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. Factors that might be contained in the error term like religion, occupation, and age. Yes, level of education are correlated to occupation and age.

ii With the simple regression, other variable or the omitted X variable will be represent as u. Resulting in, when finding the effect of education on kids, as it does not hold other factor constant, it will find the overall effect of eauction on kids but the impact of only education on kids

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)?

4. (i.) when cigs = 0 $\widehat{bwght} = 119.77 - 0.514(0)$
$= 119.77$

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

As more cigarettes the mother smoked per day during pregnancy increase, infant birth weight tend to decrease.

(ii) No, as there are other factors that would affect a child's birth weight. Like the mother's health and this correlate to the mother smoking.

(iii) $125 = 119.77 - 0.514\, cigs$
$125 - 119.77 = -0.514\, cigs$
$cigs = \frac{5.23}{-0.514} = -10.175$

cigs should be -10.175 per day but it is imposible to smoke negative cigs.

(iv) The proportion of women who smoke during pregnancy is too small compare to not smoke one. Increase the sample of woman who smoke can increase accuracy.

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) Because the higher hsperc, the lower the rank in the class. This make GDP lower so, it should be negative

(ii) hsperc = 20 sat = 1050

$\widehat{colgpa} = 1.392 - .0135(20) + .00148(1050)$
$= 2.676$

(iii) The college GPA expected to be .00148(140) or 0.2072 or it is 140 times the coefficient on sat.

(iv) $0.5 = 0.00148 sat$
$sat = 337.8378$
Holding hsperc fixed to have colgpa different of .50, a student should have 337.8378 score higher on SAT.

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) Yes, because it might cause budget constraint holding other factors constant
$-1 = -.094 sibs$
$sibs = 10.6383$
to reduce 1 year of education. He might have to have 10 siblings.

(ii) As 1 year more of mother education, .131 year of son education increase.

(iii) A sibs = 0 meduc = 12 feduc = 12

$\widehat{educ}_A = 10.36 - .094(0) + .131(12) + .210(12) = 14.452$

B sibs = 0 meduc = 16 feduc = 16

$\widehat{educ}_B = 10.36 - .094(0) + .131(16) + .210(16) = 15.816$

A is predicted to have about 1.4 less education than B.