Data Analysis for (Policy) Consulting

Chapter 1: Introduction

Time and Date  
Lecture:  
Wednesday, 12:15-13:45, SR226  
Start: 16 October 2019

Tutorial:  
Thursday, 14:15-15:45, CIP1  
Start: 17 October 2019

Credit  
Wahlpflichtbereich: "Ausgewählte Probleme der Volkswirtschaftslehre" (5 ECTS)

Exam  
Written Exam (60 min.) or oral exam (20 min.), depending on number of participants.
Course Material

The lecture slides and tutorial exercises will be available in WUEcampus one week in advance:

https://wuecampus2.uni-wuerzburg.de/moodle/course/view.php?id=34683

Datasets and further Material for the lab tutorials will be available at the course network drive:

\wifak-server.rz.uni-wuerzburg.de\WIFAK\USERS\Public\Kurse\DAPolicyConsulting\Stud_R
Literature:

This course follows the textbook (a few copies are available at the Teilbibliothek WiWi):


Additional sources which will be provided as we go along

In this chapter:


During the lecture, applications of the material will be discussed in **Case Studies**. These often relate to other sources as well
1.1 Motivation

- Applied economics is as exciting as any science can be!

We study *causes* and *effects* in many important questions:

- Does a mandatory health insurance make people healthier?
- Do graduates from private colleges make more money than graduates from public colleges?
- Why are women’s wages lower than men’s wages?
- Will loose monetary policy actually spark economic growth?
- Does a minimum wage destroy jobs?
- Should the police arrest suspects of domestic abuse?
- ...

- Applied economics is a science of measurement:

  ‘The Economy’ generates a lot of data on people, jobs, firms, etc. that potentially offer answers to the questions we ask.

But the true answers don’t reveal themselves easily. Often we see some relationship in the data and are tempted to interpret this as cause and effect. *This is extremely dangerous!*
Correlation vs. Causality (1)

- Imagine your boss shows you this data and wants to know if there is a causal link?

What if the issue is less clear?
Case Study: The effect of police on crime

- Imagine you consult a politician who wants to know if he can reduce crime by hiring more policemen. After some googling, you find data on robberies and police forces for US states:

![Graph showing the relationship between robberies per 100,000 citizens and sworn officers per 1,000 citizens. The graph includes states such as Nevada, Maryland, New York, Washington, Arizona, Ohio, Tennessee, Mississippi, and Wyoming. The slope of the line is 38.2 with a standard error of 12.1.]

Sources:
own calculations

- Would you conclude that more cops increase the crime rate?

*You think this is a stupid question??*
Correlation vs. Causality (2)

- According to a folktale, a czar found out that the most disease-ridden province in his empire was also the one with the most doctors. His solution? Kill all doctors!

_How about a more complicated example?_

- The election candidate who has more campaign money usually wins. Does this mean that campaign money determines election outcomes?

- A **correlation** simply means that there is a relation between two factors, e.g. X and Y. In terms of **causality**, this can mean that:

- Our tools to measure correlations are the huge variety of statistical and econometric methods. Yet, there is not a single statistical test that indicates which of the above alternatives is true!
Other things equal

• Imagine you consider donating 1,000 Euros to a political candidate. When would you do this?

• Picture two candidates. One is very charismatic, the other one is totally bland. The first one raises more money and is more likely to win

• How can we answer the question “Does campaign money determine election outcomes?”

• We need to make an “other things equal” comparison. (Latin: “ceteris paribus” – sounds familiar?)

• The other things equal, i.e. causal interpretation is hard to engineer. There is no econometric method that does this automatically

• In this course, we discuss strategies to identify causal effects
1.2 Identification strategies

- On the search for causal effects, we gather and analyze data using the econometric methods you have learnt in the compulsory module “Grundlagen der Quantitaven Wirtschaftsforschung” (Don’t panic!! We’ll repeat what you need to know on an intuitive basis!)
- But the focus here will be on how to implement them in order to identify the causal effects we are looking for. This is what we call an identification strategy.
- The gold standard is a laboratory experiment. In order to find out whether a medicine works, a pharmacist would divide a population of cloned mice and administer a medicine to one group only. (Un)fortunately we can’t clone politicians
- Our best alternative is a randomized experiment. Researchers change the variable of interest for a randomly selected group (e.g. by a coin toss). The variable of interest is then likely to be unrelated to other factors that determine the outcome we study. This does not mean that subjects are equal, but other things equal holds on average between the groups

Unfortunately, no politician would participate in an experiment randomly assigning her/his money
• In addition, randomized experiments are very expensive, take a lot of time and might even raise ethical concerns

• Sometimes, nature or the government creates a “natural experiment” by accident
   (A prominent example: The US government conducted two lotteries to determine the order of call to military service during the Vietnam War)

• Fortunately, we have other research designs that are less powerful but more accessible:
  ▪ regression analysis
  ▪ instrumental variables
  ▪ regression discontinuity design
  ▪ difference-in-differences

• Still it is useful to start a research project with the consideration “What would the ideal experiment look like?”. This disciplines our use of these econometric tools and avoids making naïve mistakes
1.3 Approach of the course

There will be three types of lessons:

- Discussion of research designs and results on the basis of prominent studies
- Discussion of the relevant technical issues
- Replication of studies using real data and the statistics package STATA

Exam:

- You will be presented different research questions and how someone else tried to tackle them
- You will be asked to comment on this other person’s approach
- You will be asked to interpret and discuss the other person’s results tables and figures
- You will be asked to suggest your own strategy to answer the research question
1.4 Why should you be in this course?

- You learn how to draw *causal inference* from data
- Many wrong decisions have been made because decision makers took advice from people who can’t *distinguish correlation from causality*. You learn to avoid common *pitfalls* in the interpretation of results that are omnipresent in the media, the private and the public sectors
- You will be able to apply for job postings that require knowledge in “*data science*”
- We will talk about *interesting questions* and the creative ways smart people found to answer them