

| y | $X_1$ | $X_2$ |
|-----|-----|-----|
| 140 | 60 | 22 |
| 155 | 62 | 25 |
| 159 | 67 | 24 |
| 179 | 70 | 20 |
| 192 | 71 | 15 |
| 200 | 72 | 14 |
| 212 | 75 | 14 |
| 215 | 78 | 11 |

Use the following steps to fit a multiple linear regression model to this dataset.

## Step 1: Calculate $X_1^2$, $X_2^2$, $X_1y$, $X_2y$ and $X_1X_2$.

| y | $X_1$ | $X_2$ | | $X_1^2$ | $X_2^2$ | $X_1y$ | $X_2y$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|---|
| 140 | 60 | 22 | | 3600 | 484 | 8400 | 3080 | 1320 |
| 155 | 62 | 25 | | 3844 | 625 | 9610 | 3875 | 1550 |
| 159 | 67 | 24 | | 4489 | 576 | 10653 | 3816 | 1608 |
| 179 | 70 | 20 | | 4900 | 400 | 12530 | 3580 | 1400 |
| 192 | 71 | 15 | | 5041 | 225 | 13632 | 2880 | 1065 |
| 200 | 72 | 14 | | 5184 | 196 | 14400 | 2800 | 1008 |
| 212 | 75 | 14 | | 5625 | 196 | 15900 | 2968 | 1050 |
| 215 | 78 | 11 | | 6084 | 121 | 16770 | 2365 | 858 |
| **Mean** 181.5 | 69.375 | 18.125 | **Sum** | 38767 | 2823 | 101895 | 25364 | 9859 |
| **Sum** 1452 | 555 | 145 | | | | | | |

## Step 2: Calculate Regression Sums.

Next, make the following regression sum calculations:

- $\Sigma x_1^2 = \Sigma X_1^2 - (\Sigma X_1)^2 / n = 38{,}767 - (555)^2 / 8 = \mathbf{263.875}$
- $\Sigma x_2^2 = \Sigma X_2^2 - (\Sigma X_2)^2 / n = 2{,}823 - (145)^2 / 8 = \mathbf{194.875}$
- $\Sigma x_1 y = \Sigma X_1 y - (\Sigma X_1 \Sigma y) / n = 101{,}895 - (555*1{,}452) / 8 = \mathbf{1{,}162.5}$
- $\Sigma x_2 y = \Sigma X_2 y - (\Sigma X_2 \Sigma y) / n = 25{,}364 - (145*1{,}452) / 8 = \mathbf{-953.5}$
- $\Sigma x_1 x_2 = \Sigma X_1 X_2 - (\Sigma X_1 \Sigma X_2) / n = 9{,}859 - (555*145) / 8 = \mathbf{-200.375}$

| y | $X_1$ | $X_2$ | | $X_1^2$ | $X_2^2$ | $X_1y$ | $X_2y$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|---|
| 140 | 60 | 22 | | 3600 | 484 | 8400 | 3080 | 1320 |
| 155 | 62 | 25 | | 3844 | 625 | 9610 | 3875 | 1550 |
| 159 | 67 | 24 | | 4489 | 576 | 10653 | 3816 | 1608 |
| 179 | 70 | 20 | | 4900 | 400 | 12530 | 3580 | 1400 |
| 192 | 71 | 15 | | 5041 | 225 | 13632 | 2880 | 1065 |
| 200 | 72 | 14 | | 5184 | 196 | 14400 | 2800 | 1008 |
| 212 | 75 | 14 | | 5625 | 196 | 15900 | 2968 | 1050 |
| 215 | 78 | 11 | | 6084 | 121 | 16770 | 2365 | 858 |
| **Mean** 181.5 | 69.375 | 18.125 | **Sum** | 38767 | 2823 | 101895 | 25364 | 9859 |
| **Sum** 1452 | 555 | 145 | | | | | | |

| **Reg Sums** | 263.875 | 194.875 | 1162.5 | -953.5 | -200.375 |
|---|---|---|---|---|---|

**Step 3: Calculate $b_0$, $b_1$, and $b_2$.**

The formula to calculate $b_1$ is: $[(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)] / [(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]$

Thus, $\mathbf{b_1}$ = $[(194.875)(1162.5) - (-200.375)(-953.5)] / [(263.875)(194.875) - (-200.375)^2]$ = **3.148**

The formula to calculate $b_2$ is: $[(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)] / [(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]$

Thus, $\mathbf{b_2}$ = $[(263.875)(-953.5) - (-200.375)(1152.5)] / [(263.875)(194.875) - (-200.375)^2]$ = **-1.656**

The formula to calculate $b_0$ is: $y - b_1 X_1 - b_2 X_2$

Thus, $\mathbf{b_0}$ = $181.5 - 3.148(69.375) - (-1.656)(18.125)$ = **-6.867**

**Step 5: Place $b_0$, $b_1$, and $b_2$ in the estimated linear regression equation.**

The estimated linear regression equation is: $\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$

In our example, it is $\hat{y}$ = **-6.867 + 3.148$x_1$ − 1.656$x_2$**

## Interpretation of a Multiple Linear Regression Equation

Here is how to interpret this estimated linear regression equation: $\hat{y}$ = -6.867 + 3.148$x_1$ − 1.656$x_2$

$\mathbf{b_0}$ = **-6.867**. When both predictor variables are equal to zero, the mean value for y is -6.867.

**$b_1$ = 3.148**. A one unit increase in $x_1$ is associated with a 3.148 unit increase in y, on average, assuming $x_2$ is held constant.

**$b_2$ = -1.656**. A one unit increase in $x_2$ is associated with a 1.656 unit decrease in y, on average, assuming $x_1$ is held constant.