Bruce Campbell ST-617 Discussion Problem #2

Wed Jun 29 16:30:13 2016

Chapter 3

Problem 3

```
Suppose we have a data set with five predictors, X1 =GPA, X2 = IQ,
X3 = Gender (1 for Female and 0 for Male), X4 = Interaction between
GPA and IQ, and X5 = Interaction between GPA and Gender. The
response is starting salary after graduation (in thousands of dollars).
Suppose we use least squares to fit the model, and get
```

$$\beta_0 = 50, \beta_1 = 20, \beta_2 = 0.07, \beta_3 = 35, \beta_4 = 0.01, \beta_5 = -10$$

The linear model is then

$$Salary = 50 + 20 * GPA + 0.07 * IQ + 35 * Gender + 0.01 * (GPA * IQ) - 10 * (GPA * Gender)$$

```
rm(list = ls())
GPA <- (0:0.2:4)
```

a) Which answer is correct, and why?

```
i. For a fixed value of IQ and GPA, males earn more on average
than females.
ii. For a fixed value of IQ and GPA, females earn more on
average than males.
iii. For a fixed value of IQ and GPA, males earn more on average
than females provided that the GPA is high enough.
iv. For a fixed value of IQ and GPA, females earn more on
average than males provided that the GPA is high enough.
```

Let's rewrite our regression equation, gathering the gender term

$$Salary = Gender * \big(35 + -10GPA)\big) + \big(50 + 20 * GPA + 0.07 * IQ + 0.01 * (GPA * IQ)\big)$$

Note that for all the values of IQ and GPA the term on the right is positive, that gender is a binary predictor and that range of GPA is $[0, 4]$. We see that when Gender=1 (female) and $GPA < 3.5$ the contribution from gender to salary is positive. Therefore the correct statement is i).

b) Predict the salary of a female with IQ of 110 and a GPA of 4.0.

```
GPA <- 4
IQ <- 110
Gender <- 1
Salary <- 50 + 20 * GPA + 0.07 * IQ + 35 * Gender + 0.01 * (GPA * IQ) - 10 *
    (GPA * Gender)
```

The predicted salary is 137.1

c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

This is a true statement. To validate it we need to consider the scale of the perdictors, and ideally we'd like to see p-values as well.

$$range(GPA * IQ) = [min(GPA) * min(IQ), max(GPA) * max(IQ)] \approx [0, 800]$$

More realistically, the range is probably $[50, 800]$, at this level all other variables held fixed for every unit increase in GPA*IQ we expect a 0.01 increase in salary (in units of $1,000) - or an additional $10. At the midpoints of the range of IQ* GPA (400) we have an increase of o fonly 4000. Given that the $\beta_0 = 50$, and the other coefficients and perdictor ranges, we are lead to the conclusion that this interaction term is not significant.