Special Topics Comp Stat & Pro MAT5999 and Computational Stats & Prob. AIM 5002
Written Assignment 10 (12.5 points)
5/2/22

Solutions to be returned by the beginning of class on Tuesday, 5/9.

1. A response Y is a function of three independent variables x_1, x_2, and x_3 that are related as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon,$$

Our goal is to fit this model to the $n = 7$ data points shown in the accompanying table.

y	x_1	x_2	x_3
1	-3	5	-1
0	-2	0	1
0	-1	-3	1
1	0	-4	0
2	1	-3	-1
3	2	0	-1
3	3	5	1

(a) (3 points) Find the least squares estimates of the model parameters β_i, $i = 1, 2, 3$. Hint: You can use R to perform the necessary computations.

(b) (3.5 points) Estimate $\mathbb{E}(Y)$ when $x_1 = 1, x_2 = -3, x_3 = -1$. Compare with the observed response in the original data. Why are these two not equal?

(c) (3 points) Give a 95% prediction interval for Y when $x_1 = 1, x_2 = -3, x_3 = -1$.

2. (3 points) **Correlation does not mean causation**. Assume that in a simple linear model, we can reject the null hypothesis $\beta_1 = 0$ at very high confidence level. That is, we are confident that Y is correlated with X. It is important to remember that this does not necessarily mean that the increase of X *causes* Y to increase. E.g., if X is the number of sunglasses sold and Y is the amount of ice cream sold, we can easily find a correlation between X and Y but an increase in X does not cause an increase in Y. Indeed, there is a hidden variable, a.k.a. confounding variable (in this example being hot sunny weather) that causes both X and Y to increase.

   Think about this and give 3 more examples where correlation does not mean causation (no need to give data, just examples. You can look up examples online).

   Just for fun: https://imgs.xkcd.com/comics/correlation.png