Introduction to Data Analysis 2

Ágoston Reguly

Data Analysis 2: Finding Patterns with Regressions

2024

This course

- This course introduces uncovering patterns of associations with regression analysis.
- Modelling with cross-sectional data where dependent variable is continuous or binary and basic time-series analysis.

Discovering

- We assume you have already an approximately well cleaned and organized data.
- Data Analysis 2 is about discovering patterns.

Discovering

- We assume you have already an approximately well cleaned and organized data.
- Data Analysis 2 is about discovering patterns.
  - This is the fun part!
  - Unfortunately, this is only a fraction of your working time.

Discovering

- We assume you have already an approximately well cleaned and organized data.
- Data Analysis 2 is about discovering patterns.
  - This is the fun part!
  - Unfortunately, this is only a fraction of your working time.
- Proper discovery means strong knowledge on statistical tools
  - Understanding the theory takes time.
  - Using theory in computer takes few seconds...

Data Analysis 2: Patterns - topics

1. Simple Regression (non-parametric and parametric, simple linear regression's anatomy, model summary)
2. Complicated patterns and messy data (transformations and more advanced functional forms, influential observations, measurement errors, weighted regression)
3. Generalizing results of a regression (SE of coeff, CI, prediction intervals, hypothesis testing, external validity)
4. Multiple linear regression (using more xs, omitted variable bias, inference, variable selection)
5. Probability models (binary regression models: LPM, probit, logit, non-linear regression, marginal differences, model evaluation)
6. Time series models (time series properties, (non)-stationarity and random walk, seasonality, type of trends, serial correlation, leads and lags, SARIMA models)

Office hours

Contact me any time!

- regulya@ceu.edu OR areguly6@gatech.edu
- *Office hours*: by appointment.

We have a great TA:

- Bálint Thaler: thaler.balint@gmail.com
- Feel free to reach him with any questions!
- He will help you with problems and questions and conduct a seminar on 25th of November.

Course Material - Data Analysis Textbook

Gábor Békés and Gábor Kezdi:

- *Data Analysis for Business, Economics and Policy*
- Mixing intro statistics and key ideas from data science with case studies
- Official website: https://gabors-data-analysis.com/
- Github repo for case studies, codes and more

Part 2: Finding patterns with regression is for DA2

- 4 times - six chapters - handouts and book Part II.
- One seminar on the 3rd occasion - interpretation and coding with Bálint
- Slides on moodle

Coding with Python is a complementary to this course.

Quiz and Assignments

- Start-of-the-class Quizzes (only for lectures)
  - Past lecture material
  - Simple question, close to practice questions at the end of handout chapter and similar to the exam.
  - 2 quiz 5-5 points $\implies$ 10/100p
    - 11th of November and 2nd of December
  - 5 short questions: multiple answers OR True/False.
- Individual assignments
  - Assignment 1 (10p) – individual short analysis on a given topic.
    - Deadline: 24th of November (Sunday) 23.55 on moodle.
  - Assignment 2 (20p) – analysis of a selected topic in pairs.
    - joint with Coding in Python (lecturer: Péter Duronelly)

Assignment 2 - 20p

Deadlines:

- 8th of December: approved pairs and dataset by me via email. (you have to receive an OK email from me)
  - Best if you bring your own data, that you have already used or you plan to use in your work!
  - If this is not an option, you may want to use data from world bank, OECD, Eurostat or other such statistical resources.
  - Should be your work and not using popular data/code repositories such as Kaggle, unless you can prove that it is your work.
  - Minimum requirements: 40 or more observation, cross-section OR time-series. If cross section, continuous or binary outcome. Explanatory variables: at least 4 variables which can be included in the regression. Out of 4 variables minimum 2 variables which are continuous.
- 5th of January, Sunday 23.55.am. (upload to ceu-learning site)
  - Finished work.
  - Late submission: 1-2 day delay -50%, after that no points.

Exams

- Mid-term exam (20p) – on 25th of November, first 45 mins
  - Topic is based on BK Chapter 7-9, all sections unless otherwise noted
  - Closed book exam
  - 45 minutes, short questions, interpretations and T/F
  - May practice using the practice question section after each chapter.
- Closed book exam on 16th of December 13.30-15.10 pm CET
  - Textbook chapter 7-12 (all material covered in the course)
  - All sections unless otherwise noted
  - Closed book exam
  - 90 minutes, short questions, interpretations and T/F

Grading policy

- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the final exam.
- Lectures - can not miss more than 2 - measured as quiz, exam submitted.

Extra

- End of chapter - Data exercises
  - Submit any 3 to get bonus points
  - * Easy/quicker - 1p
  - ** Harder/longer - 2p
  - Deadline: 15th of December, 23.59
- Suggesting other useful resources/materials with short presentation (2-5 min in class)
  - Reference your resource/material on slack channel
  - Scientific article - 2p
  - Useful forum/community - 1p
  - Deadline: 2nd of December, last class
- Max extra points in DA2 overall is 6p for everybody.