

Data Analytics II, Spring 2022

Instructor: Rickard Sandberg

Course Description

In this course, the students will learn the basics of Econometrics (for cross-sectional data) and get an introduction to Machine Learning (ML). These learnings will form the basis for methods and applications in the subsequent course Data Analytics III. This course also includes further essential mathematical methods, building on the mathematical methods taught in the Data Analytics I course, that are used in economic and financial analysis and other areas. The importance of data science to businesses is discussed, with a primary focus on the opportunities and challenges that business school students face when encountering this field. Applications and practical implications are addressed, and the students will get further inspiration by a guest lecture (by people from the Swedish industry).

This course covers important econometric concepts such as: random sample, central limit theorem, estimation, inference (hypothesis testing), regression models, causality, randomized experiments, as well as the ML concepts: supervised learning (prediction and classification), unsupervised learning (clustering). This course also introduces some mathematics: integration, topics in financial mathematics, and matrices.

In this course the program R is used to facilitate the learning of Econometrics and ML. The students will use different R-packages for the analysis of regression models, causality, randomized experiments, prediction, classification, clustering, etc., and the students will also learn how data and results can be visualized.

Intended Learning Outcomes

After successfully completing this course students should be able to:

- solve problems that require knowledge about standard concepts of Econometrics and ML
- recognize the importance of data science to businesses
- perform basic Econometric and ML analysis on real-world data sets
- solve problems in Financial Mathematics and understand the usefulness and power of matrices

Format & Structure

Written assignments, cases, group work, lectures, seminars, tutoring, guest lecture(s) and exercises, both with and without the use of statistical software (R). Explanations and applications relevant to business and economics. An individual case study on real-world data using R.

Prerequisites

One course in both micro and macroeconomics, statistics, calculus, and working knowledge of statistical software (R).

It is highly recommended that the prerequisite high school mathematics is well reviewed before the start of the course.

Literature

Newbold, Paul, Carlson, W.L. & Thorne, B., Statistics for Business and Economics, 9th ed., 2020. Chapters 6 – 14. ISBN 9781292315034.

Introduction to Data Science - Data Analysis and Prediction Algorithms with R (<https://rafalab.github.io/dsbook/>) (Available online.)

Supplementary material on Causality and Randomized Experiments (to be handed out).

Examination

The first and second intended learning outcomes (see below) will be assessed through a written exam at the end of the course. The third intended learning outcome will be assessed through a written report as well as the written exam.

The fourth intended learning outcome will be assessed through a separate written exam.

The final grade for the course is based on hand in assignments, 10%, and the written exam, 90%.