

# 8<sup>th</sup> GESIS Summer School in Survey Methodology Cologne, August 2019

## Syllabus for Course 04: "Mathematical Tools for Social Scientists: A Refresher Course with R"

Instructors:	Michael Greenacre	Oleg Nenadić
E-mail:	michael.greenacre@upf.edu	onenadi@uni-goettingen.de
Homepage:	<a href="http://www.upf.edu">www.upf.edu</a>	<a href="http://www.uni-goettingen.de">www.uni-goettingen.de</a>

Date: 05.–09. August 2019  
Time: 09:00–13:00, 14:00–16:00  
Course starts Monday morning at 09:00

### About the Instructors:

*Michael Greenacre* is professor of Statistics at the University Pompeu Fabra, Barcelona. He specialized in multivariate data analysis, principally in the social and environmental sciences. Apart from over 100 published articles in international journals, book chapters and encyclopedias (see <http://www.researcherid.com/rid/D-7401-2013>), he has written six books on correspondence analysis and related methods and co-edited four books (with Jörg Blasius) on data visualization. His latest book is *Compositional Data Analysis in Practice* (Chapman & Hall / CRC, 2018). He has given courses in 15 countries in Europe, North and South America, Africa and Australia.

*Oleg Nenadić* is presently a Postdoc at the University of Göttingen and has been a visiting professor at the universities of Göttingen and Erfurt. His interest lies in Applied Statistics across various disciplines such as Economics, Social Sciences and Biology, and his main focus is on Computational Statistics with the statistical software environment R (<http://www.r-project.org>). He is author of a number of R-packages and has given workshops on R on four continents.

### Selected Publications:

- Korneliusson, T. and Greenacre, M. (2018). Information sources used by European tourists: a cross-cultural study. *Journal of Travel Research* 57: 193–205.
- Greenacre, M. (2017). "Size" and "shape" in the measurement of multivariate proximity. *Methods in Ecology and Evolution* 8: 1415–1424. Video on YouTube at <https://www.youtube.com/watch?v=d5erjODvHvo>
- Greenacre, M. (2016). *Correspondence Analysis in Practice* (Third Edition). Chapman & Hall / CRC, Boca Raton, USA.
- Blasius, J., Nenadić, O., Thiessen, V. (2018): The Dirty Data Index – Assessing the Quality of Survey Data in International Comparison, *Italian Journal of Statistics*, 29 (2-3), 137–152.
- Zucchini, W., Schlegel, A., Nenadić, O., Sperlich, S. (2009): *Statistik für Bachelor- und Masterstudenten – Eine Einführung für Wirtschafts- und Sozialwissenschaftler*. Springer, Berlin, Heidelberg.
- Greenacre, M. and Nenadić, O. (forthcoming): *Mathematical Tools for Social Scientists – An Introduction with R*. BBVA Foundation, Bilbao.

### Short Course Description:

This course aims to refresh mathematical concepts which are required for the understanding and the application of recent developments in empirical research methodology. The course covers the fundamentals of Mathematics (functions, linear algebra, calculus and optimization), focusing on the understanding of the concepts. Instead of pursuing a formal approach, this course will help participants to familiarize themselves with what we consider as essential and useful mathematical knowledge. Thus, one aim of the course is to remove the commonly experienced uncertainty when researchers are dealing with mathematical concepts and expositions in their research.

An important part of this course is the active use of the open-source statistical programming language R. Since its introduction in the 1990s, R has become a *de facto* standard for statistical computing. As such, this workshop combines communicating mathematical concepts with an introduction and the active application of R, both of which are taught in parallel.

The course sessions basically comprise of two continuously alternating parts: the "theoretical part" covers the mathematical explanations while the "applied part" re-elaborates and reinforces the theoretical part by actively using R. In this way, the course lays the foundations for advanced empirical research methodology by covering the theoretical background as well as providing the participants with an insight and competence in R.

## Keywords:

Mathematics, empirical social research, R

## Course Prerequisites:

- The only prerequisite for the workshop is the motivation to learn (or to repeat) the fundamentals of Mathematics and R;
- Note: Participants may bring a laptop computer in order to perform the practical exercises in this course. In the PC lab sessions however, you will have access to preinstalled hard- and software. Familiarity with R is not a prerequisite. In case you wish to work on your own computer, downloading and installing R (<http://cran.r-project.org>) and Rstudio (<http://rstudio.org>) prior to the course is strongly suggested. Participants are encouraged to play around with the software before the course in order to get a feeling for it.

## Target Group:

The course is conceptualized to support social scientists who want to improve their mathematical background knowledge as a prerequisite for advanced empirical research. It is commonly found that mathematical prerequisites are an obstacle for understanding advanced statistical methods. In many cases, this may be due to the fact that the last Mathematics course has been taken at least some time ago. Thus, the course is aimed at students and researchers, especially in the social sciences, who wish to renew their mathematical competence as well as learn or fine-tune their programming skills using R.

## Course and Learning Objectives:

By the end of the course, participants will:

- have an understanding of the fundamentals of Mathematics, as well as several advanced topics;
- have learnt how several basic as well as advanced statistical methods are conceptualized mathematically and how they are implemented analytically and computationally;
- give R beginners an introduction to the most powerful and universal computing tool available;
- give existing R users new insights into the R environment.

## Organizational Structure of the Course:

Summer school courses are typically structured around four hours of classroom instruction and two hours of hands-on exercises, tutorials, or lab sessions per day. Lecturers shall be available for group and/or individual support during these two hours that – depending on the scheduled rooms – may either be organized as one longer session or multiple shorter sessions spread over the day.

This course is innovative in that the "theoretical" and the "practical" parts will alternate continuously and not be totally separated within each day. Typically, after a theoretical chunk (approximately 20-30 Minutes), the discussed topics will be elaborated in R in the practical chunk (approximately 20-30 Minutes). This makes the learning process dynamic, as opposed to delivering all the theory in the morning, for example, and doing the practical part separately in the afternoon. In addition, exercises for each topic will be provided, a few of which participants will have to solve during the class. Most exercises will be done after class, and participants will be strongly encouraged to solve as many of these as they can on their own after each daily session. Problems arising from the exercises will be discussed at the start of the following day, which also provides a revision of the previous day's material before starting with new topics.

## Software and Hardware Requirements:

Required hardware:

- Participants will need to bring their own laptops.
- A printed copy of our book "Mathematical Tools for Social Scientists - An Introduction with R" will be distributed to every participant (approximately 160 pages)

Required software:

- R (<https://cran.r-project.org>) and RStudio (<https://www.rstudio.org>). No additional R packages are required.

## Long Course Description:

- |   |   |
|---|---|
| <p>1 Introduction to Mathematics &amp; R</p> <ul style="list-style-type: none"> <li>1.1 Numbers</li> <li>1.2 Powers and common transformations</li> <li>1.3 Equations and inequalities</li> <li>1.4 Some mathematical notation</li> <li>1.5 Sets, categories and logical variables</li> <li>1.A Exercises</li> </ul>  | <p>2 Functions</p> <ul style="list-style-type: none"> <li>2.1 Examples of functions</li> <li>2.2 Inverse functions</li> <li>2.3 Functions of two or more variables</li> <li>2.A Exercises</li> </ul>  |
| <p>3 Linear algebra</p> <ul style="list-style-type: none"> <li>3.1 Many into one</li> <li>3.2 Linear transformations</li> <li>3.3 Operations on vectors and matrices</li> <li>3.4 Linear equations and matrix inverses</li> <li>3.5 More matrix-vector operations</li> <li>3.6 Some special matrix transformations</li> <li>3.A Exercises</li> </ul>  | <p>4 Calculus</p> <ul style="list-style-type: none"> <li>4.1 A simple example</li> <li>4.2 Limits</li> <li>4.3 Basic forms and rules of derivatives</li> <li>4.4 Second derivatives</li> <li>4.5 Partial derivatives and derivatives of vectors</li> <li>4.6 Integration</li> <li>4.7 Difference and differential equations</li> <li>4.A Exercises</li> </ul>   |
| <p>5 Optimization</p> <ul style="list-style-type: none"> <li>5.1 Finding the best</li> <li>5.2 Optimizing functions with "closed form" solutions</li> <li>5.3 Optimizing single-variable functions by iterative methods</li> <li>5.4 Optimizing multivariable functions: direct search method</li> <li>5.5 Optimizing multivariable functions: gradient methods</li> <li>5.A Exercises</li> </ul> | <p>6 Statistical applications</p> <ul style="list-style-type: none"> <li>6.1 Simple linear regression: least-squares fitting</li> <li>6.2 Multiple linear regression</li> <li>6.3 Maximum likelihood estimation of simple regression model</li> <li>6.4 Maximum likelihood estimation of Poisson regression model</li> <li>6.5 Principal component analysis and factor analysis</li> <li>6.6 Mathematics of networks</li> </ul> |

## Day-to-day Schedule and Literature:

Day	Topic(s)
1	1.) Introduction to Mathematics & R 2.) Functions
	<u>Suggested reading:</u> Chapters 1 and 2, "Mathematical Tools for Social Scientists - An Introduction with R"
2	3.) Linear Algebra
	<u>Suggested reading:</u> Chapter 3, "Mathematical Tools for Social Scientists - An Introduction with R"
3	4.) Calculus
	<u>Suggested reading:</u> Chapter 4, "Mathematical Tools for Social Scientists - An Introduction with R"
4	5.) Optimization
	<u>Suggested reading:</u> Chapter 5, "Mathematical Tools for Social Scientists - An Introduction with R"
5	6.) Statistical Applications
	<u>Suggested reading:</u> Chapter 6, "Mathematical Tools for Social Scientists - An Introduction with R"

## Preparatory Reading:

The workshop follows the book "Mathematical Tools for Social Scientists - An Introduction with R" by M. Greenacre and O. Nenadić (forthcoming). The participants will be provided with a copy of the book.

## Additional Recommended Literature:

- Nenadić, O. & Greenacre, M. (2007) Correspondence analysis in R, with two- and three-dimensional graphics: the *ca* package. *Journal of Statistical Software* 20(3). URL: <http://www.jstatsoft.org/v20/i03/>
- Greenacre, M. (2010). *Biplots in Practice*. BBVA Foundation, Bilbao. Downloadable for free from [www.multivariatestatistics.org](http://www.multivariatestatistics.org).
- Greenacre, M. (2016). *Correspondence Analysis in Practice (Third Edition)*. Chapman & Hall / CRC, Boca Raton, USA.