Microeconometrics Using Stata

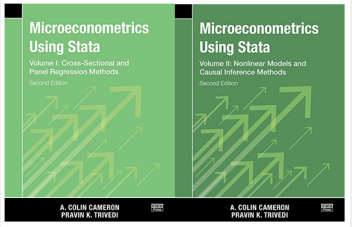
Linear regression basics: Exercises

Jhon R. Ordoñez [1] [2] [3]

[1] National University of San Cristóbal de Huamanga
[2] Faculty of Economic, Administrative and Accounting Sciences
[3] Professional School of Economics

Research Assistant

September 20, 2024



Outline

1 Exercise 1

2 Exercise 2

3 Exercise 3

4 Exercise 4

5 Exercise 5

6 Exercise 6

Exercise 1

Exercise 1

Exercise 1

Fit the model in **section 3.5** using only the first 100 observations. Compute standard errors in three ways: default, heteroskedastic, and cluster-robust where clustering is on the number of chronic problems. Use **estimates** to produce a table with three sets of coefficients and standard errors, and comment on any appreciable differences in the standard errors. Construct a similar table for three alternative sets of heteroskedasticity-robust standard errors, obtained by using the **vce(robust)**, **vce(hc2)**, and **vce(hc3)** options, and comment on any differences between the different estimates of the standard errors.

Exercise 2

Exercise 2

Exercise 2

Fit the model in **section 3.5** with robust standard errors reported. Test at 5% the joint significance of the demographic variables **age**, **female**, and **income**. Test the hypothesis that being male (rather than female) has the same impact on medical expenditures as aging 10 years. Fit the model under the constraint that $\beta_{phylim} = \beta_{actlim}$ by first typing **constraint 1 phylim = actlim** and then by using **cnsreg** with the **constraints(1)** option.

Exercise 3

Exercise 3

Exercise 3

Fit the model in **section 3.6**, and implement the RESET test manually by regressing y on x and $\hat{y}^2$, $\hat{y}^3$, and $\hat{y}^4$ and jointly testing that the coefficients of $\hat{y}^2$, $\hat{y}^3$, and $\hat{y}^4$ are 0. To get the same results as **estat ovtest**, do you need to use default or robust estimates of the **VCE** in this regression? Comment. Similarly, implement **linktest** by regressing y on $\hat{y}$ and $\hat{y}^2$ and testing that the coefficient of $\hat{y}^2$ is 0. To get the same results as **linktest**, do you need to use default or robust estimates of the **VCE** in this regression? Comment.

Exercise 4

Exercise 4

Fit the model in section 3.6, and perform the standard Lagrange multiplier test for heteroskedasticity by using **estat hettest** with $\mathbf{z} = \mathbf{x}$. Then, implement the test manually as 0.5 times the explained sum of squares from the regression of y_i^* on an intercept and z_i, where

$$y_i^* = \left\{ \frac{\hat{u}_i^2}{\frac{1}{N}\sum_j \hat{u}_j^2} \right\} - 1 \tag{1}$$

and $\hat{u}_i$ is the residual from the original **OLS regression**. Next, use **estat hettest** with the iid option, and show that this test is obtained as $N \times R^2$, where R^2 is obtained from the regression of $\hat{u}_i^2$ on an intercept and $\mathbf{z}_i$.

Exercise 5

Exercise 5

Using the **DGP** of **section 3.7.6**, generate a sample of size 100. Regress y on x and an intercept. Apply the **RESET** omitted variable test. Does the test indicate functional form misspecification? Next, use the **estat imtest** postestimation command to generate **IM** omnibus diagnostic test statistics. If you had applied only this test, what changes to your model specification would you make? Would you make additional tests before changing the model specification?

Exercise 6

Exercise 6

One type of data transformation that is sometimes used in linear regression, though not much in econometrics, is variable standardization. To standardize a variable you subtract the sample mean of that variable from each observation and divide the result by the sample standard deviation of that variable. The resulting variable with mean zero and standard deviation one is independent of the measurement scale of original data. A standardized regression is a regression in which the dependent and regressor variables are all standardized before running the regression. The standardized regression coefficients are interpreted as measuring the effect of one standard deviation change in the regressor and hence are in a sense comparable across variables. Note that when a regressor is discrete, interpreting its impact in units of standard deviation is not natural. Reestimate the regression equation of **section 3.5.2** after standardizing all variables. Are the insignificant coefficients the same as those in the previous regression? Which variable has the largest coefficient? Does that mean it is the most important variable in the regression?

References I


A. Colin Cameron, P. K. T. (2022a).
Microeconometrics Using Stata, Second Edition, Volume I: Cross-Sectional and Panel Regression Models.
Stata Press, 2 edition.


A. Colin Cameron, P. K. T. (2022b).
Microeconometrics Using Stata, Second Edition, Volume II: Nonlinear Models and Casual Inference Methods.
Stata Press, 2 edition.

Microeconometrics Using Stata

Linear regression basics: Exercises

Jhon R. Ordoñez [1] [2] [3]

[1] National University of San Cristóbal de Huamanga
[2] Faculty of Economic, Administrative and Accounting Sciences
[3] Professional School of Economics

Research Assistant

September 20, 2024