Warm Up

1. Wooldridge (2015), C.11 (located at end of Appendix C)

Solution: Since $\overline{y} = .132$, $s = 1.27$, and $n = 400$, we have

$$\text{se}\left(\overline{y}\right) = .064.$$

The t statistic is

$$t = \frac{.132 - 0}{.064} \approx 2.08$$

The p-value is 0.019, which is less than 0.05. Therefore, we conclude that the average change in GPAs is statistically greater than zero.

Exercises

3. *Wooldridge*, 1.1

Solution: Part (i): Ideally, we could randomly assign students to classes of different sizes. That is, each student is assigned a different class size without regard to any student characteristics such as ability and family background. For reasons we will see in Chapter 2, we would like substantial variation in class sizes (subject, of course, to ethical considerations and resource constraints).

Part (ii): A negative correlation means that a larger class size is associated with lower performance. We might find a negative correlation because a larger class size actually hurts performance. However, with observational data, there are other reasons we might find a negative relationship. For example, children from more affluent families might be more likely to attend schoolswith smaller class sizes, and affluent children generally might score better on standardized tests. Another possibility is that, within a school, a principal might assign the better students to smaller classes. Or, some parents might insist their children to be placed in smaller classes, and these same parents tend to be more involved in their children's education.

Part (iii): Given the potential for confounding factors—some of which are listed in (ii)—finding a negative correlation would not be strong evidence that smaller class sizes actually lead to better performance. Some way of controlling for the confounding factors is needed, and this is the subject of multiple regression analysis.

Computer Exercises You should use `R` to complete these exercises. Any data set referred to in the question should be available in the `wooldridge` package in `R`. You do not need to turn in an R-script for these questions, but you are welcome to do so if you would like to.

4. *Wooldridge*, 1.C2 (Chapter 1, Computer Exercise 2)

Solution: Part (i): There are 1,388 observations in the sample. Tabulating the variable cigs shows that 212 women have $cigs > 0$.

Part (ii): The average of $cigs$ is about 2.09, but this includes the 1,176 women who did not smoke. Reporting just the average masks the fact that almost 85 percent of the women did not smoke. It makes more sense to say that the "typical" woman does not smoke during pregnancy; indeed, the median number of cigarettes smoked is zero.

Part (iii): The average of $cigs$ over the women with $cigs > 0$ is about 13.7. Of course, this is much higher than the average over the entire sample because we are excluding 1,176 zeros.

Part (iv): The average of $fatheduc$ is about 13.2. There are 196 observations with a missing value for $fatheduc$, and those observations are necessarily excluded in computing the average.

Part (v): The average and standard deviation of $faminc$ are about 29.027 and 18.739, respectively, but $faminc$ is measured in thousands of dollars. So, in dollars, the average and standard deviation are \$29,027 and \$18,739.

5. *Wooldridge*, 2.C3

Solution: Part (i): The estimated equation is

$$\widehat{sleep} = 3,586.4 - .151 totwrk$$

with $n = 706, R^2 = .103$.

The intercept implies that the estimated amount of sleep per week for someone who does not work is 3,586.4 minutes, or about 59.77 hours. This comes to about 8.5 hours per night.

Part (ii): If someone works two more hours per week, then $\Delta totwrk = 120$ (because $totwrk$ is measured in minutes), and so $\widehat{\Delta sleep} = -.151(120) = -18.12$ minutes. This is only a few minutes a night. If someone were to work one more hour on each of five working days, $\widehat{\Delta sleep} = -.151(300) = -45.3$ minutes, or about five minutes a night.

Cool Down

6. *Wooldridge*, 2.1.

Solution: Part (i): Income, age, and family background (such as number of siblings) are just a few possibilities. It seems that each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

Part (ii): Not if the factors we listed in part (i) are correlated with $educ$. Because we would like to hold these factors fixed, they are part of the error term. But if u is correlated with $educ$, then $E(u|educ) \neq 0$, and so $SLR.4$ fails.

References

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 6 ed.