EE325: Homework3

6104641060 Suchaya Ch.

Chapter 2

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) u is the error terms of the observation, age are contained in u and correlated with educ. It means that this equation cannot be BLUE as violating SLR4 $E(u_i|x_i) = 0$ which is not be true in this equation.

(ii) Ceteris paribus is holding other things constant, only education cannot explain well about the fertility.

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

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

The effect of smoking an additional 20 ciggarettes will cause an estimated $(119.77 - 109.49)$ decrease in the birth weight.

(ii) As the amount of cigarettes is independent variable and the infant weight is dependent, it implies that there is a casual effect. However, there are other factor to be considered as well like genetics or parent's weight. Smoking habits can partially explain a baby's weight

(iii) $125 = 119.77 - 0.514(cigaretts)$

$\frac{125 - 119.77}{-(0.514)} = -10.175$ cigaretts

which is not possible. The highest possible weight according to the model would be its β_0 or interest, which is 119.77 ounces.

(iv) To get an accurate sample, we should consider more smokers to get more data points which will make our regression less limited and more accurate. We need more sample variation in the explanatory variable or SLR3 because 85% of the sample smoke cigaretts = 0. Moreover, SLR2 means we need the sample to be a better representative. Thus, we can improve our regression function so it is not as limited and will give us possibly inaccurate data.

6104641060 Suchaya Ch.

Question 1

(i) Yes, it does make since because if you are top of class, the "hsperc" will be low, and top class usually get high "colgpa". Hence, the coefficient is negative.

(ii) $\widehat{colgpa} = 1.392 - 0.0135\ hsperc + 0.00148\ sat$

→ Thus, increasing in SAT score for 1 mark will make the value of $\widehat{colgpa}$ changes for 0.00148

→ $\frac{\partial \widehat{colgpa}}{\partial sat} = 0.00148$

→ A is predicted to have a score of $0.207 = (140)(0.00148)$ higher. It is quite large if we consider the s.d.

(iv) $0.5 = 0.00148(sat)$

$sat = 337.8378$

Question 2

(i) Yes, the higher the number of year, the lower the year schooling. This is because of the budget constrain

→ $1 = -0.094\ sibs$

→ $sibs = 10.638$ of reduce years of education.

(ii) If mother's years of schooling ↑ by 1, the predicted years of school would measure by 131

(iii) Man A predicted years of education

$= 10.36 - 0.94(0) + 0.131(2) + 0.210(12)$

$= 10.36 + 1.572 + 2.52$

$= 14.452$

Man B predicted years of education

$= 10.36 - 0.094(0) + 0.131(16) + 0.210(16)$

$= 15.816$

∴ Predicted different is 1.364