CH. 2

Natchaya Laokham 6104641425

1. Let *kids* denote the number of children ever born to a woman, and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is

$$kids = \beta_0 + \beta_1 educ + u,$$

where *u* is the unobserved error.

i. What kinds of factors are contained in *u*? Are these likely to be correlated with level of education?

Other factors that might be contained in the error term would be age of the mother, income level, religion. Surely age and income would be highly correlated with education.

ii. Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

If we ran the above regression, you wouldn't be controlling for any other effects (there are no other effects in the model: no other RHS variables) so you would not be looking at the ceteris paribus effect of education on fertility. A simple regression would tell the OVER-ALL effect of education on kids (controlling for nothing else at all)

4. The data set BWGHT contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression was estimated using data on $n = 1{,}388$ *births*:

$$\widehat{bwght} = 119.77 - 0.514\, cigs$$

i. What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.

The predicted birth weight is 119.77 when $Cigs = 0$,
then predited birth weight $= 119.77 - 0.514(20) = 109.49$
Therefore, the effect of smoking on additional 20 cigs will cause an estimated $119.77 - 109.49 = 10.28$ ounces decrease in the infant birth weight.

ii. Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

ก่อ.สาเหตุ (original)

As the amount of cigarettes is the independent variables and the infants weight is dependent. It implies that there is a causal effect. However, there are other factors to be considered as well like genetics or parent's weight. Smoking habits can partially explain the infant weight.

iii. To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.

$125 = 119.77 - 0.514\ Cigs$

$Cigs \approx -10.175$ which is clearly impossible!

The highest possible weight of infant according to the model would be its β_0 or intercept, which is 119.77 ounces.

iv. The proportion of women in the sample who do not smoke while pregnant is about .85. Does this help reconcile your finding from part (iii)?

ประนีประนอม

We know that it would be nicer, in terms of sample design, to have more smokers in the sample because this would add to the variation in the explanatory variable (cigs) and we might get a more accurate estimate of the intercept which gives us the average birthweight of children of non-smoking mothers. It is possible that the intercept estimate might take on a more accurate estimate in this case and indicate the greater likelihood of having larger babies.

CH. 3

1. Using the data in GPA2 on 4,137 college students, the following equation was estimated by OLS:

fixed

$$\widehat{colgpa} = 1.392 - .0135\, hsperc + .00148\, sat$$
$$n = 4{,}137,\ R^2 = .273,$$

where *colgpa* is measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (defined so that, for example, $hsperc = 5$ means the *top* 5% of the class), and *sat* is the combined math and verbal scores on the student achievement test.

i. Why does it make sense for the coefficient on *hsperc* to be negative?

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 is his/her expected college GPA.

ii. What is the predicted college GPA when $hsperc = 20$ and $sat = 1{,}050$?

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

iii. Suppose that two high school graduates, A and B, graduated in the same percentile (hsperc same) from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

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. Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of .50, or one-half of a grade point? Comment on your answer.

With hsperc fixed, $\Delta\widehat{colgpa} = .00148\,\Delta sat$

Now, to find Δsat such that $\Delta\widehat{colgpa} = .5$

$$.5 = .00148\,\Delta sat$$
$$\Delta sat = \frac{.5}{.00148} \approx 338$$

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.

2. The data in WAGE2 on working men was used to estimate the following equation:

$$\widehat{educ} = 10.36 - .094\, sibs + .131\, meduc + .210\, feduc$$
$$n = 722,\ R^2 = .214,$$

where *educ* is years of schooling, *sibs* is number of siblings, *meduc* is mother's years of schooling, and *feduc* is father's years of schooling.

i. Does *sibs* have the expected effect? Explain. Holding *meduc* and *feduc* fixed, by how much does *sibs* have to increase to reduce predicted years of education by one year? (A noninteger answer is acceptable here.)

Yes, because of budget constraints, it makes sense that, the more siblings there are in a family, the less education any one child in the family has.

To find the increase in the number of siblings that reduces predicted education by one year, we solve

$$1 = .094(\Delta sibs)$$
$$\Delta sibs = \frac{1}{.094} \approx 10.6 \ \#$$

ii. Discuss the interpretation of the coefficient on *meduc*.

Holding sibs and feduc fixed,
one more year of mother's education implies .131 years more of predicted to have about a half a year (.524) more years of education.

iii. Suppose that Man A has no siblings, and his mother and father each have 12 years of education. Man B has no siblings, and his mother and father each have 16 years of education. What is the predicted difference in years of education between B and A?

$$\text{Man A: } \widehat{educ} = 10.36 - \cancel{.094(0)} + .131(12) + .210(12) = 14.452$$

$$\text{Man B: } \widehat{educ} = 10.36 - \cancel{.094(0)} + .131(16) + .210(16) = 15.816$$

The predicted difference in education between B & A is $15.816 - 14.452 = 1.364$ #