

B365 Homework 7

Zac Monroe

November 2018

1. (a) See prob1.r to view my matrix X . Try sourcing prob1.r into R and asking for the first 20 rows of X by prompting $X[1:20,]$.
(b) Source prob1.r into R to get a printout of my α and β values. I got roughly $\alpha = 0.3318736$ and $\beta = 1.677939$.
(c) Yes, it appears that people with more education tend to have larger vocabularies; this is clear because the α -value that I got by solving the normal equations is positive and not very close to zero, so a larger education number seems to result in a higher test score.
(d) From the data I can reason that 3 extra years of education would seem to give a random participant a boost of one point on this particular vocabulary test.
2. (a) Source prob2.r into R to get a printout of all my regression coefficients. I included a bias column (the last column in my matrix X) just to be safe.
(b) My initial sse is roughly 5.905161. This can be seen by sourcing prob2.r into R.
(c) The variable *ferr*'s omission seems to give the greatest in sse, and thus I believe that it appears most important in predicting rcc with these data. If, by omitting it, we have the hardest time giving proper regression values for the test data, it is most important.
3. (a) Source prob3.r into R to get the plot.
(b) Source prob3.r into R to get the line on top of the plot.
(c) I have commented out the line in prob3.r that adds this model to the plot, because it was too confusing to look at with also the model that accounts for linear growth.
Red is the data points. Blue is the model that doesn't account for linear growth.
(d) Source prob3.r into R to get this plot with the model that accounts for linear growth. Green data points are those of the model. The results do indeed suggest that the company is experiencing at least a little bit of growth in sales because the coefficient d on the linear term x_i (seen by sourcing prob3.r into R) is positive, and also because the sse of the model with the linear growth is lower than the sse of the model that did not account for linear growth (again, source prob3.r into R to see these results).