Introduction to DA2

Ágoston Reguly

Data Analysis 2: Finding Patterns with Regressions

2020

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

As usual feel free to write me any time!

- reguly_agoston@phd.ceu.edu
- *Office hours*: Monday 10:30-12:00 or by appointment at N13 220.
  - Except: 16th and 30th of November: no office hours, instead I will be available on Wednesday during the same time-slot.
- Weekends - typically, I am not checking my email.

Course Material - Data Analysis Textbook

Gábor Békés (CEU) and Gábor Kezdi (U. Michigan):

- *Data Analysis for Business, Economics and Policy* (It is almost done...)
- Mixing intro statistics and key ideas from data science with case studies
- Official website: `https://gabors-data-analysis.com/`
- Github repo for case studies
- Data also available - read the 'readme.rtf' files about availability and for ethical usage!

Part 2: Finding patterns with regression is for DA2

- Ten lectures - six chapters - handouts and read-only pdf
- Two seminars - similar to coding
- Slides on moodle

Coding with R (2 credits / 12 weeks) is a complement to this course.

Quiz, Exam, Assignment

- 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.
  - 9 quiz - consider best 8, each 1.25 points.
  - Short answers: 1 or 2 questions OR True/False.
- Individual assignments
  - Wait 3 slides...
- Closed book exam on 17th of December 5.30-7.00 pm CET
  - Textbook chapter 6-12
  - All sections unless otherwise noted
  - 75 minutes, short questions and T/F, like in DA1

Grading policy

- 10% quiz, 40% assignments, 50% exam
- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the exam.
- Lectures - can not miss more than 2 - measured as quiz submitted.
  - In case of online participation: write me with the reason and in the next class you can do both quizzes.

Extra

- End of chapter - Data exercises
  - Submit any 3 to get bonus points
  - * Easy/quicker - 1%
  - ** Harder/longer - 2%
- Suggesting other useful resources/materials with short presentation (2-5 min in class)
  - Reference your resource/material on slack channel
  - Scientific article - 2%
  - Useful forum/community - 1%
- Max extra points in DA2 overall is 6% for everybody.
- Extra, above 6%: pointing out typos or mistakes in the book
  - Competition: 1st 3%, 2nd 2%, 3rd 1%
- Deadline: 17th of December, 23.59

Assignment 1 - 10%

Carry out simple linear regression analysis with given dataset (10%).

- Joint assignment with Coding 1.
- Create proper folder structure with: data (raw and clean), codes (rmd and if want an .R) and docs (.html and .pdf generated from .rmd)
- For DA 2 the pdf/html is graded.
- Exact grading system TBA.
- Deadline: 29th November Sunday 23.55.am. (upload to ceu-learning site)

Assignment 2 - 30%

Carry out a complete data analysis for your chosen dataset. This assignment should contain all the materials that we have learned in DA1, DA2 and in Coding! (30%)

- Create proper folder structure with: data (raw and clean), codes (rmd and if want an .R) and docs (.html and .pdf generated from .rmd)
- Joint assignment with Coding 1 - thus you will use github.
- For DA 2 the pdf/html is graded.
- Exact points TBA.

Assignment 2 - 30%

Deadlines:

- 6th of December: approved dataset by me via email.
  - 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.
  - 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.
- 3rd of January, Sunday 23.55.am. (upload to ceu-learning site)
  - Finished work.
  - Late submission: 1-2 day delay -50%, after that no points.