Parametric Statistics

Week 2 - Research Design

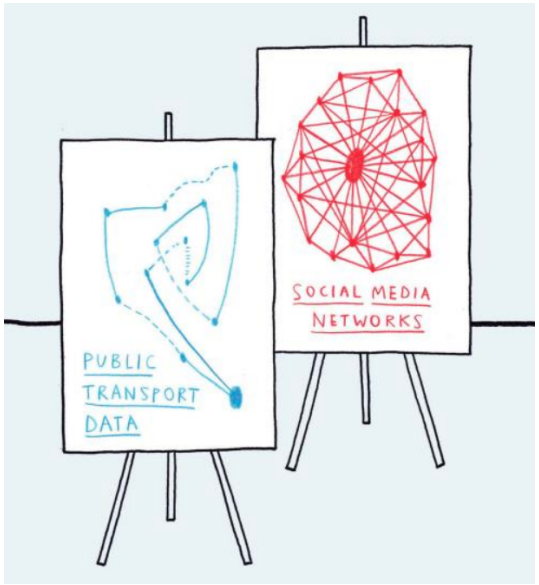
Juan Carlos Medina Serrano

Technische Universität München
Hochschule für Politik
Political Data Science

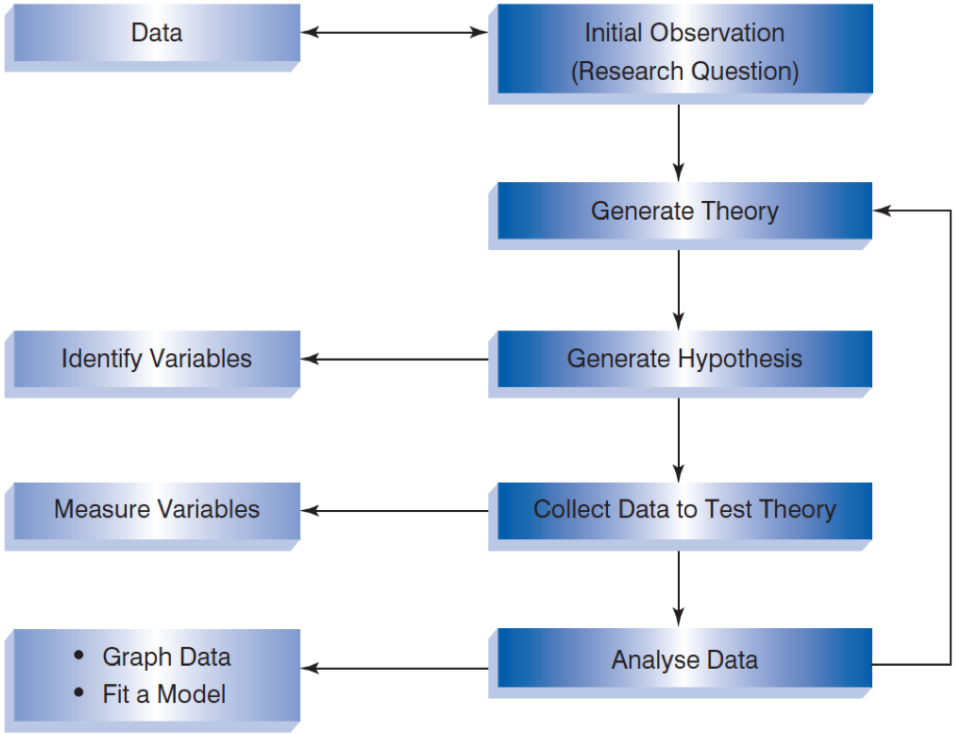
Munich, 23. October 2019

political data science

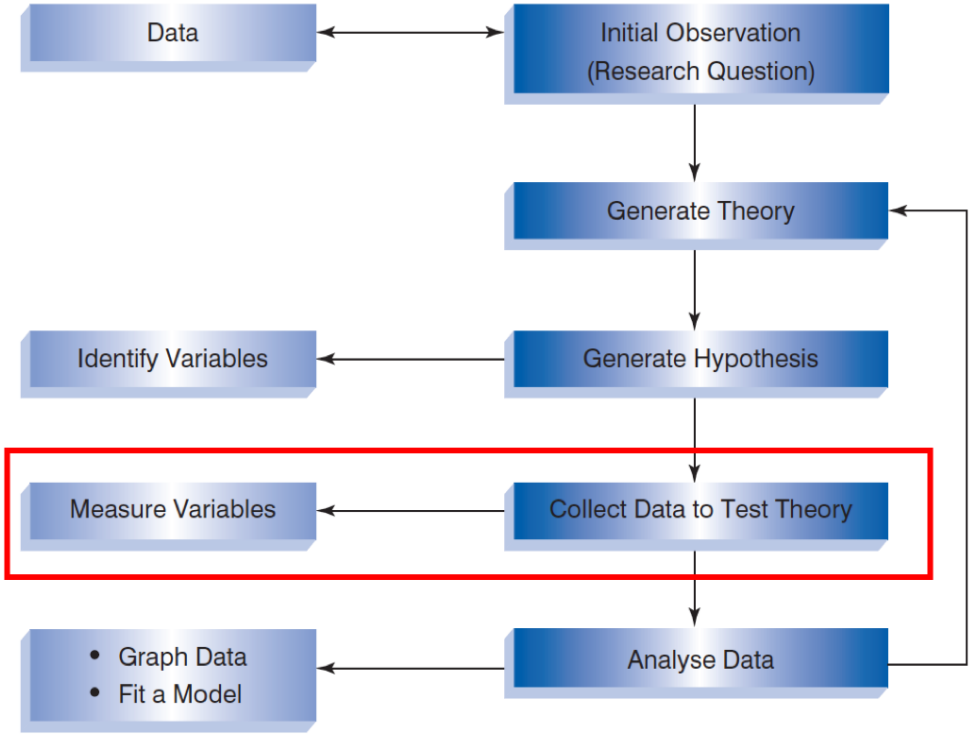
https://politicaldatascience.blogspot.de



The Research Process



The Research Process



Variables

Predictor variables: A variable thought to be the cause of some effect. It helps predict an outcome variable. (a.k.a independent variables)

Outcome variables: A variable thought to change as a function of changes in a predictor variable. Our hypothesis are based on these variables (a.k.a dependent variables)

Measurement Errors

Discrepancy between the numbers we use to represent the thing we're measuring and the actual value of the thing we're measuring

Types of error:

- **Random error:** due to chance. Not related to the true value.
- **Systematic error:** has an observable pattern. Should be identified and remedied.

Measurement Standards

Two standards: Reliability and Validity

Reliability: Quantifies how consistent are repeated measurements.

- Test-retest Reliability: Repeated tests over time. Measure temporal stability.
- Parallel-forms Reliability: Perform different versions of a test that should come to the same result.

How to calculate it? Correlation (more on this in a couple of weeks) or percentage agreement (same results/all results)

Validity: Are we measuring what we intended to measure?
Are IQ tests a real measure of intelligence?
Proxy Variables substitute one measure that is hard (or impossible) to measure for another which we can measure.
Hard to prove validity! Theories are needed!

Measurement Bias

Bias is disproportionate weight in favor of or against an idea or thing.

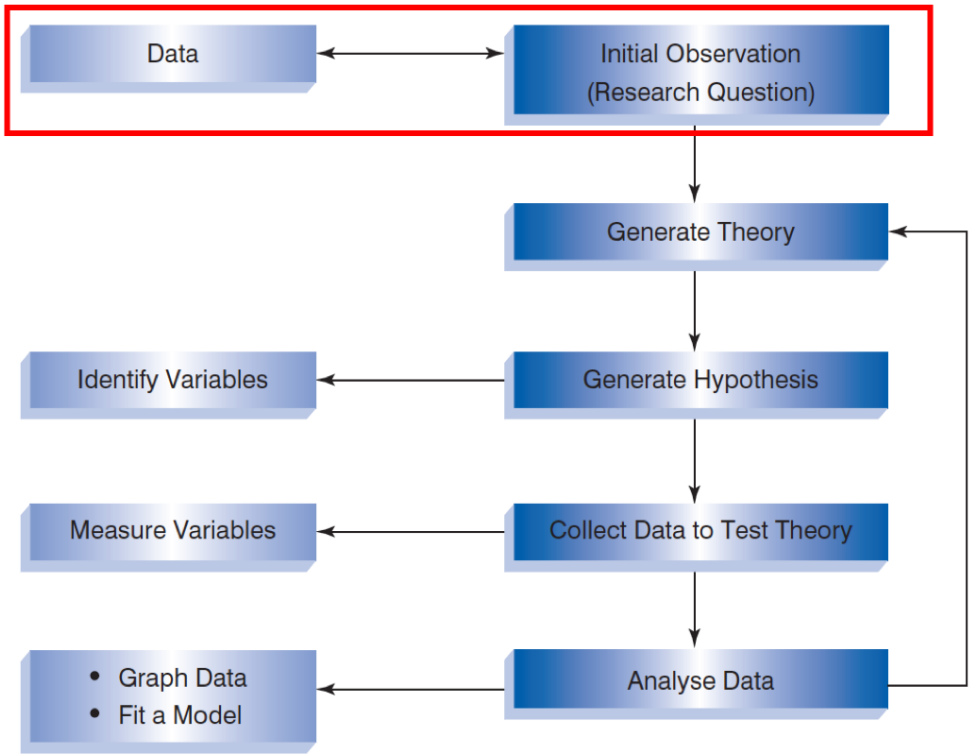
Bias can enter experiments in two ways (especially with survey data) :

- **Sample selection bias:** The sample is not representative of the population.
  - Selection bias: Some subjects are more likely than others.
  - Volunteer bias: Only taking into account volunteers, does not mean they represent the population.
  - Non-response bias: People who decline to participate may introduce bias.

- **Information bias:** Bias through the data collection methods.
  - Interviewer bias: Is the interviewer biased into selecting the information/people to interview.
  - Recall bias: False importance to past events.
  - Social desirability bias: People desire to present themselves in a favourable light.

Be aware! Bias leads to false conclusions, even if we apply the statistical tests correctly.

The Research Process



Finding a research question

This is why we are here
Finding something interesting in the data!

Most scientific questions imply a causal link between variables. X causes Y.
Proving causality is very hard outside the natural sciences

Problems: Causality or just correlation? `https://www.youtube.com/watch?v=VQ2bjy71PAc`

Confounding variables: Extraneous factors that affect the predictor and outcome variables.
Example Suicidal patients eat less, does this mean that eating less predicts suicidal traits? Probably both are caused by another psychological factor

Two types of research

Correlational research: Observing what naturally goes in the world. Researcher has no influence. Variables should not be biased by the presence of the observer!

Experimental research: Measurements and manipulations are under research control. They provide a comparison of situations (treatments) in which a proposed cause is present or absent. They use a control group and a treatment group.

- within-subjects design: same persons, different treatments.
- between-subjects design: different groups of persons, different treatments.

Two types of research

Correlational research: Observing what naturally goes in the world. Researcher has no influence. Variables should not be biased by the presence of the observer!

Experimental research: Measurements and manipulations are under research control. They provide a comparison of situations (treatments) in which a proposed cause is present or absent. They use a control group and a treatment group.

- within-subjects design: same persons, different treatments.
- between-subjects design: different groups of persons, different treatments.

Goal of a good research design: Minimize systematic errors and biases.
How?
Randomization: Randomly assigning subjects to treatments.
Blinding: The subjects (and sometimes also the researchers) do not know who is in the control and who is in the treatment group.

Experimental research: Variability

Idea: Measure variation between control and treatment groups

- Systematic variation: This variation is due to the experimenter doing something to all of the participants in one condition but not in the other condition.
- Unsystematic variation: This variation results from random factors that exist between the experimental conditions

The role of statistics is to discover how much variation there is in performance, and then to work out how much of this is systematic and how much is unsystematic.

Literature

- Discovering statistics using R : Chapter 1 (except 1.7)
- Statistics for people who hate statistics : Chapter 6
- Statistics in a nutshell: Chapter 1 and 5