Additional Empirical Exercise 6.2

Using the data set **CollegeDistance** described in Empirical Exercise AEE4.3, carry out the following exercises.

a. Run a regression of years of completed education (*ED*) on distance to the nearest college (*Dist*). What is the estimated slope?

b. Run a regression of *ED* on *Dist*, but include some additional regressors to control for characteristics of the student, the student's family, and the local labor market. In particular, include as additional regressors *Bytest, Female, Black, Hispanic, Incomehi, Ownhome, DadColl, Cue80,* and *Stwmfg80*. What is the estimated effect of *Dist* on *ED*?

c. Is the estimated effect of *Dist* on *ED* in the regression in (b) substantively different from the regression in (a)? Based on this, does the regression in (a) seem to suffer from important omitted variable bias?

d. Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and $\bar{R}^2$. Why are the R^2 and $\bar{R}^2$ so similar in regression (b)?

e. The value of the coefficient on *DadColl* is positive. What does this coefficient measure?

f. Explain why *Cue80* and *Swmfg80* appear in the regression. Are the signs of their estimated coefficients (+ or −) what you would have believed? Interpret the magnitudes of these coefficients.

g. Bob is a black male. His high school was 20 miles from the nearest college. His base-year composite test score (*Bytest*) was 58. His family income in 1980 was $26,000, and his family owned a home. His mother attended college, but his father did not. The unemployment rate in his county was 7.5%, and the state average manufacturing hourly wage was $9.75. Predict Bob's years of completed schooling using the regression in (b).

h. Jim has the same characteristics as Bob except that his high school was 40 miles from the nearest college. Predict Jim's years of completed schooling using the regression in (b).