Homework 2 Solution

3.1 (i) *hsperc* is defined so that the smaller it is, the lower the student's standing in high school. Everything else equal, the worse the student's standing in high school, the lower his/her expected college GPA.

(ii) Just plug these values into the equation

$$\widehat{colgpa} = 1.392 - .0135(20) + .00148(1050) = 2.676.$$

(iii) The difference between A and B is simply 140 times the coefficient on *sat*, because *hsperc* is the same for both students. So A is predicted to have a score .00148(140) $\approx$ .207 higher.

(iv) With *hsperc* fixed, $\Delta\widehat{colgpa} = .00148\Delta sat$. Now, we want to find Δsat such that $\Delta\widehat{colgpa} = .5$, so $.5 = .00148(\Delta sat)$ or $\Delta sat = .5/(.00148) \approx 338$. Perhaps not surprisingly, a large ceteris paribus difference in SAT score – almost two and one-half standard deviations – is needed to obtain a predicted difference in college GPA or a half a point.

3.5 (i) No. By definition, *study* + *sleep* + *work* + *leisure* = 168. Therefore, if we change *study*, we must change at least one of the other categories so that the sum is still 168.

(ii) From part (i), we can write, say, *study* as a perfect linear function of the other independent variables: *study* = 168 − *sleep* − *work* − *leisure*. This holds for every observation, so MLR.3 is violated.

(iii) Simply drop one of the independent variables, say *leisure*:

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + u.$$

Now, for example, β_1 is interpreted as the change in *GPA* when *study* increases by one hour, where *sleep*, *work*, and *u* are all held fixed. If we are holding *sleep* and *work* fixed but increasing *study* by one hour, then we must be reducing *leisure* by one hour. The other slope parameters have a similar interpretation.

Computer Problems

C3.1 (i) Probably β_2 > 0, as more income typically means better nutrition for the mother and better prenatal care.

(ii) On the one hand, an increase in income generally increases the consumption of a food, and *cigs* and *faminc* could be positively correlated. On the other, family incomes are also higher for families with more education, and more education and cigarette smoking tend to be negatively correlated. The sample correlation between *cigs* and *faminc* is about –.173, indicating a negative correlation.

(iii) The regressions without and with *faminc* are

$$bwght = 119.77 - .514\,cigs$$

$$n = 1{,}388,\ \ R^2 = .023$$

and

$$bwght = 116.97 - .463\,cigs + .093\ faminc$$

$$n = 1{,}388,\ \ R^2 = .030.$$

The effect of cigarette smoking is slightly smaller when *faminc* is added to the regression, but the difference is not great. This is due to the fact that *cigs* and *faminc* are not very correlated, and the coefficient on *faminc* is practically small. (The variable *faminc* is measured in thousands, so $10,000 more in 1988 income increases predicted birth weight by only .93 ounces.)

C3.6(i) The slope coefficient from the regression *IQ* on *educ* is (rounded to five decimal places) $\tilde{\delta}_1 = 3.53383.$

(ii) The slope coefficient from log(*wage*) on *educ* is $\tilde{\beta}_1 = .05984.$

(iii) The slope coefficients from log(*wage*) on *educ* and*IQ* are $\hat{\beta}_1 = .03912 \text{ and } \hat{\beta}_2 = .00586,$ respectively.

(iv) We have $\hat{\beta}_1 + \tilde{\delta}_1\hat{\beta}_2 = .03912 + 3.53383(.00586) \approx .05983,$ which is very close to .05984; the small difference is due to rounding error.

C3.10 (i) The variable *educ* ranges from 6 to 20. Out of 1,230 men, 512, or 41.63%, completed 12th grade, but no higher. The average level of education for the men in the sample is about 13.04, which is higher than the average of *motheduc* (12.18) and *fatheduc* (12.45).

(ii) The regression results are

$$\widehat{educ} = 6.96 + .304\, motheduc + .190\, fatheduc$$

$$n = 1{,}230 \quad R^2 = .249.$$

About 25% of the variation in *educ* is explained by parents' education. The coefficient on *motheduc* means that, holding father's education fixed, another year of the mother's education is associated with about .304 additional years of education for the child, or slightly less than one-third of a year.

(iii) When *abil* is added to the regression, we get

$$\widehat{educ} = 8.45 + .189\, motheduc + .111\, fatheduc + .502\, abil$$

$$n = 1{,}230 \quad R^2 = .428$$

The three explanatory variables together explain almost 43% of the variation in *educ*. This is much more than in part (ii); clearly *abil* is helping.

(iv) When $abil^2$ is added to the regression, the estimated equation is

$$\widehat{educ} = 8.24 + .190\, motheduc + .109\, fatheduc + .401\, abil + .051\, abil^2$$

$$n = 1{,}230 \quad R^2 = .444$$

The derivative with respect to *abil* is $.401 + .102\, abil$. Setting equal to zero and solving gives

$$abil^* = -.\frac{401}{.102} \approx -3.93,$$

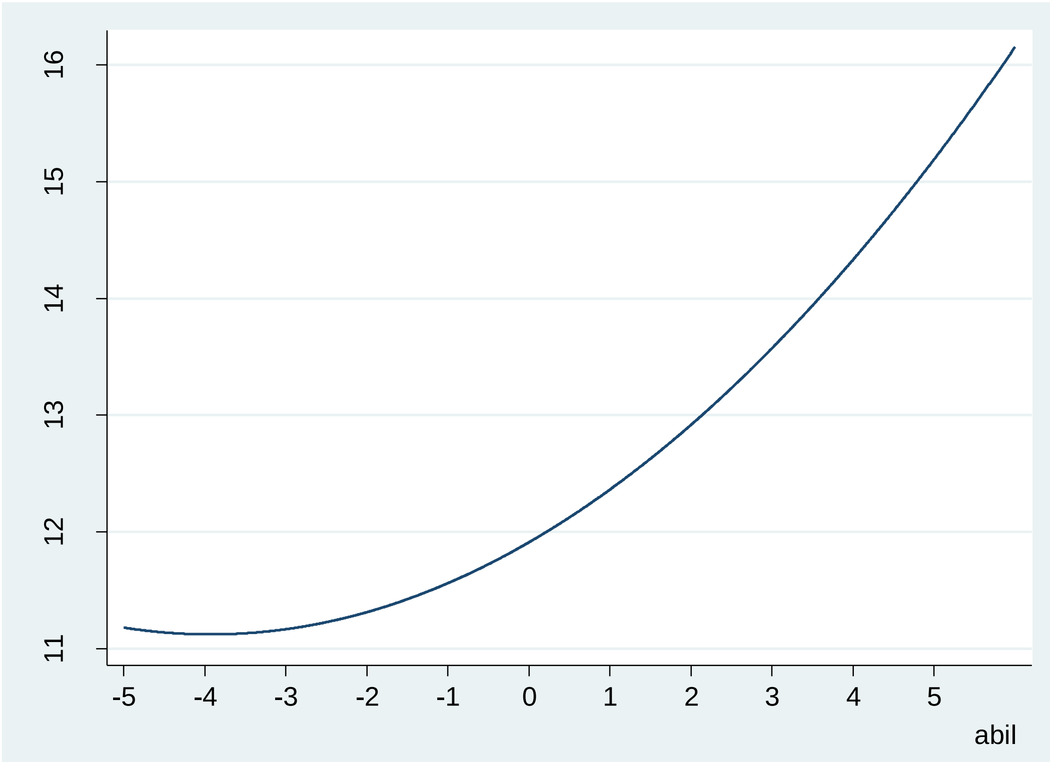
so about -4. The second derivative is $.102$, and so we know we have found the global minimum.

(v) Out of 1,230 men, only 15 have $abil < -3.93$, or only about 1.2 percent of the sample. This is reassuring because it means we can effectively ignore what is happening to the left of -3.93. The important story is that the level of education increases with ability at an increasing rate.

(vi) I used the equation from part (iv) and plugged in the mean values for *motheduc* and *fatheduc*. Thus, I used the equation

$$\widehat{educ} = 8.24 + .190(12.18) + .109(12.45) + .401\, abil + .051\, abil^2$$

which has an intercept of about 11.9. I generated 2,000 values of *abil*, equally spaced between -5 and 6, to generate the following graph:



Its minimum is at roughly -4, as the calculation in part (iv). The slope of the function is increasing.