Linda T.M. Bui

Sketch of Solutions for the First Six Problems.

1. Suppose that a score on a final exam depends upon attendance and unobserved factors that affect exam performance (such as student ability). Then,

$$score = \beta_0 + \beta_1 attendance + \epsilon.$$

When would you expect this model to satisfy the assumption $E(\epsilon \mid \text{attendance}) = 0$? (That is, when would you expect the conditional expectation of the unobserved error term to be zero?)

Answer: you would expect this to be the case if the unobserved variables that can affect exam performance (such as student ability) are NOT correlated with attendance. This seems pretty unlikely, though.

2. Let kids denote the number of children ever born to a woman, and let educ denote years of education for this woman. A simple model relating fertility to years of education is:

$$kids = \beta_0 + \beta_1 educ + \epsilon.$$

a. What kinds of factors are contained in ϵ? Are these likely to be correlated with the level of education?

b. Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

Answer:

a. Other factors that might be contained in the error term would be age of the mother, income level, religion. Surely age and income would be highly correlated with education.

b. If you ran the above regression, you wouldn't be controlling for any other effects (there are no other effects in the model: no other RHS variables) so you would NOT be looking at the ceteris paribus effect of education on fertility. A simple regression would tell you the OVER-ALL effect of education on kids (controlling for nothing else at all).

3. In the estimated linear consumption function: $\widehat{cons} = \hat{\beta}_0 + \hat{\beta}_1 income$ the (estimated) marginal propensity to consume (MPC) out of income is simply the slope, $\hat{\beta}_1$ and the average propensity to consume out of income (APC) is given by $\widehat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$. Using 100

families randomly chosen and taking their annual income and consumption data (both measured in dollars) the following equation is obtained:

$$\widehat{cons} = -124.84 + 0.853\,inc \quad R^2 = 0.692, \quad N = 100$$

a. Interpret the intercept in this equation, and comment on its sign and magnitude.

b. What is the prediction consumption when family income is $30,000?

Answer: a. Remember that in the bivariate model, the OLS estimator for the intercept term is given by: intercept = ybar - beta hat*xbar. So, if we have a negative intercept, and we assume that consumption MUST be at least zero, in our sample, the average income level must be at least 146.35. (This is one of the reasons we don't worry too hard about the sign of the intercept.) On the other hand, that's a pretty small average income level – some people in our sample must have an income level LOWER than that.

b. -124.84 + 0.853(30,000)

4. The OLS fitted line explaining college GPA in terms of high school GPA and ACT score is estimated as:

$$colGPA = 1.29 + 0.453\,hsGPA + 0.0094ACT$$

If the average high school GPA is about 3.4 and the average ACT score is about 24.2, what is the average college GPA in the sample?

Answer: 1.29 + 0.453(3.4) + 0.0094(24.2) = 3.06

5. Suppose there are two candidates for office and we wanted to explain the share of votes that candidate A gets and we model it as:

$$voteshareA = \beta_0 + \beta_1 expendA + \beta_2 expendB + shareA + \epsilon$$

Where expendA is the campaign expenditures spent by candidate A (and expendB is the campaign expenditures of candidate B), and shareA = expendA/totalexpend where totalexpend is the TOTAL amount of campaign expenditures put out by both candidates. Does this model violate the perfect multicollinearity assumption?

Answer: No. There is no perfect linear relationship between the explanatory variables.

6. Suppose that you postulate a model explaining final exam score in terms of class attendance.

Thus, the dependent variable is final exam score, and the key explanatory variable is number of classes attended. To control for student abilities and efforts outside the classroom, you include among the explanatory variables cumulative GPA, SAT score, and measures of highschool performance. Someone says, "You cannot hope to learn anything from this exercise because cumulative GPA, SAT score and highschool performance are likely to be highly collinear." What should be your response?

Answer: You're interested in how class attendance affects final exam scores. Even if GPA, SAT score and highschool performance are highly collinear (and therefore might have low t-statistics because OLS can't sort out their relative contribution to explaining the variation in final exam score), that will NOT affect the OLS estimator on class attendance. The OLS estimator will still be unbiased, as will the variance on the coefficient estimator. If we were to DROP those variables and they were correlated with attendance, then we'd really be in trouble: we'd have omitted variable bias, then.