

Additional Empirical Exercise 4.3

The data file **CollegeDistance** contains data from a random sample of high school seniors interviewed in 1980 and re-interviewed in 1986. In this exercise, you will use these data to investigate the relationship between the number of completed years of education for young adults and the distance from each student's high school to the nearest four-year college. (Proximity to college lowers the cost of education, so that students who live closer to a four-year college should, on average, complete more years of higher education.) A detailed description is given in **College Distance_Description**, also available on the Web site.¹

- a. Run a regression of years of completed education (ED) on distance to the nearest college ($Dist$), where $Dist$ is measured in tens of miles. (For example, $Dist = 2$ means that the distance is 20 miles.) What is the estimated intercept? What is the estimated slope? Use the estimated regression to answer this question: How does the average value of years of completed schooling change when colleges are built close to where students go to high school?
- b. Bob's high school was 20 miles from the nearest college. Predict Bob's years of completed education using the estimated regression. How would the prediction change if Bob lived 10 miles from the nearest college?
- c. Does distance to college explain a large fraction of the variance in educational attainment across individuals? Explain.
- d. What is the value of the standard error of the regression? What are the units for the standard error (meters, grams, years, dollars, cents, or something else)?

¹ These data were provided by Professor Cecilia Rouse of Princeton University and were used in her paper "Democratization or Diversion? The Effect of Community Colleges on Educational Attainment," *Journal of Business and Economic Statistics*, April 1995, 12(2): 217–224.