ØAMET4100 · Spring 2019
Worksheet 3A

Instructor: Fenella Carpena

January 24, 2019

1 Omitted Variable Bias

Exercise 1.1 At Banana Corporation, a fictional computer manufacturing company, the process of hiring manufacturing workers involves a written test. A human resources manager at the company is interested in understanding whether this written test is a reliable predictor of job performance. To do so, the manager drew a random sample of 500 workers. She compiles data on two variables: the worker's test score and the percentage of non-defective computers the worker produced in the last week.

(a) Write down the population regression equation that the manager would like to estimate using the sample data she collected.

(b) Consider the regression you wrote in part (a). Will $\widehat{\beta}_1$ be an unbiased estimator? Why or why not? If you think $\widehat{\beta}_1$ is biased, is it biased up or down? Why?

Exercise 1.2 (Stock & Watson, Review the Concepts 6.1) A researcher is estimating the effect of studying on the test scores of students from a private school. She is concerned, however, that she does not have information on the class size to include in the regression. What effect would the omission of the class size variable have on her estimated coefficient on the private school variable? Will the effect of this omission disappear if she uses a larger sample of students?

2 Multiple Regression Model

Exercise 2.1 (Stock & Watson, Review the Concepts 6.2) A multiple regression includes two regressors: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

(a) What is the expected change in Y if X_1 increases by 8 units and X_2 is unchanged?

(b) What is the expected change in Y if X_2 decreases by 3 units and X_1 is unchanged?

(c) What is the expected change in Y if X_1 increases by 4 units and X_2 decreases by 7 units?

Exercise 2.2 Data were collected from a random sample of 220 home sales from a community in 2013. 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

$$\widehat{Price} = 119.2 + 0.485BDR + 23.4Bath + 0.156Hsize + 0.002Lsize + 0.090Age - 48.8Poor$$

(a) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. 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"?

3 $\overline{R}^2$ (Adjusted R^2)

Exercise 3.1 (Stock & Watson, Review the Concepts 6.3) What are the measures of fit that are commonly used for multiple regressions? How can an adjusted R^2 take on negative values?

Exercise 3.2 Consider the regression $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Does a higher R^2 mean that it is more likely that X_{1i} causes Y_i? What about a higher $\overline{R}^2$?

4 Perfect and Imperfect Multicollinearity

Exercise 4.1 In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours the spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168.

(a) In the model $GPA_i = \beta_0 + \beta_1 study_i + \beta_2 sleep_i + \beta_3 work_i + \beta_4 leisure_i + u_i$, does it make sense to hold *sleep*, *work* and *leisure* fixed, while changing *study*?

(b) Explain why this model violates the fourth least squares assumption.

(c) How could you reformulate the model so that its parameters have a useful interpretation and it satisfies the fourth least squares assumption? Provide an interpretation of one of the parameters in your proposed model.

Exercise 4.2 Suppose we want to estimate the effect campaign spending on campaign outcomes. Let $voteA_i$ denote the percent of the vote for candidate A in county i, let $expendA_i$ denote the campaign expenditures for candidate A, let $expendB_i$ denote the campaign expenditures for candidate B. Let $totexp_i$ be the total expenditures across both candidates, and let $shareA_i$ denote the percentage of total campaign expenditures made by candidate A (in other words $shareA_i = 100 \cdot (expendA_i/totexp_i)$. Consider the regression model

$$voteA_i = \beta_0 + \beta_1 expendA_i + \beta_2 expendB_i + \beta_3 shareA_i + u_i.$$

Does this regression violate the fourth least squares assumption? Explain why or why not.

Exercise 4.3 Explain why two perfectly multicollinear regressors cannot be included in a linear multiple regression. Give two examples of a pair of perfectly multicollinear regressors.

Exercise 4.4 Explain why it is difficult to estimate precisely the partial effect of X_1 holding X_2 constant, if X_1 and X_2 are highly correlated.

Exercise 4.5 (Stock & Watson, Review the Concepts 6.4) Explain the dummy variable trap and how is it related to multicollinearity of regressors? What is the solution for this form of multicollinearity?

Exercise 4.6 (Stock & Watson, Review the Concepts 6.5) How is imperfect multicollinearity of regressors different from perfect multicollinearity? Compare the solutions for these two concerns with multiple regression estimation.