pavita kriathkungwalkai
6104640674

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 term of the observation, Let's say that age are contained in u and correlated with education. So this equation cannot be BLUE as violating SLR4 E(ui|xi)=0 which is not be true in this equation.
ii) ceteris parisbus 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) $\widehat{bwght}$ = 119.71 - 0.514(0) = 119.77
Swght = 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 casval effect. However, there are other factor to be considered as well like genetics or parent's weight. Smoking habit's can partially explain a baby's weight.
iii) 125 = 119.77 - 0.514 cigs
cigs = -10.1751, which is not possible. The highest possible weight according to the model would be its β0 or interest, which is 119.77 ounces
iv) To get an accurate sample, we should consider more smokers to get more data points variation in the explainatory variable or SLR3 b/c 85% of the sample smoke cigs=0
In addition, SLR 2 means we need the sample for be a better representative. Thus, we can improve our regression fn so it is not as limited and will give us possibly inaccruate data

CHAPTER 3

Q1) i) Yes, it does make sense b/c if you are top of the class, and top class usually get high 'colgpa' Hence, the coefficient is negative.

ii) $\widehat{colgpa} = 1.392 - 0.135(20)P.00148(1050)$

$= 1.392 - 0.27 + 1.554$

$= 2.676$

iii) $\widehat{colgp} = 1.392 - 0.0135\ hsperc + 0.00148\ sat$

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

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

→ 14 is predited to have a score sat at 0.207 $= (140)(0.00148)$ higher.

It is avite large if we consider the s.d.