6109641177 Napakarn Thongyai

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) error terms maybe anything such as ages, income or names and these kind of errors is unlikely to correlated with level of education.

ii) SLR 4 fails, due to the fact that u is correlated with education.

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) If $cigs = 0$, $\widehat{bwght} = 119.77$. It means that mom smokes 0 number of cigarette then expected weight of baby would be 119.77.

when mumy smoke 20 Cigarettes, $\widehat{bwght} = 109.49$. As Cigs increase from 0 to 20, $\widehat{bwght}$ decrease from 119.77 to 109.49.

ii) As an increase in one Cig, it decreases expected weight of baby by 0.514.

iii) $125 = 119.77 - 0.514$ Cigs

$\text{Cigs} = -10.1751$

→ It means that mother has to smoke -10.751 cigarettes. In order to gain weight 125 ounces. However, it does not make sense that one can consume negative value. ∴ This regression fn. is not enough to measure, there must be other factor. In order to use this function precisely.

iv) No, it does not give any useful information to make this function more precise.

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

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

$1 \rightarrow 0.00148$

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) it would make sense because, if you get a higher score, hsperc would be low, and has high colgpa. ∴ Coefficient will be negative.

ii) $\widehat{colgpa} = 1.392 - 0.135(20) + 0.00148(1{,}050)$

$= 0.246$ *

iii) If A, B graduated in the same percentile.

If SAT change by 1 : $\frac{\partial\, colgpa}{\partial\, Sat}$ would change by 0.00148

∴ student A's SAT score is higher by 140 : $0.00148 \cdot 140$

$= 0.2072$

iv) $0.5 = 0.00148\, SAT$

$SAT = 337.837838$

∴ 0.5 point means different in SAT score by 337.837838 *

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 & feduc is fixed

if education change by 1 year.

$Sibs = 10.638$

ii) If mother's years of schooling increase by 1, the predicted year of school would increase by 0.131 *

iii) Man A predicted year of education

$= 10.36 - 0.094(0) + 0.131(2) + 0.210(12)$

$= 10.36 + 1.572 + 2.52 = 14.452$

Man B predicted year of education

$= 10.36 - 0.94(0) + 0.131(16) + 0.210(16)$

$= 10.36 + 2.096 + 3.36 = 15.816$

$\therefore$ The predicted different is 1.364 *