

## Empirical Exercises for Chapter 5

In this exercise you will investigate the relationship between a worker's age and earnings.

On the Web site you will find the data file **CPS98** (in Excel and Stata formats). This file contains an extended version of the dataset for 1998 used in Table 3.1 of the text. It contains data on full-time, full-year workers, aged 25-34, with a high school diploma or B.A./B.S. as their highest degree. A detailed description is given in **CPS98 \_Description**, available on the Web site.

Use these data to answer the following questions.

1. Run a regression of average hourly earnings (*AHE*) on age (*Age*). What is the estimated intercept? What is the estimated slope?
2. Run a regression of *AHE* on *Age*, gender (*Female*), and education (*Bachelor*). What is the estimated effect of *Age* on earnings? Construct a 95% confidence interval for the coefficient on *Age* in the regression.
3. Are the results from the regression in (2) substantively different from the results in (1) regarding the effect of *Age* and *AHE*? Based on this, does the regression in (1) seem to suffer from important omitted variable bias?
4. Bob is a 26-year-old worker with a high school diploma. Predict Bob's earnings using the estimated regression in (2). Alexis is a 30-year-old worker with a college degree. Predict Alexis's earnings using the regression.
5. Compare the fit of the regression in (1) and (2) using the regression standard errors,  $R^2$  and  $\bar{R}^2$ . Why are the  $R^2$  and  $\bar{R}^2$  so similar in regression (2)?
6. Are gender and education important determinants of earnings? Test the null hypothesis that *Female* can be deleted from the regression. Test the null hypothesis that *Bachelor* can be deleted from the regression. Test the null hypothesis that both *Female* and *Bachelor* can be deleted from the regression.
7. A regression will suffer from omitted variable bias when two conditions hold. What are these two conditions? Do these conditions seem to hold here?