Homework Set 4

Due on May 3rd, 2017

All questions are from the end-of-chapter exercises. The question numbers refer to those in the book. I highly recommend you reading the textbook and lecture notes before completing the homework questions. When reading the textbook, please pay attention to the sections on how to interpret the estimated coefficients.

Exercises

6.5 Data were collected from a random sample of 220 home sales from a community in 2003. Let *Price* denote the selling price (in \$1000), *BDR* denote the number of bedrooms, *Bath* denote the number of bathrooms, *Hsize* denote the size of the house (in square feet), *Lsize* denote the lot size (in square feet), *Age* denote the age of the house (in years), and *Poor* denote a binary variable that is equal to 1 if the condition of the house is reported as "poor". An estimated regression yields

$$\begin{aligned}\widehat{Price} =& 119.2 + 0.485BDR + 23.4Bath + 0.156Hsize + 0.002Lsize \\ &+ 0.090Age - 48.8Poor,\ \bar{R}^2 = 0.72,\ SER = 41.5\end{aligned}$$

a. Suppose that a home owner converts part of an existing family room in the house into a new bath room. What is the expected increase in the value of the house?

b. Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?

c. What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

d. Compute the R^2 for the regression.

6.6 A researcher plans to study the causal effect of police on crime using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

b. Use your answer to (a) and the expression for omitted variable bias given in Equation (6.1) to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. (That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?)

6.10 (Y_i, X_{1i}, X_{2i}) satisfy the assumptions in Key Concept 6.4; in addition, $\text{Var}(u_i|X_{1i}, X_{2i}) = 4$ and $\text{Var}(X_{1i}) = 6$. A random sample of size $n = 400$ is drawn from the population.

a. Assume that X_1 and X_2 are uncorrelated. Compute the variance of $\hat{\beta}_1$. (*Hint*: Look at Equation (6.17) in the Appendix 6.2)

b. Assume that $\text{corr}(X_1, X_2) = 0.5$. Compute the variance of $\hat{\beta}_1$.

c. Comment on the following statements: "When X_1 and X_2 are correlated, the variance of $\hat{\beta}_1$ is larger than it would be if X_1 and X_2 were uncorrelated. Thus if you are interested in β_1, it is best to leave X_2 out of the regression if it is correlated with X_1."

6.11 (Require calculus) Consider the regression model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

for $i = 1, \ldots, n$. (Notice that there is no constant term in the regression.)

a. Specify the least squares function that is minimized by OLS.

b. Compute the partial derivatives of the objective function with respect to b_1 and b_2.

c. Suppose $\sum_{i=1}^{n} X_{1i}X_{2i} = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^{n} X_{1i}Y_i / \sum_{i=1}^{n} X_{1i}^2$.

d. Suppose $\sum_{i=1}^{n} X_{1i}X_{2i} \neq 0$. Derive an expression for $\hat{\beta}_1$ as a function of the data (Y_i, X_{1i}, X_{2i}), $i = 1, \ldots, n$.

e. Suppose that the model includes an intercept: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Show that the least squares estimators satisfy $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$.

f. As in (e), suppose that the model contains an intercept. Also suppose that $\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) / \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2$. How does this compare to the OLS estimator of β_1 from the regression that omit X_2?

Empirical Exercises

E6.1 Using the data set **TeachingRatings** described in Empirical Exercise 4.2, carry out the following exercise.

a. Run a regression of `Course_Eval` on `Beauty`. What is the estimated slope?

b. Run a regression of `Course_Eval` on `Beauty`, including some additional variables to control for the type of course and professor characteristics. In particular, include as additional regressors `Intro`, `OneCredit`, `Female`, `Minority`, and `NNEnglish`. What is the estimated effect of `Beauty` on `Course_Eval`? Does the regression in (a) suffer from important omitted variable bias?

c. Estimate the coefficient on `Beauty` for the multiple regression model in (b) using the three-step process in Appendix 6.3 (the Frisch-Waugh theorem). Verify that the three-step-process yields the same estimated coefficient for `Beauty` as that obtained in (b).

d. Professor Smith is a black male with average beauty and is a native English speaker. He teaches a three-credit upper-division course. Predict Professor Smith's course evaluation.