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) It would contain wealth, age, housing.
Using these example, age & wealth would be correlated to level of edu

ii) With the simple regression, other variable or the omitted x variable will be represented 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 education on kids but the true 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)?

i) cigs = 0, $\widehat{bwght}$ = 119.77
cigs = 20, $\widehat{bwght}$ = (119.77 - 0.514(20) = 109.49

ii) no, as there are other factors that would affect a child's birth weight. for example, the mother's health and this correlate to the mother's smoking habits.

iii) $\widehat{bwght}$ = 125

125 = (119.77 - 0.514 cig

cigs = -10.175,

the function shows a negative relationship between weight and cigs, with the intercept of 119.77, this mean that the max weight of the baby is supposed to be 119.77. Not to mention one could not smoke negative amount of cigarettes.

iv) This help reconcil that this function does not give an accurate impact of only smoking baby weight.

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) The GPA max out a 4.0. student who are the top of their class would receive a higher GPA than their peers. Thus as hsperc increase, the lower your standing in class, therefore lower GPA, and resulting in negative coefficient.

ii) hsperc = 20, sat = 1050

GPA = 2.678

iii) Suppose let A hsperc = 1 SAT = 140 & 0

A = 0.2072

B = 1.3765

1.1713

predicted diff = 1.1713

iv) 1.5 = 1.392 - 0.0135 + 0.00148 SAT

SAT = 82.0945

0.5 = 1.392 - 0.0135 + 0.00148 SAT

SAT = -593.58

Δ = 675 .6755

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) suppose a family only have a certain amount of money to lend their children to a year no. of schooling more children means that lesser money for each education.

ii) when the mother's year of schooling increase by 1, the year of schooling will increase by 0.131.

iii) $\widehat{educ}_A = (10.36 - 0.094(6) + 0.131(12) + 0.210(12)$

$= 14.452$

$\widehat{educ}_B = 10.36 - 0.094(6) + 0.131(16) + 0.210(16)$

$= 15.816$

$\Delta = 15.816 - 14.452 = 1.364$ years