Reproducible research on child and youth development (pt. 1)

Michael Kühhirt

michael.kuehhirt@uni-koeln.de

Summer 2018

Office Hours: WED 02:00-03:00 p.m.   Class Hours: TUE 02:00-5:30 p.m.
Office: Room 0.05, Greinstraße 2   Class Room: SocioLab, Greinstraße 2
Web: goo.gl/rszqWJ

Class content

Are there differences between boys and girls in terms of behavior, verbal ability, or math skills, and if so, why? How do factors like parental divorce or poverty affect children? What are the consequences of maternal smoking during pregnancy in terms of children's health and development?

In this two-part research lab, participants carry out their own empirical project on such or similar topics in the area of child and youth development using data from the Children of the National Longitudinal Survey of Youth. During the first semester, participants develop a research question based on a previous study using these data and start with data preparation. The semester concludes with a proposal that specifies the research question, outlines the motivation and theoretical background, and describes the data and methods used to answer it. The second semester is devoted to finalizing data preparation, to carrying out the data analysis, and to writing the research report.

A key goal of the course is to establish a workflow for *transparent* and *reproducible* empirical social research (on any topic). To this end, participants practice how to prepare research data (e.g., merging data sets, recoding variables, constructing analytical samples), how to efficiently produce formatted research output (e.g., tables and figures), and, most

importantly, how to rigorously document these tasks so that others (including one's future self!) can understand and repeat each step of the research process and build on the work.

Goals and learning outcomes

The class has three broad goals, which I enumerate below. Each goal is associated with a number of specific learning outcomes, which you should be able to perform as the class concludes after two semesters. Students who complete the class should be able to

1. formulate an empirical research question based on existing literature. This includes
   - providing an overview of existing research on a specific topic,
   - motivating their research question with coherent theoretical arguments,
   - systematically building on (and possibly extending) existing research on a specific topic.
2. prepare and analyze existing survey data. For this, students
   - understand different data structures,
   - explore, clean, and combine different data,
   - use statistical methods appropriate to answer their research question.
3. document their empirical research project by
   - writing a research report,
   - providing materials to reproduce their research from data acquisition through data preparation to data analysis,
   - presenting their research project in class.

Although the substantive focus of the class is on child and youth development, all learning outcomes can be readily transferred to other topics in empirical social research. Many of the skills practiced in class (e.g., summarizing existing knowledge, script-based data preparation and analysis, writing and presenting) are also valuable for jobs outside academia.

Requirements and grading

To successfully complete the class, you must meet the following requirements. In the first semester (Summer 2018), you

- present an empirical study on a specific topic in child or youth development that you want to replicate in your research project,
- write a proposal summarizing your research question, theoretical background, and analytical approach.

In the second semester (Winter 2018), you

- present a progress report on your research project in class,
- write a research report along with documentation to reproduce your project.

In addition, there will be homework assignments designed to facilitate successful class completion. Although these assignments aren't graded, I highly recommend to work through them within the appropriate time period in order to regularly review and practice class contents. In some cases, the homework is also immediately related to the class requirements outlined above. Homework 1 contains information on how to find an appropriate empirical study for replication. For details on the content and structure of the presentation of your empirical study, see Homework 8. In Homework 10 you'll find instructions for writing the research proposal.

Research project

In your research project, you replicate an existing empirical study on child or youth development in whole or in part. The primary goal is to competently carry out some or all of the empirical analyses in the study, to appropriately motivate the analyses, and to document your project workflow so that it can be understood and reproduced by anyone who has access to the original data. An exceptional research project additionally extends the original analysis in a meaningful way, for example, by

- looking at children at a different age,
- exploring group differences,
- investigating the mechanisms underlying some statistical relation,
- using additional or alternative control variables,
- applying alternative statistical methods.

The final report should emulate an empirical article in a typical social science journal. Guidelines and techniques for project documentation will be discussed and applied as the class progresses. I'll provide more details on the structure, layout, and grading of the report at the beginning of next semester.

Bonus points

You can improve your final grade by one decimal increment (e.g., from 1.7 to 1.3 or from 3.0 to 2.7) by adequately answering a total of 10 questions posted by your classmates on Piazza, an online platform for class Q&A (see class resources below). For your answer to count towards your grade

it must completely solve the initial query, including follow-up questions and clarifications, within 24 hours of the first post. Answers to trivial questions (e.g., "When is Homework 1 due?" or "Where do I find the slides?") don't qualify in this context. This bonus point policy is subject to review after the first semester.

Workload

Upon successful completion of the class you are awarded 15 credit points, which equals a total workload of 450 units à 45 minutes. Per semester, there are only 60 units (i.e., 3 hours/week) of class time. The remaining 165 units per semester (i.e., roughly 8 hours/week) are allotted to preparing and reviewing class contents, including homework assignments, reading, writing etc. Therefore, please schedule sufficient time outside of class hours.

Class schedule

April 10:
- Introduction
- Lecture: Research on Child and Youth Development
- Homework 1 (Find An Empirical Study)

April 17:
- Lecture: The National Longitudinal Survey of Youth (NLSY)
- Lecture: A Workflow for Reproducible Research
- Homework 2 (Set Up The Workflow)

April 24:
- Discussion: Homework 2
- Lecture: Data Exploration, Commenting Code, Capturing Output
- Homework 3 (Explore The Data)

May 1: NO CLASS

May 8:
- Discussion: Homework 3
- Lecture: Data Cleaning I
- Homework 4 (Clean The Data)

May 15:
- Discussion: Homework 4
- Lecture: Data Cleaning II
- Homework 5 (Clean More Data)

- *Homework 1 Due*

May 22: NO CLASS

May 29:

- Discussion: Homework 5
- Lecture: Combining Data Using The Merge Command
- Homework 6 (Combine Data)

June 5:

- Discussion of Homework 6
- Lecture: Loops and Basic Automation
- Homework 7 (May The Loop Be With You)

June 12:

- Discussion: Homework 7
- Exercise: Putting It All Together
- Homework 8 (Present Your Study)

June 19:

- Presentations of Empirical Studies I
- Presentations of Empirical Studies II

June 26:

- Presentations of Empirical Studies III
- Presentations of Empirical Studies IV
- Homework 9 (Be An Investigator)

July 3: NO CLASS

July 10:

- Discussion: Homework 9
- Lecture: Working With NLSY Raw Data
- Homework 10 (Write A Proposal)

August 12: *Homework 10 Due*

Class resources

Materials on Ilias (UoC e-learning platform)

Here you'll find slides and other class materials:
https://www.ilias.uni-koeln.de/ilias/goto_uk_crs_2459216.html

Q&A on Piazza

Piazza is an online platform to facilitate class discussion. Please use it to post any query regarding class content and organization, including but not limited to questions about homework assignments. The system is highly catered to getting you help fast and efficiently from your class-mates and me. Use email only for personal matters such as making an office hour appointment or reporting an illness. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class page at:
https://piazza.com/uni-koeln.de/spring2018/specialisationmodulei/home

Piazza is also available cost free as an app for Android and iOS.

National Longitudinal Survey of Youth 1979 (NLSY79)

For practical exercises and class projects, we'll use data from the NLSY79, a repeated survey of men and women in the United States born between 1957 and 1964, along with data from the children born to NLSY79 women.

To access these data, use the NLS Investigator (after registration):
https://www.nlsinfo.org/investigator/pages/login.jsp

Following the link below you'll find a database collecting publications that have used data from *Children of the NLSY79*. You can search this database using keywords that reflect your specific research interest.

https://www.nlsinfo.org/bibliography/search/cohort+keyword/chrt=Children+of+the+NLSY79

Open Science Framework (OSF)

OSF is an open source software project for collaboration in scientific re-search. We'll use it to regularly back up our code files and to submit completed homework assignments.

To use OSF, please create an account at https://osf.io and then fork the project template at https://osf.io/nx7q8/.

Other resources

Birth cohorts worldwide:
http://www.birthcohorts.net/

Developing a Child Cohort Research Strategy for Europe:
http://www.chicosproject.eu/

Project TIER - Teaching Integrity in Empirical Research:
http://www.projecttier.org/