

Empirical Exercises for Chapter 6

In this exercise you will investigate the relationship between the number of completed years of education for young adults and the distance from their high school to the nearest four-year college.

On the Web site you will find the data file **CollegeDistance** (in Excel and Stata formats). This file contains information on a random sample of high school seniors in 1980. These same students were interviewed again in 1986 to determine how many years of education they had completed. A detailed description is given in **CollegeDistance_DataDescription**, available on the Web site.

Use these data to answer the following questions.

1. Run a regression of ED on $Dist$, $Tuition$, $Black$, $Hispanic$, $Incomehi$, $Ownhome$, $DadColl$, $MomColl$, $Cue80$ and $Stwmfg80$. If $Dist$ increases from 2 to 3 (that is, 20 to 30 miles), how are years of education expected to change? If $Dist$ increases from 6 to 7 (that is, 60 to 70 miles), how are years of education expected to change?
2. Run a regression of $\ln(ED)$ on $Dist$, $Tuition$, $Black$, $Hispanic$, $Incomehi$, $Ownhome$, $DadColl$, $MomColl$, $Cue80$ and $Stwmfg80$. If $Dist$ increases from 2 to 3 (that is, 20 to 30 miles), how are years of education expected to change? If $Dist$ increases from 6 to 7 (that is, 60 to 70 miles), how are years of education expected to change?
3. Run a regression of ED on $Dist$, $Dist^2$, $Tuition$, $Black$, $Hispanic$, $Incomehi$, $Ownhome$, $DadColl$, $MomColl$, $Cue80$ and $Stwmfg80$. If $Dist$ increases from 2 to 3 (that is, 20 to 30 miles), how are years of education expected to change? If $Dist$ increases from 6 to 7 (that is, 60 to 70 miles), how are years of education expected to change?
4. Do you prefer the regression in (3) to the regression in (1)? Explain.
5. Consider a Hispanic female with $Tuition = \$950$, $Bytest = 58$, $Incomehi = 0$, $Ownhome = 0$, $DadColl = 1$, $MomColl = 1$, $Cue80 = 7.1$, and $Stwmfg = \$10.06$.
 - (a) Plot the regression relation between $Dist$ and ED from (1) and (3), for $Dist$ in the range of 0 to 10 (that is, 0-100 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?
 - (b) How does the regression function (3) behave for $Dist > 10$? For how many observations is $Dist > 10$?
6. Add the interaction term $DadColl \times MomColl$ to the regression in (3). What does the coefficient on the interaction term measure?
7. Mary, Jane, Alexis, and Bonnie have the same values of $Dist$, $Tuition$, $Female$, $Black$, $Hispanic$, $Fincome$, $Ownhome$, $Cue80$ 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.

(a) What is the regression's prediction of the difference between Jane's and Mary's years of education?

(b) What is the regression's prediction of the difference between Alexis's and Mary's years of education?

(c) What is the regression's prediction of the difference between Bonnie's and Mary's years of education?

8. Is there any evidence that the effect of *Dist* on *ED* depends on the family's income?

9. After running all of these regressions (and any others that you want to run), summarize the effect of *Dist* on years of education.