Computergestützte Statistik
WS 2021/2022

Lena Janys
Assistant Professor Institute for Finance and Statistics
ljanys@uni-bonn.de

October 11, 2021

No lecture in week two (no lecture 19th and 21st of October)!
Zoom Link for the lecture: https://uni-bonn.zoom.us/j/92546054860?pwd=Tmx2Rlp5aDN6R3ZiaWxRd0lEcGFSUT09
Zoom Link for the office hours: https://uni-bonn.zoom.us/j/99778704920?pwd=cVl0enZITUpWVzVydkJKaW9jQjRhQT09

Syllabus

In general I assume that you have taken Econometrics and Statistics A and Statistics B and have a working knowledge of the methods discussed there.

The goal of this course is to introduce

1. basic methods from Statistics/Econometrics
2. how to implement these methods in the language R
3. how to simulate from relevant data generating processes
4. how to interpret empirical results
5. also introduce some methods from machine learning/

This is a very applied course, where we want to emphasize how to *implement* a particular method and how to *interpret* the results.

To this end, the course will be divided into three blocks

1. Introduction to R
   (a) How to write basic code
   (b) Functions and loops
   (c) Basics of simulation/DGP
2. Basic Concepts Econometrics
   (a) Linear Model (review and implementation)
   (b) Bias/Variance tradeoff
   (c) Nonparametric density estimation
   (d) Nonparametric regression
   (e) Maximum likelihood estimation
3. Machine learning/regularization

(a) resampling methods: bootstrap and cross-validation

(b) Regularization: ridge regression, lasso

(c) Joker: possibly random forests

I will interweave applications from social science, as well as common pitfalls when implementing these methods and particular problems that occur in social science applications.

Literature and languages:

1. Programming: we will use R for all implementations. Please download any distribution of R for your machine https://cran.r-project.org and you may also download any other editor than the build-in one, I prefer https://rstudio.com/products/rstudio/.
2. The main reference will be my own script, I will publish an up-to-date version every week with the relevant chapters. I will add some more reading material

Format:

- The lectures will be held live at the times specified in BASIS on Tuesday and Wednesday.
- All materials will be available online on the GitHub repo https://github.com/LJanys/Computergest_Statistik. I will also upload all other materials on this repo.
- Wednesdays are reserved for implementation and presentation of problem sets. Once we finish the R introduction/ revision of some econometric concepts, I would like for everyone to present their results of the problem set by sharing their screen and discussing the results in class, so please expect to do that at some point and to participate actively in class, even if you are not presenting. This means (1) be prepared (2) think of one or two questions you would like to ask the presenters and/or that you had while working on the problem set. Be prepared to be called to ask those questions at the end of the presentation.

Feedback: I wanted to provide an environment where you can leave quick, anonymous and helpful feedback for the running of the course: Survey. This is not a replacement for the typical course evaluations, but since I will not get them until the beginning of the next semester, any useful info about the content and format will be too late for you to benefit from.

Office Hours: I will have virtual office hours every week before the lecture on Tuesday. If you want to sign up for a slot, please do so until Monday evening under (https://signup.com/go/JWjJrCB)

Grade: Instead of an exam there will be a project/Hausarbeit, for which you will need to pick a topic in a couple of weeks, once you have decided that you want to stay in the course. You need to register for the course with the Prüfungsamt until **TBA**. Once everyone has registered for the course, you will be assigned a problem set to present (possibly in groups). 20% of your grade will consist of that presentation. The problem sets will involve implementation of different methods in R and interpretation of the results.

Project: The deadline for the individual project will be TBA, as the deadline has to be least 4 and at most 6 weeks after the topics were approved/assigned (this is stipulated by the examination office). This would mean that the topics will be finally chosen in the middle of January, where we will do a "mini-workshop" ($\approx$ two lectures), where everyone introduces their topic and we can discuss some issues that might arise. The project will involve a simulation study with a realistic empirical set-up using one of the methods discussed in class in a social science setting.