Introduction to Econometrics with R

Boris Demeshev

Course description

This course presents a short introduction to econometrics with hands-on applications using R statistical software. Prospective students are required to have basic knowledge of probability and statistics. No previous knowledge of programming is assumed.

We cover standard topics as linear regression and logistic regression. We also include topics usually omitted from introductory econometrics courses: LASSO, classification trees, random forest, clusterisation and dimensionality reduction.

Prospective students are strongly encouraged to install R and Rstudio beforehand and to study the first chapter of the course "Introduction to the tidyverse" at https://www.datacamp.com/courses/introduction-to-the-tidyverse.

Topics

1. Data manipulation in R

Importing and exporting data. Filtering observations, selecting variables, arranging observations. Grouping operations and summarising dataset.

2. Graphs in R

Plots for continous variables. Plots for cathegorical variables. Plots for combinations of cathegorical and variables.

3. Linear regression

Non statistical properties of estimators Statistical inference in regression. Comparing nested models with F-test. Doing a lot of regressions. LASSO.

4. Reporting with R markdown

Markdown without R. Code chunk parameters. Short introduction to LaTeX. Introduction to github (if time permits).

5. Tree related methods

Trees for cathegorical data. Trees for continuous data. Random forest. Visualising a tree.

6. Logistic model

Logit and probit models. Ordered logit model. Visualising logit model.

7. Dimensionality reduction

Principal component method. Singular value decomposition approach (if time permits). Stochastic neighborhood embedding.

8. Clusterisation techniques

K-means clusterisation. K-means regression. Hierarchical clusterisation.

Grading policy

Four practice minicourses at datacamp — 40%, final written exam — 60%.

Literature

1. Hadley Wickham, R for data Science, http://r4ds.had.co.nz
2. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Introduction to Statistical Learning, http://www-bcf.usc.edu/~gareth/ISL/