

## Additional Empirical Exercise 8.2

Use the data set **CollegeDistance** described in Empirical Exercise AEE4.3 to answer the following questions.

- a. Run a regression of  $ED$  on  $Dist$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (that is, from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (that is, from 60 to 70 miles), how are years of education expected to change?
- b. Run a regression of  $\ln(ED)$  on  $Dist$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (from 60 to 70 miles), how are years of education expected to change?
- c. Run a regression of  $ED$  on  $Dist$ ,  $Dist^2$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (from 60 to 70 miles), how are years of education expected to change?
- d. Do you prefer the regression in (c) to the regression in (a)? Explain.
- e. Consider a Hispanic female with  $Tuition = \$950$ ,  $Bytest = 58$ ,  $Incomehi = 0$ ,  $Ownhome = 0$ ,  $DadColl = 1$ ,  $MomColl = 1$ ,  $Cue80 = 7.1$ , and  $Stwmfg80 = \$10.06$ .
  - i. Plot the regression relation between  $Dist$  and  $ED$  from (a) and (c) for  $Dist$  in the range of 0 to 10 (from 0 to 100 miles). Describe the similarities and differences between the estimated regression functions. Would your answer change if you plotted the regression function for a white male with the same characteristics?
  - ii. How does the regression function (c) behave for  $Dist > 10$ ? How many observations are there with  $Dist > 10$ ?
- f. Add the interaction term  $DadColl \times MomColl$  to the regression in (c). What does the coefficient on the interaction term measure?
- g. Mary, Jane, Alexis, and Bonnie have the same values of  $Dist$ ,  $Bytest$ ,  $Tuition$ ,  $Female$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $Cue80$  and  $Stwmfg80$ . Neither of Mary's parents attended college. Jane's father attended college, but her mother did not. Alexis's mother attended college, but her father did not. Both of Bonnie's parents attended college. Using the regressions from (f):

- 
- i. What does the regression predict for the difference between Jane's and Mary's years of education?
  - ii. What does the regression predict for the difference between Alexis's and Mary's years of education?
  - iii. What does the regression predict for the difference between Bonnie's and Mary's years of education?
- h. Is there any evidence that the effect of *Dist* on *ED* depends on the family's income?
- i. After running all these regressions (and any others that you want to run), summarize the effect of *Dist* on years of education.