

Week 4

Sai Kumar Murarishetti
Lewis ID: L30079224

1. (By Hand) For the dependent variable Y and the independent variables X1 and X2, the linear regression model is given by:

$Y = 0.08059 \cdot X_1 - 0.16109 \cdot X_2 + 5.26570$. Complete the following table:

Actual Y	x1	x2	Predicted Y	Residuals (Predication Error)
6	6.8	4.7		
3.1	5.3	5.5		
5.8	4.5	6.2		
4.5	8.8	7		
4.5	6.8	6.1		
3.7	8.5	5.1		
5.4	8.9	4.8		
5.1	6.9	5.4		
5.8	9.3	5.9		
5.7	8.4	5.4		

Is this a good model? Why? Why not?

Actual Y	X1	X2	Predicted Y	Residuals (Predication Error)
6	6.8	4.7	5.05658	-0.94341
3.1	5.3	5.5	4.80683	1.70683
5.8	4.5	6.2	4.62959	-1.17040
4.5	8.8	7	4.84726	0.34726
4.5	6.8	6.1	4.83106	0.33106
3.7	8.5	5.1	5.12915	1.42915
5.4	8.9	4.8	5.20971	-0.19028
5.1	6.9	5.4	4.95186	-0.14811
5.8	9.3	5.9	5.06475	-0.73524
5.7	8.4	5.4	5.07277	-0.62723

We can simplify it by using Root Mean Square Error = $\sqrt{\sum(x - y)^2/n}$

$$= (-0.94341)^2 + (1.70683)^2 + (-1.17040)^2 + (0.34726)^2 + (0.33106)^2 + (1.42915)^2 + (-0.19028)^2 + (-0.14811)^2 + (-0.73524)^2 + (-0.62723)^2$$

$$= 8.4380$$

$$N = 10$$

$$= 8.4380 / 10$$

$$= 0.8438$$

$$= \sqrt{0.8438}$$

$$= 0.918586$$

Here we can say that it is low.

Our Data Model is good.

2. For the data set associated with this homework (HBAT). Using X19 as the dependent variable and (X6, X7, X9, X11, X12 and X16) as the independent variables:

a. Find the parameters (coefficients) for the Linear Regression Model, then write down the equation of the model.

The equation for Linear regression

$$Y = B_0 + B_1(x_6) + B_2(x_7) + B_3(x_9) + B_4(x_{11}) + B_5(x_{12}) + B_6(x_{16})$$

b. Find the coefficient of determination and the standard error of the estimate. How accurate is the model?

Standard error of the estimate it is called SEE.

The SEE can calculate as = Sqrt (MSE)

c. If you are asked to remove two independent variables, which two variables would you choose and why?

d. After removing the two variables found in part c, re-run parts and b. Compare the results. Which model is more accurate and why?

linear regression equation is, $Y = B_0 + B_1(x_6) + B_2(x_9) + B_3(x_{11}) + B_4(x_{12})$