

Syllabus



Department
of Economics
and Business

CENTRAL
EUROPEAN
UNIVERSITY

Data Analysis 3: Pattern discovery and regression analysis

- **Instructor:** Gábor Békés (bekesg@ceu.edu office hours: Wednesday 16-17.30 by appointment)
- **Credits:** 2 (4 ECTS)
- **Term:** Fall 2017-2018
- **Course level:** [MA/MSc]
- **Prerequisites:** Data Analysis 1 and Data Analysis 2

Course description

Uncovering patterns in the data can be an important goal in itself, and it is the prerequisite to establishing cause and effect and carrying out predictions. The course starts with simple regression analysis, the method that compares expected y for different values of x to learn the patterns of association between the two variables. It discusses nonparametric regressions and focuses on the linear regression. It builds on simple linear regression and goes on to enriching it with nonlinear functional forms, generalizing from a particular dataset to other data it represents, adding more explanatory variables, etc. We also cover regression analysis for time series data, binary dependent variables, as well as nonlinear models such as logit and probit.

Learning outcomes

By successfully completing the course the students will be able to:

- Successfully formulate research questions that are answerable by empirical analysis;
- Produce meaningful descriptive statistics and informative graphs;
- Carry out simple regression analysis;
- Discuss and interpret results, understand validity and constraints.
- Present empirical analysis and write short reports with data;

Reading list

Data, codes and handouts will be provided. The exam material is composed of slides. All material will be available at ceulearning.

Assessment

- Start-of-the-class quizzes (10%), based on the material of previous class. 4 quizzes in class, 1 extra after the exam. Best 4 out of 5 will be considered.
- 3-4 assignments including a term project (40%). Individual and in teams. The term project deadline is 15 December.
- Closed book, pen and paper exam (50%)

Grading policy

- Students shall not miss more than 2 lectures and more than 1 seminar. Failing to do so will yield an administrative fail grade. (If you have a major impediment please contact the Instructor.)
- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the exam. Failure to do so, will yield a Fail grade.

Course schedule and materials for each session

1. Simple regression analysis. Nonparametric regression. Linear model. Uncovering the parameters of simple linear regression. Predicted dependent variable and the residual. Goodness of fit.
2. Feature engineering, uncovering non-linear patterns in regression analysis. Measurement error.
3. Generalization, external validity. Confidence interval of regression coefficients.
4. Multiple linear regression analysis. Categorical explanatory variables, interactions. Omitted variable bias, bad controls. Modelling and interpretation.
5. Probability models. Linear regression with binary outcome: the linear probability model. Nonlinear probability models: logit and probit. Coefficients and marginal differences.
6. Analysis of time series data