Tutorial exercises

Econometrics

2022/2023

All the tutorial exercises are from Wooldridge “Introductory Econometrics: A Modern Approach”, 5th edition.

The data for the computer exercises can be downloaded from:

<https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041>

where you have to select the relevant chapter, then download Data-Sets Stata (as a zip file).

**Week 1:**

Problems 1-6 (Ch.2), Introduction to STATA

**Week 2:**

Problems: 7, 8, 12 (Ch.2); Problem 1 (Ch.3)

**Week 3:**

Problems: 2-5, 7, 8 and 10 (Ch.3)

Computer Exercises: C2, C3 (Ch.2); C1, C2, C5, C6 (Ch.3)

Week 3 exercises not from Wooldridge:

**Problem A**: Consider the data from problem 3 from Chapter 2 of Wooldridge (5th edition).

1. Calculate the OLS estimators of the parameters in the simple linear regression model: GPA=ACT +u. Denote the OLS estimators by and .
2. Calculate the variance of the OLS estimators from a), conditional on the regressors
3. Assume the true model is the one given in a). Estimate by OLS the parameters of the simple linear regression model without intercept: GPA=ACT +u. Denote the OLS estimator by .
4. Assume that and . Calculate E() conditional on the regressors. Briefly comment your result.
5. Calculate the variance of the OLS estimator from c), conditional on the regressors. Compare this variance with the one in b). Briefly comment your results.
6. Calculate the R-squared for the model in a) and c). Briefly comment.

**Multiple choice questions:**

1. Data that have been collected on one or more variables at a single point in time is referred to as

(a) Cross-sectional data

(b) Time-cross-sectional data

(c) Time series data

(d) Panel data

2. Data that have both time series and cross-sections is referred to as

(a) Cross-sectional data

(b) Time-cross-sectional data

(c) Time series data

(d) Panel data

3. The linear relationship between two variables (*y* and *x*) can be represented by the equation. Which of the following statements is true?

(I) Parameter *a* is termed the intercept

(II) Parameter *a* is termed the slope

(III) Parameter *b* is termed the gradient

(IV) Parameter *b* is termed the constant

(a) I and IV only

(b) I and III only

(c) II and III only

(d) II and IV only

4. What does a positive linear relationship between *x* and *y* in a simple regression imply?

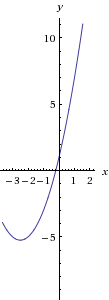
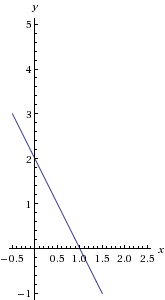
(a) Increases in the independent variable are usually accompanied by increases in the regressor

(b) The relationship between *x* and *y* cannot be explained by a straight line

(c) Decreases in the independent variable is usually accompanied by increases in the regressors

(d) Increases in the regressor are usually accompanied by increases in the dependent variable

5. Consider the following graphs:



(*A*) (*B*)

Which of the following statements is true?

(a) *A* is depicts a non-linear relationship between *y* and *x*

(b) *B* is depicts a linear relationship between *y* and *x*

(c) *A* and *B* depict linear and non-linear relationships between *y* and *x* respectively

(d) *A* and *B* depict non-linear and linear relationships between *y* and *x* respectively

6. Which of the following statements is true about graph (A) above?

(a) The intercept of the graph is positive and its slope is negative

(b) The intercept of the graph is negative and its slope is positive

(c) Both the intercept and slope of the graph are positive

(d) It is impossible to say anything about the intercept and slope without seeing the mathematical equation

7. Which one of the following is NOT an assumption of the classical linear regression model?

1. The explanatory variables are uncorrelated with the error terms.
2. The disturbance terms have zero mean
3. The dependent variable is not correlated with the disturbance terms
4. The disturbance terms are independent of one another.

8. What is the relationship, if any, between the normal and *t*-distributions?

1. A *t*-distribution with zero degrees of freedom is a normal
2. A *t*-distribution with one degree of freedom is a normal
3. A *t*-distribution with infinite degrees of freedom is a normal
4. There is no relationship between the two distributions.

9. Which one of the following is the most appropriate as a definition of *R*2 in the context that the term is usually used?

1. It is the proportion of the total variability of *y* that is explained by the model
2. It is the proportion of the total variability of *y* about its mean value that is explained by the model
3. It is the correlation between the fitted values and the residuals
4. It is the correlation between the fitted values and the mean.

10. Suppose that the value of *R*2 for an estimated regression model is exactly one. Which of the following are true?

1. All of the data points must lie exactly on the line
2. All of the residuals must be zero
3. All of the variability of *y* about is mean have has been explained by the model
4. The fitted line will be horizontal with respect to all of the explanatory variables
5. (ii) and (iv) only
6. (i) and (iii) only
7. (i), (ii), and (iii) only
8. (i), (ii), (iii), and (iv)

**Week 4:**

Problems: 1, 2, 6 (Ch. 4)

Computer exercises: C6 (Ch. 4)

**Week 5:**

Problems: 4, 8, 9 (Ch. 4)

Computer exercises: C1 (Ch. 4), C5 (Ch.6)

**Week 6:**

Problems: 10, 11 (Ch. 4)

**Problem B:** The parameter estimates obtained in a linear model and their standard errors (given between the brackets) are:

(0.316) (0.241)

The sample size is .

1. For each parameter calculate the *t*-statistic (t-ratio) that is used to test the significance of the parameter in a two-sided test.
2. Assume you are using a 5% significance level. Select the appropriate critical value from the table of critical values from the *t* distribution and decide whether or not you would reject the hypotheses tested in (a).
3. Calculate the 95% confidence interval for each parameter.
4. Using the confidence level approach (from (c)) carry out tests of the hypotheses that the parameters are zero against a two-sided alternative.
5. Is your answer in (b) similar to or different from your answer in (d)? Briefly explain why.

**Multiple choice questions:**

11. The following regression is estimated on 64 observations:

Which of the following null hypotheses could we test using an *F*-test?

(i) *β*2 = 0

(ii) *β*2 = 1 and *β*3 *+ β*4 = 1

(iii) *β*3*β*4 = 1

(iv) *β*2 -*β*3 -*β*4 = 1

(a) (i) and (ii) only

(b) (ii) and (iv) only

(c) (i), (ii), (iii) and (iv)

(d) (i), (ii), and (iv) only

12. If you are interested in conducting a multiple hypotheses test to determine whether  and  are each 1 for a regression , what would the restricted regression be?

(a) 

(b) 

(c) 

(d) 

13. What would the restricted regression be if you are interested in testing the null hypothesis  and  against the alternative hypothesis or  for a regression ,?

(a) 

(b) 

(c) 

(d) 

14. If the residuals of a regression on a large sample are found to be heteroscedastic which of the following might be a likely consequence?

(i) The coefficient estimates are biased

(ii) The standard error estimates for the slope coefficients may be too small

(iii) Statistical inferences may be wrong

(a) (i) only

(b) (ii) and (iii) only

(c) (i), (ii) and (iii)

(d) (i) and (ii) only

**Week 7:**

Problems: 1 and 4 (Ch. 6)

**Week 8:**

Problems: 1, 2, 4, 8, 9 (Ch. 7)

Computer problem: C2 (Ch. 7)

**Week 9:**

Problems: 1, 2, 3, 7 (Ch. 8)

Computer problem: C1 (Ch. 8)

**Week 10:**

Problems: 1 (Ch.10)

Computer problems: C6 (Ch. 10)

Stata exercise: Use hprice1.dta and regress *price* on *lotsize*, *sqrft* and *bdrms*. Compute the 90% confidence interval for the predicted (average) price (assuming homoskedasticity and heteroskedasticity) given the following characteristics: *lotsize*=2000, *sqrft*=1000 and *bdrms*=6.

**Week 11:**

Problems: 2, 3, 4, 5 (Ch. 11)

Problem: Consider the AR(1): , where is a white noise with variance and . Derive cov(). Hint: use the fact that implies cov()= cov().

**Week 12:**

Problems: 2, 4 (Ch. 12)

Computer exercises: C8 (Ch.12)

Problems: A) Compute the variance of the OLS estimator for the slope in a simple linear regression model when the errors are AR(1).

B) What are the steps for testing for AR(q) serial correlation in a multiple linear regression with k explanatory variables when the regressors are not strictly exogeneous?

C) Assume in a multiple linear regression model (with k regressors) the errors are AR(2). Describe the steps of the Cochrane-Orcutt approach to (feasible) GLS (ignoring the first two observations).

**Week 13**

See Revision exercises