MCS-Model and Inference
Assignment # 4
Course Supervisor: Ms.Shaista Rais
University of Karachi

Assignment # 4

Q. In a small scale regression study, the following data were obtained:

y	42	33	75	28	91	55
x1	7	4	16	3	21	8
x2	33	41	7	49	5	31

Calculate Line of Best Fit.

Calculate Multiple R and R^2

Calculate Adjusted R^2 *and state the importance of adjusted* R^2 *in multiple regression model.*

Hint for Adjusted R^2

Coefficient of Determination, R-squared, and Adjusted R-squared

- As in simple linear regression, $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$, and represents the proportion of variation in y (about its mean) "explained" by the multiple linear regression model with predictors, $x_1, x_2, \ldots$
- If we start with a simple linear regression model with one predictor variable, x_1, then add a second predictor variable, x_2, SSE will decrease (or stay the same) while $SSTO$ remains constant, and so R^2 will increase (or stay the same). In other words, R^2 always increases (or stays the same) as more predictors are added to a multiple linear regression model, *even if the predictors added are unrelated to the response variable*. Thus, by itself, R^2 cannot be used to help us identify which predictors should be included in a model and which should be excluded.
- An alternative measure, adjusted R^2, does not necessarily increase as more predictors are added, and can be used to help us identify which predictors should be included in a model and which should be excluded. Adjusted $R^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right)(1 - R^2)$, and, while it has no practical interpretation, is useful for such model building purposes. Simply stated, when comparing two models used to predict the same response variable, we generally prefer the model with the higher value of adjusted R^2 – see Lesson 11 for more details.

Note: SSTO is the same as SST