Experimental and Statistics Software

SUMMER TERM 2021

Prof. Dr. Thomas-Patrick Schmidt
Faculty of Society an Economics

COPYRIGHT

This presentation is only for internal use at Rhine-Waal University of Applied Sciences in the summer term 2021.

It is not permitted to record, pass on or publish the video and audio of this lecture without the written permission of the lecturer.

EXPERIMENTAL AND STATISTICS SOFTWARE

Lecture 2

Organization Update

Introduction to R | Basics in R

COURSE OUTLINE

- Part I: Introduction to Data Analysis with R
  - **Introduction to R & Basics in R [CW 16, 17]**
  - Descriptive Statistic & Normal Distribution [CW 18, 19]
  - Hypothesis Testing [CW 21, 22]
  - Linear Regression [CW 23, 24]
- Part II: Applied Econometrics with R (we have to pick sth.) [CW 25, 26, 27]
  - Regression with Panel Data **or**
  - Regression with Binary Dependent Variable **or**
  - Instrumental Variable Regression **or**
  - Experiments and Quasi-Experiments

ORGANIZATION UPDATE LECTURE & CASE STUDY

- Live-Sessions only on demand (!!)
- **"lectures": me & you (online, video shared via moodle)**
  - **"bigger picture"**
  - **(some) basic statistics but no focus**
  - **hands-on programming in RStudio**
  - **presentation of the case study**
- Case Study: you & me
  - Questions / Work on the questions yourself
  - you are in the driver's seat, I'll support you
  - however, since we will use the result of the case study for your examination, you'll have to perform it individually (no group work, no solutions given by me)

ORGANIZATION UPDATE EXAM

Topic	Lecture	Case Study (+1)	Submission deadline (+1)	Points
Introduction to R & Basics in R	CW 16	CW 17	CW 18 [2021 – 05 – 09 23:59 CET]	10
Descriptive Statistic	CW 18	CW 19	CW 20 [2021 – 05 – 23 23:59 CET]	20
Hypothesis Testing	CW 21	CW 22	CW 23 [2021 – 06 – 13 23:59 CET]	20
Linear Regression	CW 23	CW 24	CW 25 [2021 – 06 – 27 23:59 CET]	20
Applied Econometrics	CW 25 & 26	CW 27	CW 28 [2021 – 07 – 18 23:59 CET]	30

- Open Source → free (important at least since we cannot use the computer lab efficiently because of the covid pandemic)

- R is a platform, analysis based on application → packages
  - broad community: put your (econometric) "problem" +"R" into google/youtube

- But: R appears not that user friendly

- RStudio
  - standalone program based on R (free)
  - modern user interface

- R Commander → Graphical Interface for R (we will not use this)

- Download & Installation: https://www.r-project.org/
- Some functions:

Function	Description	Example
+	Addition of values	5 + 3 = 8
−	Subtraction of values	2 − 4 = −2
*	Multiplication of values	8 * (−2) = −16
/	Division of values	−16/16 = −1
sqrt()	Square root of a number	sqrt(9) = 3
$(y)^{(1/x)}$	x-th root of the number y	$7776^{(1/5)} = 6$
^	Power of a number	$3^3 = 27$
log()	Natural logarithm	log(120) = 4.79
exp()	Exponential function	exp(10) = 22,026.47

INTRODUCTION TO R
INSTALL RSTUDIO

- Download & Installation: https://www.rstudio.com/

- Script window: document our analysis & store it

- R Console:

  - results from executing script

  - enter commands and look if they do what we wand

- Data and object window

  - active datasets

  - other objects

- Environment flow

  - Files, plots, packages, help

INTRODUCTION TO R
FIRST ANALYSIS WITH R STUDIO

```
#First Analysis

#Author: Thomas Schmidt
#Date: 2021-04-25

#lecture 1 part 1
#Introduction to R

setwd("d:/r/")

library(readxl)

hound <- read_excel("D:/r/hound.xlsx")

str(hound)

head(hound)

summary(hound)

hound$weight

mean(hound$weight)

hist(hound$weight)

#End
```

INTRODUCTION TO R
HOW TO INSTALL PACKAGES

- R packages for statistical analysis
- installation
  - command: *install.packages(), example: install.packages("ggplot2")*
  - menu-driven: easy with Rstudio
- activation
  - command: library(ggplot2)
  - part of our code
- How to know which package we need?
  - help in R
  - community

BASICS IN R
IMPORT DATA

- menu-driven

  – data and object window → environment

  – import dataset

- command-driven

  – part of our code ("end-to-end" R Script)

  – example
    library(readxl)
    hound <- read_excel("D:/r/hound.xlsx")

- head(): display to six rows of the dataset
  - head(hound)
- tail(): display last rows of the record
  - tail(hound)
- View(): display the whole record in the window
  - View(hound)
- str(): general information about the dataset, variables, data types of the characteristics, characteristic values
  - str(hound)

BASICS IN R
CREATING AND MODIFYING VARIABLES

- nominal and ordinal variables should be recorded as factors (=data type in R for both nominal and ordinal)

  - example: sex should be recorded into a factor variable and the new variable should be added to the dataset

    f_sex <- as.factor(hound$sex)

- add a variable to the dataset

  hound<- cbind(hound,f_sex)

- we could also use arithmetic operations

  - One dogyear equals six human years

    humanage<- dogs$age*6
    hound<- cbind(hound,humanage)

BASICS IN R
SELECTING CASES

- to pick rows and columns, we can command dataset[rows, colums]

- example

  - first column and row: hound[1,1]

  - several rows/columns via ":", hound[1:4, 1:4]

  - all rows: leave rows empty: hound[,1:3]

  - names instead of numbers: hound[,"sex"]

  - several rows/columns which are not one after another: hound[,c("dog","sex","age")]

- can be used to generate a reduced dataset:

  hound_reduced <- hound[,1:3]

- list of repeatedly used commands:

Command	Description
ls()	Lists all active variables and objects within the working memory
rm()	Deletes selected variables and objects within the working memory
rm(list=ls())	Deletes all variables and objects within the working memory
str()	Displays important information about the structure of the dataset
View()	Displays the dataset in a new window
help()	Gives help for commands

BASICS IN R
SCRIPT FILES & REPORTING

- Script file

  - First Section: rm(list=ls())   setwd("d:/r/")

  - Second Section: library(readxl) read_excel("D:/r/hound.xlsx")

  - Third Part: Analysis

  - Use comments

  - Save the file

- Reports

  - PDF, HTML..

Calculate the following terms:

- (5 + 2)*(10 − 3)

- square root of 64

- third root of 64

- third root of 512

- 4 to the third power

- 8 to the third power

- logarithm of 500

- logarithm of 10

CASE-STUDY EXERCISE 2

- Download the dataset hound.xlsx from moodle and save the dataset in a folder on your computer

- Try to use a short path and a folder with a simple name. Remember the path and the folder; it is the way and the place to find the data.

- Read the dataset hound.xlsx via Import Dataset button. Delete the dataset with the command rm(). Read the dataset again with the import command in the script window. Delete your working memory completely with the correct command.

CASE-STUDY EXERCISE 3

Read the dataset hound.xlsx and try to do the following tasks:

- Have a look at the dataset with the relevant command. Check whether the data set is complete by looking at the beginning and at the end of the dataset with the relevant commands.

- Have a look at the general information about the data set, i.e. the data type of variables, characteristic values, and so on.

CASE-STUDY
EXERCISE 4: REDUCED DATASET

- Produce a reduced dataset with only 6 observations and without the variable breed.

- Name the dataset hound_red

- How old is the dog in the fourth observation? Try to figure it out by displaying only this value. Check the value received with the help of the command View().

- Generate with humanage one new variable that is added to the dataset hound.xlsx. You heard that one dogyear is only 5 humanyears not 6.

- Convert the variables sex, size and breed into factors.

CASE-STUDY DEADLINE

- Deadline: Please upload the result of work latest 2021 – 05 – 09 23:59 CET

- upload to moodle, I'll provide upload function

- result of work (basically one for all, of course you can also generate one for each exercise)
  - R Script
  - Report (html)