

In this course we make extensive use of the statistical software RStudio, which is an Integrated Development Environment - IDE used to make working in R easier. RStudio is a free software, so you can install it to your own PC and it is also available on most PC clusters around the University. Please visit [our R Wiki](#) for the installation guide.

Alternatively, you could use MATLAB throughout the course and for your homework submission. Please visit [our MATLAB Wiki](#) for the help on MATLAB. Both languages are very similar. The major difference comes from the availability. At the University of Manchester we do have a license for MATLAB, but in general it is not a free software. MATLAB will be used in the **lectures** and to get the most out of this course we recommend that you try and replicate the MATLAB analysis presented in the lectures.

In the online **computer tutorials** we will mainly focus on R. It is important that you attempt the computer work yourself, rather than relying on the provided solutions. Actually looking at the data, running regressions and performing statistical tests is an important part of learning econometrics.

And remember that using the R or MATLAB commands, and interpreting their output will feature in the assessed coursework, which contributes 20% to your final grade.

The purpose of this session is to gently introduce you to the RStudio application; familiarizing yourself with the environment, loading some data from a file and running some basic commands to generate simulations and matrices.

Start by downloading the `attend.csv` file from the "Computer Tutorial 1" section on the Course Content page on Blackboard; save this file on your drive. Next, open the RStudio program (you may need to search for it in the Windows start menu).

Preparation

- Create a new folder with the name 'CT1' on your PC in which you will store the R script and the data.
- Open RStudio, create a new R script and save it in the new folder under the name 'CT1_code'.
- Download the file `attend.csv`¹ from the Blackboard and save it in 'CT1' folder.
- Before you start, you need to tell RStudio the 'working directory', thus the folder on the computer, which is the reference e.g. for importing data and saving plots. Open the R script 'CT1_code', then select in the menu bar 'Session' → 'Set Working Directory' → 'To Source File Location'.²

¹ The format '.csv' stands for 'comma separated value'. Right click on the file `attend.csv` and navigate to 'open with' → 'Notepad'. You see that the values are separated by commas.

² If you want to run a different script from a different folder, you need to open this other script and set the working Directory to the source file location of this other file. Alternatively, you can specify the working directory directly in the code via the line `setwd("C:/users/.../CT1")`. However, note that this line has to be changed, if you move the folder or open the script on another computer.

Basic plotting

- Make yourself familiar with the 'Help' documentation, either directly in RStudio or online.
- Create a vector 'x' with $x = (-10, -9.5, -9, \dots, 9, 9.5, 10)$. Use the function 'seq', you can read about it in the documentation. Then create a variable 'y' with $y = x^2$ and plot 'x' against 'y' as a line, not as a scatter plot. Again, you can read about this in the documentation of 'plot'.
- Simulate a random normal vector $x \sim \mathcal{N}(5, 10)$ of length $n = 1000$. Plot the histogram.
- Select a random subsample x_1 of length $n_1 = 100$, save its sample mean $\hat{\mu}_1 = \bar{x}_1$.
- Repeat d) 500 times and plot the histogram of sample means $\hat{\mu}_1, \dots, \hat{\mu}_{500}$. What is the variance of this distribution?

Data Import and Formats: Data Frame, Matrix and Numeric

The variables contained in 'attend.csv' are:

Variable	Definition
termgpa	GPA performance measure for term, between 0 (worst) and 4 (best) Note: "Pass" corresponds to gpa = 1.
priGPA	cumulative GPA prior to term
ACT	ACT score (university entrance test)
attend	percent classes attended
soph	= 1 if student is a sophomore (3rd year), = 0 else

- Import the data stored in 'attend.csv' as a data frame with name 'Data'. We want to explain the variation in termgpa by the available variables.

- Compute the covariance of termgpa and attend. Interpret.

Consider at first the linear regression model

$$termgpa_i = \alpha + \beta \cdot attend_i + u_i \quad (1)$$

- Fit model (1) and report the regression summary. Interpret.

- Verify that it holds $\hat{\beta} = \frac{\widehat{cov}(termgpa, attend)}{\widehat{var}(attend)}$. Explain.

Consider now the linear regression model

$$termgpa_i = \alpha + \beta_1 attend_i + \beta_2 ACT_i + \beta_3 priGPA_i + u_i \quad (2)$$

- Regress termgpa on attend, priGPA and ACT and report the regression summary. Interpret.
- Compute the coefficient estimates from model (2) directly via the OLS formula.
- Create a figure with three subplots. In the first, plot the predicted against the actual values of termgpa, in the second, plot the residuals and in the third, plot the residual histogram.

Exercises to solve at home

Program a derivative of a quadratic function. Proceed as follows.

- m) Create a vector 'x' with $x = (-10, -9.5, -9, \dots, 9, 9.5, 10)$. Then create a variable 'y' with $y = x^2$.
- n) Create a vector 'dx' with consecutive x differences, i.e. $dx_1 = x_2 - x_1$. Create a vector 'dy' similarly.
- o) Compute the derivative $\frac{\partial y}{\partial x}$ and plot it against x .
- p) Compare your results with the homework answers which will be available on Blackboard on Friday.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder