6104640252
Jirathorn Chanthawekul

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?

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

i
- income , age, nationality
- Yes, they are correlated with education because people with higher income tends to have higher education

ii No, SLR 4 is failed because u is correlated with education

6104640252
Jirathorn Chanthawekul

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.

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

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

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)?

i

$cigs = 0$; $\widehat{bwght} = 119.77 - 0.514(0) = 119.77$

$cigs = 20$; $\widehat{bwght} = 119.77 - 0.514(20) = 109.49$

The baby that its mother smokes 20 cigarettes per day during the pregnancy has weight 10.28 ounces lower than the baby that its mother doesn't smoke any cigarettes while pregnancy

ii

- child's birth weight is a dependent variable
- mother's smoking habit is independent variable that will explain the child's birth weight

The more number of cigarettes mother smokes per day during pregnancy leads to the lower the infant birth weight.

However, there're other factors that affect child's birth weight such as the food that mother eats while pregnancy.
So, the number of cigarettes smoked by mom is partially explain the child's birth weight

6104640252
Jirathorn Chanthawekul

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

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

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?

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

iii. Suppose that two high school graduates, A and B, graduated in the same percentile 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?

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.

i If the hsperc are low, it means that you are the top of the class and if you are top of the class it is likely that you are going to get high colgpa.

ii
$$colgpa = 1.392 - 0.0135(20) + 0.00148(1050) = 2.676$$

iii same hsperc SAT score : $A = X + 140$, $B = X$

$$\widehat{colgpa}_A = 1.392 - 0.0135\, hsperc + 0.00148(X + 140)$$
$$\widehat{colgpa}_B = 1.392 - 0.0135\, hsperc + 0.00148(X)$$

different between $\widehat{colgpa}_A$ and $\widehat{colgpa}_B$ is

$$\widehat{colgpa}_A - \widehat{colgpa}_B = \cancel{1.392} - \cancel{0.0135\, hsperc} + 0.00148(\cancel{X} + 140) - \cancel{1.392} + \cancel{0.0135\, hsperc} - \cancel{0.00148(X)}$$
$$= 0.2072$$

A is predictable to have more 0.2072 in colgpa than B

IV difference in $\widehat{colgpa} = 0.5$, hsperc fixed

$$0.5 = \cancel{1.392} - \cancel{0.0135\, hsperc} + 0.00148\, SAT_1 - \cancel{1.392} + \cancel{0.0135\, hsperc} - 0.00148\, SAT_2$$
$$0.5 = 0.00148\,(SAT_1 - SAT_2)$$
$$\therefore \underbrace{SAT_1 - SAT_2}_{\text{difference in SAT score}} = 337.837$$

Students who have a 337.837 point different in SAT score will have the difference of 0.5 of their grade point

6104640252
Jirathorn Chanthawekul

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.)

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

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?

i Holding meduc and feduc fixed

reduce educ by 1 yr = $\widehat{educ}_1 - \widehat{educ}_2 = 1$

$\therefore \widehat{educ}_1 - \widehat{educ}_2 = (\cancel{10.36} - 0.094 sibs_1 + \cancel{0.131 meduc} + \cancel{0.210 feduc})$

$\qquad - (\cancel{10.36} + 0.094 sibs_2 - \cancel{0.131 meduc} - \cancel{0.210 feduc}$

$1 = (sibs_1 - 0.94 sibs_2)$

$-10.638 = sibs_1 - sibs_2$

So, sibs has to increase by 10.638 in order to reduce 1yr of educ

ii 0.131 meduc means if mother has 1 yr more in educ, the year of schooling will increases by 0.131 year

iii A : sibs = 0 , meduc = 12 , feduc = 12

B : sibs = 0 , meduc = 16 , feduc = 16

predited difference year of educ between B and A = $\widehat{educ}_B - \widehat{educ}_A$

$\widehat{educ}_B - \widehat{educ}_A = (\cancel{10.36} - \cancel{0.094 sibs_A} + 0.131(16) + 0.210(16))$

$\qquad - (\cancel{10.36} + \cancel{0.094 sibs_B} - 0.131(12) - 0.210(12))$

$\widehat{educ}_B - \widehat{educ}_A = 2.096 + 3.36 - 1.572 - 2.52$

$\widehat{educ}_B - \widehat{educ}_A = 1.364$

$\therefore$ The predicted difference year of educ between B and A is 1.364 years.