6104640260

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.

1) i). u is an error term, or other factors that could affect a number of children born to a woman aside from X or the years of education.

- Factors that could be in u are age, relationship status, income.
- These are likely to be correlated with years of education e.g. income is correlated with years of education

ii) SLR 4: Zero conditional mean

$E(u|X)=0$, $E(u)=0$

In this case, SLR 4 is likely to be violated. But in simple regression, we assume ceteris paribus (other factors fixed)

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) when cigs $= 0$

bwght $= 119.77 - 0.514(0) = 119.77$

when cigs $= 20$

bwght $= 119.77 - 0.514(20) = 109.49$

When a mother smokes 20 cigarettes per day on average, the baby's weight drops by 10.28 ounces

ii) cigs in an independent variable, bwght is an independent variable so this simple regression offers a casual relationship between 2 variables

iii) $125 = 119.77 - 0.514\,cigs$

$-10.175 = cigs$

This doesn't make sense as the number of cigarettes can't be negative. As the y intercept equals to 119.77 ounces, the maximum that a child can weigh is 119.77

iv) Since 85% of women in the sample don't sample, we may need to collect more data on smokers

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) It makes sense for the coefficient on hsperc to be negative. A 5% percentile means top 5% of class, the higher the percentile, the lower rank you are in class. Therefore, the lower the percentile, the more likely you are to have a high GPA.

ii) $\widehat{colgpa} = 1.392 - 0.0135(20) + 0.00148(1050)$
$= 2.676$

iii) $\frac{\partial\, colgpa}{\partial\, sat} = 0.00148$

→ This means if SAT score increases by 1 point, college GPA increases by 0.2072 (0.00148×140)

iv) $1 = 0.00148 SAT$
$0.50 = \frac{1}{0.00148} \times 0.5 = 337.838$

Holding other factors fixed, an increase of 337.838 GPA points increases GPA by 0.5

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.

Print Preview

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) Yes, the number of siblings negatively affects the number of years in education.

$\frac{\partial \widehat{educ}}{\partial sibs} = 0.94$ → if the number of siblings increases by 1, the number of years in education reduces by 0.94

$1 = -0.94 sibs$

$\frac{1}{-0.94} = sibs$

$-1.06 = sibs$

ii) The coefficient on meduc is positive. If mother's year of schooling increases by 1, the child's year of schooling increases by 0.131.

iii) Man A : $\widehat{educ} = 10.36 - 0.94(0) + 0.131(12) + 0.210(12)$

$= 14.452$

Man B $\widehat{educ} = 10.36 - 0.94(0) + 0.131(16) + 0.210(16)$

$= 15.816$

The difference is 1.364 years of education.