

CSSS/POLS 510 MLE Lab

1. Logistics and R Refresher

Ramses Llobet

October 8, 2023

Disclaimer

The current version of the lab materials is adapted from those drafted by previous TAs for this course.

About me

- ▶ Welcome to the first lab section of POLS / CS&SS 510!
- ▶ I am Ramses, a fourth year Ph.D. student in **Political Science**.
 - ▶ Research interests: political economy and quantitative methods.
- ▶ I am from Barcelona, Spain.
 - ▶ Please **do not** hesitate to stop me if you don't hear or understand me properly.
- ▶ **Do not** hesitate to ask questions. No question is silly. :)

Logistics

1. **Lab Sessions:** Fridays, 3:30 - 5:20pm via Zoom ([link](#))

- ▶ Reviews lecture materials using examples; Q & A for assignments and lectures
- ▶ Materials will be available on the [course website](#)
- ▶ Always look for these files: `lab1_slide.pdf`, `lab1_practice.rmd`, `lab1_key.rmd`, and `lab1_data.csv`
 - ▶ I will also provide a compressed *ZIP* file with all materials.

2. **Office Hours:** after labs or by appointment: rlobet@uw.edu.

- ▶ Trouble-shooting, questions about the lecture and assignments, etc.
- ▶ Please email me with time and a short comment of the topic you want to discuss.
- ▶ Zoom link: ([meeting room](#))

Logistics (Cont.)

3. Homework: 5 due every 2 weeks or so

- ▶ Must be typed up (not hand-written).
- ▶ Using \LaTeX in Overleaf or R Studio with R Markdown is an easy way to do this (will work on these next week).
- ▶ We will use two of [Chris's packages](#) extensively: *simcf* and *tile*.

4. Please make sure you name the assignment files properly as follow:

- ▶ Assignments to me (canvas):
CSSS510HW1_NameSurname.pdf
 - ▶ E.g. CSSS510HW1_RamsesLlobet.pdf
- ▶ Assignments to Chris: Make sure to include all your co-authors' names in the file name.

Logistics - Goals

1. When this course is over, you should be able to do the following (and much more):
 - ▶ Identify the proper **distribution** and model for your data (logistic, ordered, multinomial, count).
 - ▶ Run the model using both the *glm* function and programming “by hand” using *optim*, extract parameters of interest, and interpret these in probabilities.
 - ▶ Compute **predictions** and use simulation to find the confidence intervals of $\hat{\pi}_i$ across counterfactuals values of covariates x_i .
 - ▶ Use **cross-validation** to assess the predictive accuracy of several models and also compare these models across a variety of in-sample **goodness of fit** tests.
 - ▶ **Visualize** computations and quantities of interests.

R setup

- ▶ How to install R and R-studio.
 - ▶ R-4.2.3 for [Windows](#)
 - ▶ R-4.2.3 for [macOS](#)
- ▶ R-studio can be downloaded from [posit's repository](#).
- ▶ To render **PDF** files in markdown, you will need to install the package `tinytex`.
 - ▶ At the end of today's lab review script file, you will find the functions to install it.

Logistics - R

1. **The stuff in R:** For the homework assignments and project you will need to feel comfortable
 - ▶ importing (and exporting) data sets.
 - ▶ tidying and transforming data.
 - ▶ analyzing data (conceptual part of the course).
 - ▶ generating plots of your data and results.
 - ▶ writing basic functions and loops for repeated procedures.

Logistics - R

2. I have to read lots of your code. Please be considerate when writing code and submitting assignments.
 - ▶ Do not print unnecessary code and output. Learn how to use `results = "hide"` and `echo = TRUE` in R Markdown.
 - ▶ Name well
 - ▶ functions vs. all other objects
 - ▶ readability is about consistency (`dot.naming`, `CamelCaseNaming`, `pothole_naming`).
 - ▶ short, clear, consistent – help future you (and present me)
 - ▶ Be tidy in your code and your workspace/directory.

Logistics - R

2. I have to read lots of your code. Please be considerate when writing code and submitting assignments.

▶ Specify arguments fully, e.g.

```
rbinom(n = 1000, size = 30, prob = 0.49) # GOOD!
```

```
rbinom(1000, 30, 0.49) # LESS GOOD!
```

▶ See the [Google R styleguide](#) for an example.

Logistics - R Useful resources

- ▶ For R:
 - ▶ Introductory:
 - ▶ *Hands-On Programming with R* (Grolemund 2014).
 - ▶ R [cheat sheets](#).
 - ▶ Intermediate:
 - ▶ *R for Data Science* (Grolemund and Wickham 2023, 2nd edition).
 - ▶ *Data Visualization: A Practical Introduction* (Healy 2018).
 - ▶ *Graphical Data Analysis with R* (Unwin 2015).
 - ▶ Advanced:
 - ▶ *Advanced R* (Wickham 2019).
- ▶ For MLE:
 - ▶ *Maximum Likelihood for Social Science* (Ward and Ahlquist, 2018).
 - ▶ [Book Materials](#) code, data, etc.
 - ▶ *Regression and Other Stories* (Gelman et al. 2022).

Logistics - Social Sciences & Computing

1. There are best practices for computing in the social sciences. You should aim for transparency and replicability in your work in general, and clarity and consistency in your code.
 - ▶ [Best Practices \(Wilson et al. 2014\)](#)
 - ▶ [Good Enough \(Wilson et al. 2017\)](#)

R refresher

1. Overview

- ▶ R is a language and environment for statistical computing and graphics
 - ▶ *Object-oriented* style of programming
 - ▶ System-supplied or user-defined functionality as *functions*
 - ▶ Extended via *packages*
- ▶ RStudio is an integrated development environment for R, which includes:
 - ▶ a console to run R code
 - ▶ an editor to write code and text
 - ▶ tools for plotting, history, debugging and workspace management

R refresher

2. Data Types

- ▶ character, numeric (integer or double), logical, complex
- ▶ data can also be missing

3. Data Structures

▶ Matrices vs. data frames

- ▶ Matrices can only contain one **homogenous** type of vectors
- ▶ Data frames can contain **heterogeneous** types of vectors, and thus are more flexible

R review

- ▶ R basics.
- ▶ Data wrangling with `dplyr`.
- ▶ Analysis.
- ▶ Install ‘tinytext’ for RMarkdown.

Let's open RStudio and **review_scrip.R**.