Introduction to R: Base R

Cristian Huse

University of Oldenburg

Overview

- ▶ This session provides an introduction to the statistical software **R**
  - ▶ R is a "free software" which is one of the most popular choices among statisticians, data scientists and – increasingly – economists;
  - ▶ R's use became more user-friendly with the development of **IDE**'s (Integrated Development Environments). One such IDE is **RStudio**, which is the one we are adopting;

- ▶ An IDE provides a better overview of the software, the objects in memory, besides providing facilities when one is writing code or even running a simple session – just save the session as an **.R** file (also readable using the Notepad);

- ▶ RStudio enables users to combine code and writing (e.g., reporting results) in the same document via **R Markdown**;
  - ▶ The source of this document (**.Rmd**), for instance, is an R Markdown file;

Overview (cont'd)

- ▶ While economists are often interested in estimating regressions, before doing so we need to go through some basic issues which are important to know regarding R in a pragmatic way, e.g.,
  - ▶ Express overview of R, RStudio, RMarkdown
  - ▶ What are packages/libraries
  - ▶ Basic commands
  - ▶ Types of data

- ▶ In addition to estimate your regressions you will need to do other things
  - ▶ How to load the data, transform them, merge them with other data
  - ▶ Present your data using plots and tables with summary statistics
  - ▶ Present estimation output in tables
  - ▶ ...

- ▶ We will briefly cover some of the topics above

Overview (cont'd)

- Having said that, one double-edged sword when it comes to R is the existence of (thousands of) packages
  - Often written by specialists and well-designed;
  - However, one needs to search for or have heard of a package;
  - Since they are typically not part of "base R", they have to be installed then loaded called when writing code so its functions can be used later on.

A Word on R Markdown

- ▶ Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see here.
- ▶ When you click the **Knit** button, a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.
- ▶ You can embed an R code chunk by starting it with "'{r} and ending it with "'

```
summary(cars)
```

```
##     speed           dist
## Min.   : 4.0   Min.   :  2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

A Word on R Markdown (cont'd)

▶ You can also embed plots (code doesn't appear because I purposefully wrote 'echo = FALSE'):
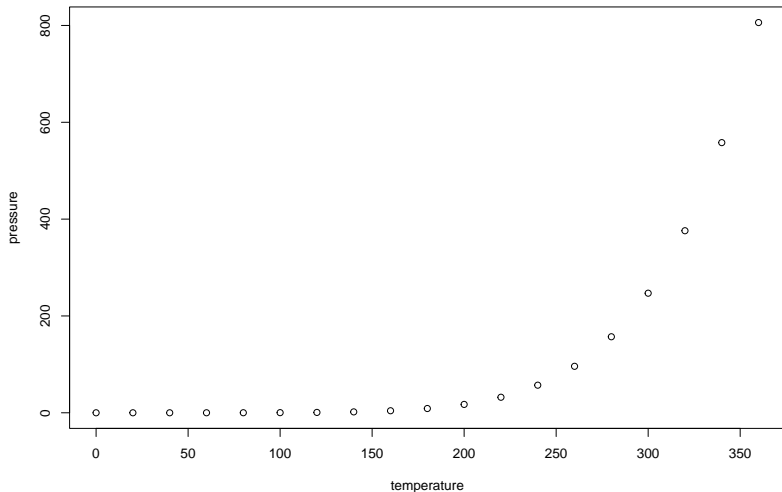
Setup: R and RStudio

- First, "base R" already provides all tools needed for basic regression analysis, plotting etc, but not necessarily in the most convenient form:
  - Install R from here (see tutorial below)
  - Then install the RStudio IDE from here;
  - You might want to install RTools (will be useful in the future) from here

- Think of RStudio as a skin, user-friendly interface, to R, which is the real engine under the hood

- Some helpful tutorials (Google and youtube are your friends):
  - Installing R and RStudio
  - Getting started with R and RStudio
  - Basic R commands
  - Reading in, Accessing, and Summarizing Data in R

Setup: Packages

- Packages provide specialized functions within R. For an incomplete list of packages, see here (some packages are available elsewhere, e.g., Github)

- The standard way to install and then call a package is

```
install.packages("ivreg")
library(ivreg)
```

- Some useful packages include:
  - **tidyverse**, **hrbrthemes**, **listviewer**;

- Other important packages include:
  - **fixest**, **estimatr**, **ivreg**, **sandwich**, **lmtest**, **mfx**, **margins**, **broom**, **modelsummary**, **vtable**, **mlogit**, **data.table**

- There are several (very helpful) cheatsheets available here

Lab Session: Base R

Structure of the Session

- ▶ Basic commands
  - ▶ Assignment operator, lists, element-wise operations, length vs. dimension

- ▶ Matrix definition and operations

- ▶ Simulating data and "regression zero"

- ▶ Basic plots

- ▶ Indexing data

- ▶ Loading libraries and their associated data
  - ▶ Transforming data: factor variables
  - ▶ Basic exploratory analysis
  - ▶ Summary statistics
  - ▶ Descriptive regressions: regressions 1 and 2

Instructions for the Lab Session Base R

▶ Please load the file **IntroR_BaseR.R**

▶ Go through the file line-by-line, consulting the help whenever needed (book 1-2 hours)

Details

- An Introduction to R

    - From the link you will see that is the "official" Introduction to R, so very complete and definitely worth going through
    - Please read Chapters 1-6
    - You might want to briefly check:
        - Appendix A ("A sample session", until "Can you see it?")
        - Chapter 7 ("Reading data from files")
        - Chapter 12.1-2 ("Graphical procedures")

Problem Set

- ▶ Please solve the problem set associated to this session
  (**PS_BaseR.pdf**) before the next lecture (book 1-2 hours)

Take-aways

- ▶ The aim of this slide deck is to provide you an overview of R and RStudio and guide you through a simple R session;

- ▶ Is is fundamental that you go through the lab session try to understand logic underlying the commands – this is pretty much like learning a **new language**;

- ▶ From my own experience, the best way to learn is to get your hands dirty with data:
  - ▶ Go through the files in detail
  - ▶ Take something you know and have done before and re-do the project using a new language
  - ▶ There are countless channels, tutorials, books, and communities, e.g., Stack Overflow

- ▶ As in everything, the contents and the approach pursued here are biased, incomplete, and reflect (my) personal taste

Selected References

- An Introduction to R

- R Data Import/Export

- Wickham & Grolemund's R for Data Science

Appendix
Setup: Package Management

- Installing packages one-by-one can become tedious, but one can use **pacman**, a package management tool to automatize the process:

```r
## Load and install several packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(mfx, tidyverse, hrbrthemes, estimatr,
               ivreg, fixest, sandwich, lmtest,
               margins, vtable, broom, modelsummary,
               data.table,fastverse)
## Make sure to have at least version 0.6.0 of ivreg
if (numeric_version(packageVersion("ivreg")) <
    numeric_version("0.6.0"))
  install.packages("ivreg")
## Optional -- ggplot2 plotting theme
theme_set(hrbrthemes::theme_ipsum())
```