

R Programming for Economics & Statistics

Emmanuel Selorm Tsyawo*

Temple University, Department of Economics

January 23, 2019

R is the most comprehensive statistical analysis package available. It incorporates all of the standard statistical tests, models, and analyses, as well as providing a comprehensive language for managing and manipulating data. New technology and ideas often appear first in R.

- Source: <http://analyticstrainings.com/?p=101>

Why learn R ?

- (a) R is free and open source; reviewed by many renowned international statisticians and computational scientists.
- (b) A lot of documentation on-line; someone has already faced the problem you are facing and the solution is out there online.
- (c) Greater flexibility in programming; you can avoid constraints with syntax-based programming languages.
- (d) Numerous contributors of packages all over the world - an exponential growth in functionalities and accessibility
- (e) Fast and lean on memory; R requires little memory to run because you only load packages you need. This allows R to execute codes rapidly.
- (f) New books for R are constantly emerging. eg. Springer Use R! series, so you never lack adequate resources.

The workshop comprises three sessions, each per meeting.

1 Session I

1.1 Basics of R

The basics of R are required for creating and handling data objects and building an R programme.

Reference: Bloomfield, 2014, Chapters 2,3 & 4

*email: estsyawo@temple.edu

- (a) Creating, indexing, sub-setting and operating vectors
- (b) Creating, indexing, sub-setting and operating matrices
- (c) Creating, indexing, sub-setting and operating lists
- (d) Handling strings and lists
- (e) Loading data into R
- (f) Data manipulation in R
- (g) Plotting data in R
- (h) Logicals (True/False checks)
- (i) If/else statements
- (j) Loops - while loop, for loop,
- (k) Writing functions in R
- (l) Generating random numbers and handling distributions/densities

1.2 Econometric models using R packages

R has a number of packages that enable the estimation of many econometric and statistical models. In this section, we explore functions in available packages for econometric and statistical models. Recall, a key here is the interpretability of results and not mere estimation.

(Reference: Kleiber and Zeileis, 2008, Chapters 3, 5, & 7 and Wooldridge, 2010, Chapter 12)

- (a) Installing and using functions in R packages
- (b) Ordinary least squares
- (c) Logit/Probit
- (d) Regression Models for Count Data
- (e) Quantile regression
- (f) The Bootstrap

2 Session II

2.1 Numerical Methods and Optimisation

This section serves to illustrate some functionalities in R for solving mathematical problems numerically. The main focus is on solving non-linear problems where analytical solutions are infeasible.

(Reference: Bloomfield, 2014)

- (a) Solving a system of linear equations
- (b) Zeros of a function
- (c) Non-linear systems of equations
- (d) Numerical differentiation and integration
- (e) Unconstrained optimisation - one- & multi-dimensional optimisation
- (f) Constrained optimisation - one- & multi-dimensional optimisation with equality and non-negativity constraints

3 Session III

3.1 Programming your own model

In this last section, we illustrate, via examples, the programming of one's own model when there are unavailable functionalities. We also consider some machine learning models (if time permits).

(References: Own codes)

- (a) OLS with heteroskedasticity-robust standard errors
- (b) Writing your own maximum likelihood
- (c) Quantile regression via the asymmetric Laplace density likelihood
- (d) ℓ_1 penalisation of the objective function, - the Lasso

Useful references for further practice

- (a) Croissant, Millo, et al., 2018 - a very good text on panel data models in R
- (b) James, Witten, Hastie, and Tibshirani, 2013 - This text is a good introduction to statistical/machine learning
- (c) Albert, 2009 - Are you a believer in the bayesian paradigm? Get on board!
- (d) Shumway and Stoffer, 2006 - This book strikes a beautiful balance between the theory of Time series analysis and programming time series models in R.

Downloads

The following software applications are required for the workshop. Download and install them for the workshop. Also, check for the compatibility with the operating system.

- <https://cran.r-project.org/bin/windows/base/> - Download and install R
- <https://www.rstudio.com/products/rstudio/download/> - Download and install Rstudio

References

- [1] Jim Albert. *Bayesian computation with R*. Springer Science & Business Media, 2009.
- [2] Victor A Bloomfield. *Using R for numerical analysis in science and engineering*. CRC Press, 2014.
- [3] Yves Croissant, Giovanni Millo, et al. *Panel Data Econometrics with R*. Wiley Online Library, 2018.
- [4] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. Springer, 2013.
- [5] Christian Kleiber and Achim Zeileis. *Applied econometrics with R*. Springer Science & Business Media, 2008.
- [6] Robert H Shumway and David S Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2006.
- [7] Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.