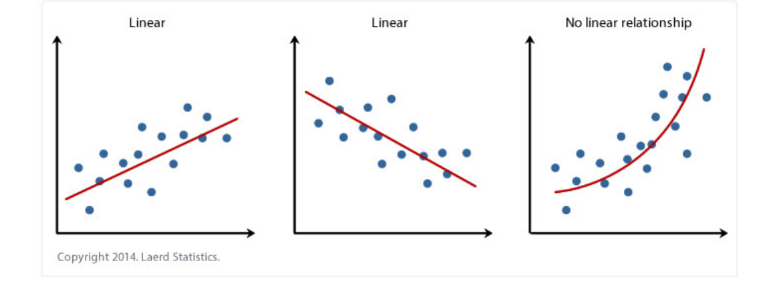


**Answer:**

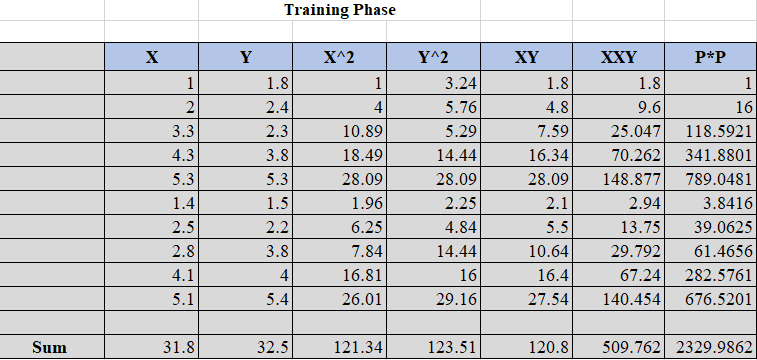
Non-linear regression has more serious overfitting issue. Linear regression is less prone to overfitting than non-linear regression because it has a simpler model structure.



**Training Phase**

X and Y values are from Real **Data Set 1**, 50% of the collected data.

As shown in the table below, values for X^2, Y^2, X\*Y, X\*X\*Y, P\*P and summation of each are calculated to be used in the equations.



Then we will calculate Slope(b) and Intercept(a) values for both **Model 1 (Linear Regression)** and **Model 2 (Non-Linear Regression)**.

**Equations for Linear Regression**

Slope(b1) = (NΣXY - (ΣX)(ΣY)) / (NΣX2 - (ΣX)2)

Intercept(a1) = (ΣY - b(ΣX)) / N

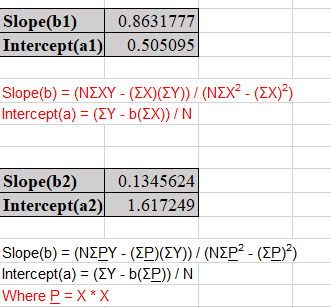
**Equations for Non-Linear Equation**

Slope(b2) = (NΣPY - (ΣP)(ΣY)) / (NΣP2 - (ΣP)2)

Intercept(a2) = (ΣY - b(ΣP)) / N

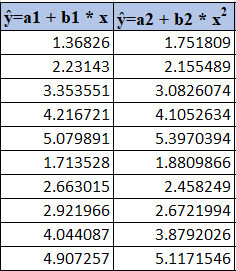
Where P = X \* X

The values are as follows:



\* The value of N is 10 for Training Phase.

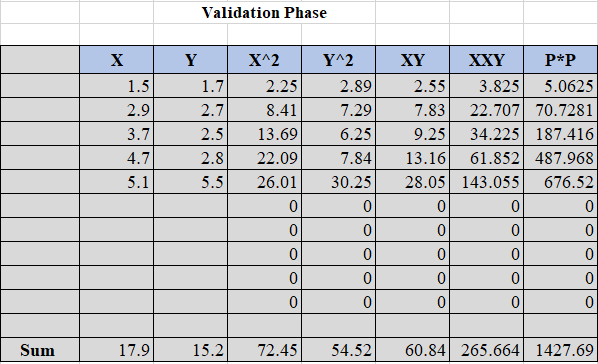
By using the values calculated above, we can calculate ŷ values for both **Model 1** and **Model 2**.



**Validation Phase**

X and Y values are from **Real Data Set 2**, 25% of the collected data.

As shown in the table below, values for X^2, Y^2, X\*Y, X\*X\*Y, P\*P and summation of each are calculated to be used in the equations.



Then we will calculate Slope(b) and Intercept(a) values for both **Model 1 (Linear Regression)** and **Model 2 (Non-Linear Regression)**.

**Equations for Linear Regression**

Slope(b1) = (NΣXY - (ΣX)(ΣY)) / (NΣX2 - (ΣX)2)

Intercept(a1) = (ΣY - b(ΣX)) / N

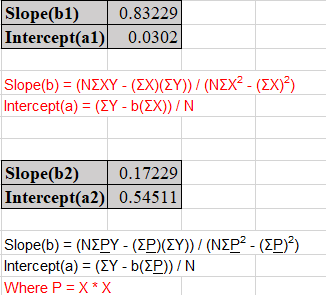
**Equations for Non-Linear Equation**

Slope(b2) = (NΣPY - (ΣP)(ΣY)) / (NΣP2 - (ΣP)2)

Intercept(a2) = (ΣY - b(ΣP)) / N

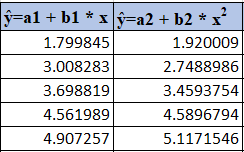
Where P = X \* X

The values are as follows:



\* The value of N is 5 for Validation Phase.

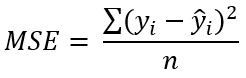
By using the values calculated above, we can calculate ŷ values for both **Model 1** and **Model 2**.

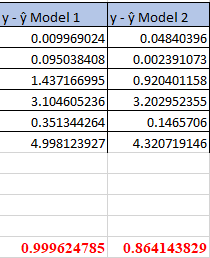
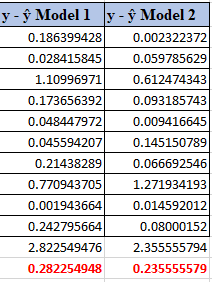


**Test Phase**

X values are from Real **Data Set 3**, 25% of the collected data.

We will calculate mean squared error (MSE) for both **Model 1** and **Model 2** by using the following equation.





Training Set Validation Set

Then we will decide which Model to choose by using the following equation.

max(Training\_Set\_MSE, Validation\_Set\_MSE) / min(Training\_Set\_MSE, Validation\_Set\_MSE)

**Model 1**

Training\_Set\_MSE = 0.28225

Validation\_Set\_MSE = 0.99962

0.99962/0.28225 = **3.54**

**Model 2**

Training\_Set\_MSE = 0.23555

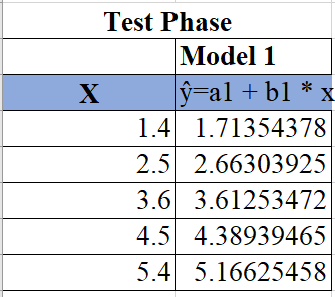
Validation\_Set\_MSE = 0.86414

0.86414/0.23555 = 3.668

According to the calculation, **Model 1** is the better model. So, we will use the values of b and a from **Model 1**.

Slope(b1) = 0.8631777

Intercept(a1) = 0.505095



**Final Answer:**

