

ECOM20001: Econometrics 1

Tutorial 5: Single Linear Regression Estimation

A. Getting Started

Please create a Tutorial5 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Tutorial5 folder:

- [tute5.R](#)
- [tute5_height.csv](#)
- [tute5_growth.csv](#)

The first file is the R code for tutorial 5, the second two files are .csv files that contain two separate datasets for the tutorial.

The first (micro) dataset,¹ [tute5_height.csv](#), has the following 5 variables:

- **id**: worker identifier
- **earnings**: annual labour earnings in \$10,000's (in real terms, 2012=100)
- **height**: height without shoes in centimetres
- **weight**: weight without shoes in kilograms
- **male**: binary variable that equals 1 if worker is male and 0 otherwise
- **age**: age of the worker at time of survey

In total, the dataset contains this information for n=17,870 U.S. workers.

The second (macro) dataset,² [tute5_growth.csv](#), has the following 5 variables:

- **country**: country name
- **growth**: average annual percentage growth rate of real GDP (1960=100) from 1960-1995
- **rgdp60**: the value of GDP per capita in 1960 (in real terms, 1960=100)

¹ This dataset is from Case, Anne and Christina Paxson (2008): "Stature and Status: Height, Ability, and Labor Market Outcomes," *Journal of Political Economy*, 116(3), pp. 499-532.

² This dataset is from Beck, Thorsten and Norman Loayza (2000): "Finance and the Sources of Growth," *Journal of Financial Economics*, 58(1), pp. 261-300.

- **tradeshare**: the average share of annual trade in the economy from 1960 to 1995, measured as the sum of gross exports plus gross imports divided by nominal GDP; that is the average of $(X+M)/GDP$ from 1960 to 1995.

In total, the dataset contains this information for $n=65$ countries.

B. Go to the Code

With the R file downloaded into your Tutorial5 folder, you are ready to proceed with the tutorial. Please go to the [tute5.R](#) file to continue with the tutorial.

C. Questions

Having worked through the [tute5.R](#) code and graphs, please answer the following:

Earnings and Height

1. Discuss the summary statistics for the variables in the sample. What does a typical worker look like?
2. Present the scatter plot [q1_scat_height_earnings.pdf](#) and discuss whether there is a relationship visually between height and earnings.
3. Compute the difference in means for workers greater and less than 170cm, conduct a two-sample t-test of the null that the difference in means is 0, and report the 95% CI for the difference in means. Do the results provide evidence of a relationship between earnings and height?
4. Estimate the following single linear regression model for worker i :

$$Earnings_i = \beta_0 + \beta_1 Height_i + u_i$$

Provide an interpretation of the slope coefficient for a one-unit increase in height, and report the R-Squared and Standard Error of the Regression.

5. Provide an alternative interpretation for the regression results.
 - For a worker with average height, what is the impact of increasing their height by one standard deviation on earnings?
 - How does this predicted increase compare to the sample mean of earnings?
 - Is this a more relevant interpretation of the results than the interpretation given in question 4 above? Why or why not?

6. Run the earnings-height regression separately for males and females. Interpret the slope coefficients in both regressions and propose a possible economic explanation for your findings.

Trade and Growth

Note: these are extra practice exercises and will potentially only be partially covered in the tutorial, depending on time remaining. Solutions will be provided.

1. Discuss the summary statistics for the variables in the sample. What does a typical country look like?
2. Present the scatter plot [q2_scat_trade_growth3.pdf](#) and discuss whether there is a relationship visually between growth and trade.
 - note: these graphs in questions 2 and 3 are created at the bottom of the R code by combining the scatter plot command and regression results from question 4 below. See discussion in the comments of [tute5.R](#) for details.
3. There appears to be a potential outlier in the scatter plot from question 2 that has an average annual growth of 2.0. This country is Malta. The scatter plot [q2_scat_trade_growth4.pdf](#) drops this observation and is otherwise constructed identically to the graph in question 2. Does dropping Malta appear to have a large impact on the results?
4. Estimate the following single linear regression using the entire sample, and for a sample that omits Malta from the dataset.

$$Growth_i = \beta_0 + \beta_1 TradeShare_i + u_i$$

For each set of results, provide an interpretation of the slope coefficient for a one-unit increase in the trade share of a country, and report the R-Squared and Standard Error of the Regression.

5. Provide an alternative interpretation for the regression results based on a sample that drops Malta. For a country with average trade share, what is the impact of increasing their trade share by one standard deviation on growth? Is this effect large relative to the sample mean for growth? Provide a potential economic explanation for your findings.