Isara Sawasilp
6104640070

No. _____

Date: / /

Chapter 2

1i. u includes the other factors that influence the number of kids rather than only years of education such as career or societal norms or age. These are somewhat likely to be correlated with education because age or career are related to education.

1ii. No, because u will still impact the regression analysis, so the regression will only show the total effect, not the ceteris paribus effect. This is because we cannot control/hold the other factors constant. $E(u_i|x_i) \neq 0$, because u_i is somewhat related to x_i so, SLR4 fails.

4i. $\widehat{bwight} = 119.77 - 0.514(0) = 119.77$

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

The effect of smoking an additional 20 cigarettes will cause an estimated 10.28 ounce (119.77 - 109.49) decrease in the infant birth weight.

4ii. As the amount of cigarettes is the independent variable and the infant weight is dependent, it implies that there is a causal effect. However, there are other factors to be considered as well like genetics or parent's weight. Smoking habits can partially explain a baby's weight

4iii. $125 = 119.77 - 0.514 cigs$

$cigs = -10.175097\overline{28}$, which is not possible. The highest possible weight according to the model would be its β_0 or intercept, which is 119.77 ounces.

4iv. To get an accurate sample, we should consider more smokers to get a more data points which will make our regression less limited and more accurate. We need more sample variation in the explantory variable or SLR3 now because 85% of the sample smoke cig: 0. In addition, 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 like in 4iii.

Isara Suwansilp
6104640070

No. _____

Date: / /

Chapter 3

1i. It makes sense for hsperc's coefficient to be negative because as the percentile of the high school graduating class increases, it means that we are moving farther and farther away from being at the top of the class. Thus, as hsperc increases, colgpa should decrease because if we are in a larger percentile, it means we did not do well in high school, thus our colgpa is expected to be lower.

1ii. $\widehat{colgpa} = 1.392 - 0.0135(20) + 0.00148(1050)$

$\widehat{colgpa} = 2.676$

1iii. The predicted difference in college GPA is $0.00148(140) = 0.2072$, which is large relative to 6.

1iv. $0.5 = 0.00148(sat)$

$sat = 337.8378378$, thus a 337.8378378 difference of SAT scores will lead to a predicted 0.5 colgpa difference.

2i. Yes, sibs has the expected effect because when you have more siblings your family may not be able to afford to pay for more of your education or you may have to work instead of get an education to support your siblings.

$\widehat{educ} = -1 = -0.094\ sibs$

$sibs = 10.63829787$, thus the number of siblings will have to increase approximately by 11 to reduce predicted years of education by one year

2ii. meduc's coefficient means that if the mother's years of schooling will increase by 1, there is an approximate/estimated increase of 0.131 years of schooling for the working men.

2iii. $\widehat{educ_A} = 10.36 - 0.094(0) + 0.131(12) + 0.210(12) = 14.452$ years

$\widehat{educ_B} = 10.36 - 0.094(0) + 0.131(16) + 0.210(16) = 15.816$ years

Man B will have a predicted $1.364\ (15.816 - 14.452)$ more years of schooling than Man A.