Lab notes for *Statistics for Social Sciences II: Multivariate Techniques*

BSc in International Studies, BSc in International Studies & Political Science, Carlos III University of Madrid

Eduardo García Portugués

2016-09-07

Contents

Introduction

Welcome to the lab notes for *Statistics for Social Sciences II: Multivariate Techniques.* Along these notes we will see how to effectively implement the statistical methods presented in the lectures. The exposition we will follow is based on learning by **analyzing datasets and real-case studies**, always with the help of **statistical software**. While doing so, we will illustrate key insights of multivariate techniques and the adequate use of advanced statistical software.

Be advised that these notes are *neither an exhaustive, rigorous nor comprehensive treatment* of the broad statistical branch know as *Multivariate Analysis.* They are just a helpful resource to implement the very specific topics covered in this short course.

Some course logistics

- **Lessons**. **Mondays** at *López Aranguren* building:
  - IS. **16:15-17:45 at INF-15.S.06**.
  - IS & PS. **18:00-19:30 at INF-15.S.04**.
- **Office hours**. **Tuesdays 16:00-17:00 at office 10.0.10** (access through 10.0.7 in *Campomanes* building). If they are incompatible with your schedule, send me an email to get an alternative appointment (preferable) or drop by my office to see if I am available (but I will remove this option if I get overwhelmed with queries).
- **Grading**.
  - Continuous evaluation is 60% of the final grade. Scored by 2 partial exams and 1 group project, each accounting for a 20%.
  - Final exam is 40%, covers mostly contents from the lectures.
- **Partials**. Cover contents from labs and lectures. The first will (likely) cover Topics 1-2. The second, Topics 3-5.
- **Group project**. Students must team up in groups of 4 (±1) to apply the statistical methodology taught to a dataset of their choice and produce a report. As a rule of thumb, all students in a group will be graded evenly, so choose wisely your coworkers. Specific details will be disclosed near to the end of the course.

Software required

The software required is available in all UC3M computer labs, including INF-15.S.04 and INF-15.S.06 (the first weeks only in these labs). We will employ two pieces of software:

- `R`. A free open-source software environment for statistical computing and graphics. Virtually all statistical analyisis you can think of is availabe in `R`.

- `R Commander`. A Graphical User Interface (GUI) designed to make `R` accessible to non-specialists through friendly menus. Essentially, it acts as a translator from 'human' to `R` code.

The only thing you need to do to run `R Commander` in any UC3M computer is:

1. Run 'Start' -> 'R3.3.1' -> 'R3.3.1 (consola)'. A black console will open. Do not panic!

2. Type inside

   `library(Rcmdr)`

If `Rcmdr` is installed, then `R Commander` will open automatically and you are ready to go. If it is not, then type

`install.packages("Rcmdr", dep = TRUE)`

and say 'Yes' to the next pop-ups. Choose a CRAN mirror, e.g. 'Spain (*) [https]', which usually work well. Wait for the downloading and installation of the packages. When it is done, type `library(Rcmdr)`.

An **important** warning about UC3M computer labs:

> Every file you save locally in the computer (including downloaded packages) will be wiped out after you close your session! So be sure to save your worth files in a pendrive or in a your `Google Drive` account.

Alternatively, you can bring your own laptop, see below.

Why this software?

There are many advanced commercial statistical packages, such as `SPSS`, `Excel` (with commercial add-ons), `Minitab`, `Stata`, `SAS`, etc. We will rely on the combo `R + R Commander` due to the its markable advantages:

1. **Free and open-source**. (Free as in beer, free as in speech.) No software licenses are needed. This means that you can readily use it outside UC3M computer labs.

2. **Scalable complexity and extensibility**. `R Commander` creates `R` code that you can see, and eventually understand. Once you begin to get a feeling of the code, you will realize it is faster to type the right commands than to navigate through menus. `R Commander` has a good number of high-quality and user-friendly plugins.

3. **`R` is the leading computer language in data science/statistics**. Any statistical analyisis you can imagine is already available in `R` through its almost 9000 free packages (September, 2016). Some of them contain a good number of ready-to-use datasets. Albeit the methodological contents covered in this course are standard, bear in mind that you will be trained on how to implement them in a top-valued statistical tool.

4. **`R Commander` produces high-quality graphs easily**. `R Commander`, through the plugin `KMggplot2`, uses the `ggplot2` library. `ggplot2` delivers publication-level graphs and is considered as one of the best and more elegant graphing packages nowadays.

5. **Great report integration**. `R Commander` integrates with `R Markdown`, which is a package able to create dynamic `.html`, `.pdf` and `.docx` reports directly from the outputs of `R Commander`.

In summary, with `R + R Commander` you will be trained in a highly-valued software that allows you to produce statistical reports efficiently. Mastering this surely will make an **important distinction in your graduate profile** (social scientists *used* to lack this kind of formation). So I encourage you to take full advantage of this opportuity!

Installation in your own laptop

You are allowed to bring your own laptop to the labs. This may have a series of benefits, such as admin privileges and the option to save all your files locally. But keep in mind:

> If you plan to use your personal laptop, **you** are responsible for the right setup of the software and laptop prior to the lab lesson.

At some point you will probably need to run the software outside UC3M computer labs. This is what you have to do to install everything:

1. Download `R` (Windows, Mac).

2. Install `R`. In Windows, be sure to select 'Startup options' and then choose the 'SDI' mode in the 'Display Mode'. Leave the rest of the options as default.

3. Open `R` and copy and paste:

   ```
   install.packages(c("Rcmdr",  "RcmdrPlugin.TeachingDemos",  "RcmdrPlugin.FactoMineR",
   "RcmdrPlugin.KMggplot2"), dep = TRUE)
   ```

   Say 'Yes' to pop-ups and choose the CRAN mirror.

4. To launch the software, run `R` ('R x64 3.3.1' in Windows, 'R' in Mac) and then

   ```
   library(Rcmdr)
   ```

If there is any Linux user, follow this and this.

Bibliography