Chapter 4: Multiple Regression Analysis: Inference Computer Exercises
Book Title: Introductory Econometrics
Printed By: Wanwiphang Manachotipong (wanwiphang@econ.tu.ac.th)

Chapter Review

Computer Exercises

C1. The following model can be used to study whether campaign expenditures affect election outcomes:

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u,$$

where *voteA* is the percentage of the vote received by Candidate A, *expendA* and *expendB* are campaign expenditures by Candidates A and B, and *prtystrA* is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

i. What is the interpretation of β_1 ?

ii. In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

iii. Estimate the given model using the data in VOTE1 and report the results in usual form. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part (ii)?

iv. Estimate a model that directly gives the *t* statistic for testing the hypothesis in part (ii). What do you conclude? (Use a two-sided alternative.)

C2. Use the data in LAWSCH85 for this exercise.

i. Using the same model as in Problem 4 in Chapter 3, state and test the null hypothesis that the rank of law schools has no ceteris paribus effect on median starting salary.

ii. Are features of the incoming class of students—namely, *LSAT* and *GPA* —individually or jointly significant for explaining *salary*? (Be sure to account for missing data on *LSAT* and *GPA*.)

iii. Test whether the size of the entering class (*clsize*) or the size of the faculty (*faculty*) needs to be added to this equation; carry out a single test. (Be careful to account for missing data on *clsize* and *faculty*.)

iv. What factors might influence the rank of the law school that are not included in the salary regression?

C3. Refer to Computer Exercise C2 in Chapter 3. Now, use the log of the housing price as the dependent variable:

$$\log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u.$$

i. You are interested in estimating and obtaining a confidence interval for the percentage change in *price* when a 150-square-foot bedroom is added to a house. In decimal form, this is $\theta_1 = 150\beta_1 + \beta_2$. Use the data in HPRICE1 to estimate θ_1 .

ii. Write β_2 in terms of θ_1 and β_1 and plug this into the $\log(price)$ equation.

iii. Use part (ii) to obtain a standard error for $\hat{\theta}_2$ and use this standard error to construct a 95% confidence interval.

C4. In Example 4.9, the restricted version of the model can be estimated using all 1,388 observations in the sample. Compute the *R*-squared from the regression of *bwght* on *cigs*, *parity*, and *faminc* using all observations. Compare this to the *R*-squared reported for the restricted model in Example 4.9.

C5. Use the data in MLB1 for this exercise.

i. Use the model estimated in equation (4.31) and drop the variable *rbisyr*. What happens to the statistical significance of *hrunsyr*? What about the size of the coefficient on *hrunsyr*?

ii. Add the variables *runsyr* (runs per year), *fldperc* (fielding percentage), and *sbasesyr* (stolen bases per year) to the model from part (i). Which of these factors are individually significant?

iii. In the model from part (ii), test the joint significance of *bavg*, *fldperc*, and *sbasesyr*.

C6. Use the data in WAGE2 for this exercise.

i. Consider the standard wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

State the null hypothesis that another year of general workforce experience has the same effect on $\log(wage)$ as another year of tenure with the current employer.

ii. Test the null hypothesis in part (i) against a two-sided alternative, at the 5% significance level, by constructing a 95% confidence interval. What do you conclude?

C7. Refer to the example used in Section 4-4. You will use the data set TWOYEAR.

i. The variable *phsrank* is the person's high school percentile. (A higher number is better. For example, 90 means you are ranked better than 90 percent of your graduating class.) Find the smallest, largest, and average *phsrank* in the sample.

ii. Add *phsrank* to (equation 4.26) and report the OLS estimates in the usual form. Is *phsrank* statistically significant? How much is 10 percentage points of high school rank worth in terms of wage?

iii. Does adding *phsrank* to (4.26) substantively change the conclusions on the returns to two- and four-year colleges? Explain.

iv. The data set contains a variable called *id*. Explain why if you add *id* to (equation 4.17) or (4.26) you expect it to be statistically insignificant. What is the two-sided *p*-value?

C8. The data set 401KSUBS contains information on net financial wealth (*nettfa*), age of the survey respondent (*age*), annual family income (*inc*), family size (*fsize*), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single-person households (so $fsize = 1$).

i. How many single-person households are there in the data set?

ii. Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u,$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

iii. Does the intercept from the regression in part (ii) have an interesting meaning? Explain.

iv. Find the *p*-value for the test $\mathrm{H}_0\colon \beta_2 = 1$ against $\mathrm{H}_1\colon \beta_2 < 1$. Do you reject H_0 at the 1% significance level?

v. If you do a simple regression of *nettfa* on *inc*, is the estimated coefficient on *inc* much different from the estimate in part (ii)? Why or why not?

C9. Use the data in DISCRIM to answer this question. (See also Computer Exercise C8 in Chapter 3.)

i. Use OLS to estimate the model

$$\log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 \log(income) + \beta_3 prppov + u,$$

and report the results in the usual form. Is $\widehat{\beta}_1$ statistically different from zero at the 5% level against a two-sided alternative? What about at the 1% level?

ii. What is the correlation between $\log(income)$ and *prppov*? Is each variable statistically significant in any case? Report the two-sided *p*-values.

iii. To the regression in part (i), add the variable $\log(hseval)$. Interpret its coefficient and report the two-sided *p*-value for $\mathrm{H}_0\colon \beta_{\log(hseval)} = 0$.

iv. In the regression in part (iii), what happens to the individual statistical significance of $\log(income)$ and *prppov*? Are these variables jointly significant? (Compute a *p*-value.) What do you make of your answers?

v. Given the results of the previous regressions, which one would you report as most reliable in determining whether the racial makeup of a zip code influences local fast-food prices?

C10. Use the data in ELEM94_95 to answer this question. The findings can be compared with those in Table 4.1. The dependent variable *lavgsal* is the log of average teacher salary and *bs* is the ratio of average benefits to average salary (by school).

i. Run the simple regression of *lavgsal* on *bs*. Is the estimated slope statistically different from zero? Is it statistically different from −1?

ii. Add the variables *lenrol* and *lstaff* to the regression from part (i). What happens to the coefficient on *bs*? How does the situation compare with that in Table 4.1?

iii. How come the standard error on the *bs* coefficient is smaller in part (ii) than in part (i)? (*Hint:* What happens to the error variance versus multicollinearity when *lenrol* and *lstaff* are added?)

iv. How come the coefficient on *lstaff* is negative? Is it large in magnitude?

v. Now add the variable *lunch* to the regression. Holding other factors fixed, are teachers being compensated for teaching students from disadvantaged backgrounds? Explain.

vi. Overall, is the pattern of results that you find with ELEM94_95 consistent with the pattern in Table 4.1?

C11. Use the data in HTV to answer this question. See also Computer Exercise C10 in Chapter 3.

i. Estimate the regression model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

by OLS and report the results in the usual form. Test the null hypothesis that *educ* is linearly related to *abil* against the alternative that the relationship is quadratic.

ii. Using the equation in part (i), test $\mathrm{H}_0\colon \beta_1 = \beta_2$ against a two-sided alternative. What is the *p*-value of the test?

iii. Add the two college tuition variables to the regression from part (i) and determine whether they are jointly statistically significant.

iv. What is the correlation between *tuit*17 and *tuit*18? Explain why using the average of the tuition over the two years might be preferred to adding each separately. What happens when you do use the average?

v. Do the findings for the average tuition variable in part (iv) make sense when interpreted causally? What might be going on?

C12. Use the data in ECONMATH to answer the following questions.

i. Estimate a model explaining *colgpa* to *hsgpa*, *actmth*, and *acteng*. Report the results in the usual form. Are all explanatory variables statistically significant?

ii. Consider an increase in *hsgpa* of one standard deviation, about .343. By how much does $\widehat{colgpa}$ increase, holding *actmth* and *acteng* fixed. About how many standard deviations would the *actmth* have to increase to change $\widehat{colgpa}$ by the same amount as a one standard deviation in *hsgpa*? Comment.

iii. Test the null hypothesis that *actmth* and *acteng* have the same effect (in the population) against a two-sided alternative. Report the *p*-value and describe your conclusions.

iv. Suppose the college admissions officer wants you to use the data on the variables in part (i) to create an equation that explains at least 50 percent of the variation in *colgpa*. What would you tell the officer?

Chapter 4: Multiple Regression Analysis: Inference Computer Exercises
Book Title: Introductory Econometrics
Printed By: Wanwiphang Manachotipong (wanwiphang@econ.tu.ac.th)