

## Answers to Empirical Exercises for Chapter 5

1. *Run a regression of average hourly earnings (AHE) on age (Age). What is the estimated intercept? What is the estimated slope?*

The estimated regression is

$$\widehat{AHE} = 4.75 + 0.31 \text{ Age}$$

(0.94) (0.032)

$$R^2 = 0.016 ; \bar{R}^2 = 0.016; SE = 6.85$$

The estimated intercept is \$4.75. The estimated slope is \$0.31/year.

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.*

The estimated regression is

$$\widehat{AHE} = 3.30 + 0.31 \text{ Age} - 2.49 \text{ Female} + 5.33 \text{ Bachelor}$$

(0.87) (0.029) (0.16) (0.17)

$$R^2 = 0.182 ; \bar{R}^2 = 0.182; SE = 6.25$$

A 95% confidence interval for the coefficient on Age is  $0.31 \pm 1.96 \times 0.029 = 0.25$  to  $0.37$ .

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?*

The coefficient on Age is the same to two decimal places in the regressions. This suggests that the regression in (1) does not suffer from omitted variable bias that can be explained by the omitted variables *Female* and *Bachelor*.

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.*

$$\text{Bob's Predicted Earnings} = 3.30 + (0.31 \times 26) - (2.49 \times 0) + (5.34 \times 0) = \$11.36$$

Alexis's Predicted Earnings =  $3.30 + (0.31 \times 26) - (2.49 \times 1) + (5.34 \times 1) = \$14.21$

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)?

The regression in (2) fits the data much better. The  $R^2$  in (1) is less than 2%, but it is 18% in (2). The standard error of the regression falls from \$6.85 in (1) to \$6.25 in (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.

Both *Female* and *Bachelor* enter the regression significantly. The  $t$ -statistic for the coefficient on *Female* is  $-15.43$ , and the  $t$ -statistic for the coefficient on *Bachelor* is  $31.21$ . Both are much larger in absolute value than the 1% critical value (2.58). The  $F$ -statistic testing that both coefficients are zero is 562.1, which is much larger than the 1% critical value of 4.61.

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?

From Key Concept 5.1:

For omitted variable bias to arise, two things must be true:

1.  $X$  must be correlated with the omitted variable; and
2. the omitted variable must be a determinant of the dependent variable,  $Y$ .

In this example, it appears that omitted variable bias relating the omitted variables *Female* and *Bachelor* in regression (1) is not important. Yet from (5), the variables *Female* and *Bachelor* are important determinants of *AHE*. Thus it seems that *Age* must not be correlated with *Female* and *Bachelor*. Indeed, the regression of *Age* onto *Female* and *Bachelor* has an  $R^2$  of .00006, suggesting a very small correlation.