Week 3: Threats to valid inferences 1

Internal and External Validity

POLS0007
Principles of Social Science Research

University College London

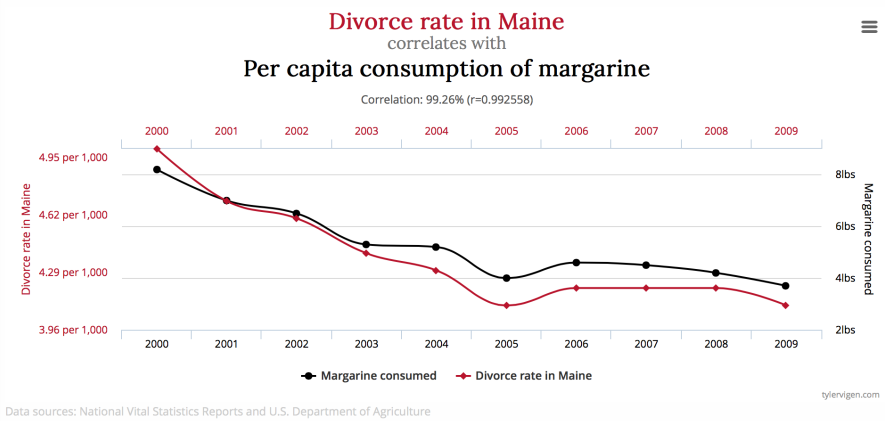
1. Causal Inference: Association vs. Causation
2. Internal Validity
3. External Validity

Causal Inference: Why is it so important?

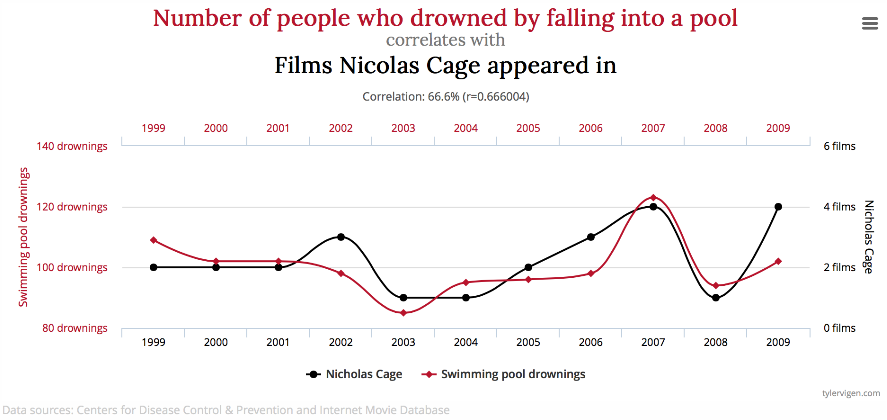
- **Research questions** that motivate science are **causal**
- The ultimate aim is to *infer the consequences* of a change in the world (treatment/policy)
- Examples:
  - Effect of an advertising campaign (X) to cut smoking (Y)
  - Effect of drinking (X) on health (Y)
  - Effect of schooling (X) on earnings (Y)
- Generally it difficult to determine **causal relationships**; More often data infer **associations/covariation/correlation** among variables
  - Recall the 4 hurdles: association is only 1/4

Association vs. Causation

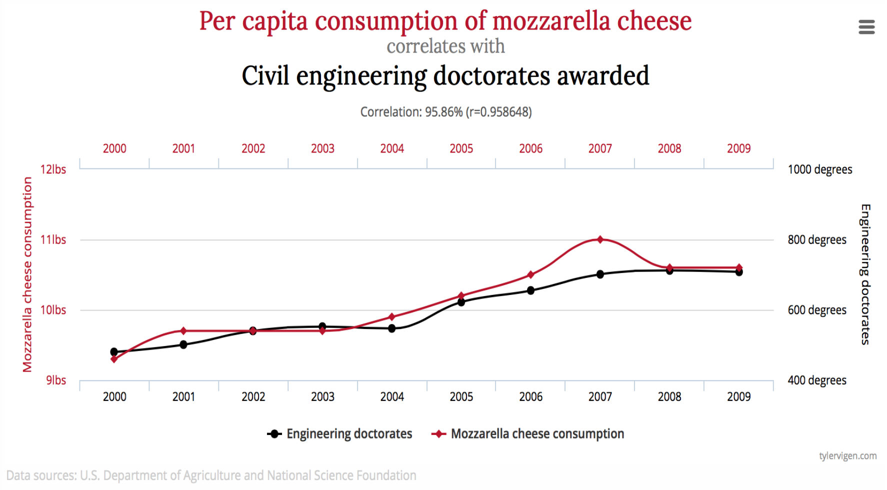
- Association/covariation/correlation means that two variables **move together**
- Causation means that one of the variables (X) is causing, at least part of, the movement in the other (Y)
- Association does not imply causation!
    - Incredible associations (hurdle 1)
    - Reverse causality: $Y -> X$ (hurdle 2)
    - Confounder: $Z -> X$ and $Z -> Y$ (hurdle 4)
- Examples of associations: `http://tylervigen.com/spurious-correlations`



This graph shows that the divorce rate in Maine moves in the same way as the per capita consumption of margarine.



This graph shows that the number of people drowning in pools is associated with the number of films Nicolas Cage appeared in over time.



This graph shows that the consumption of mozzarella is associated with the number of people who get a doctorate in engineering.

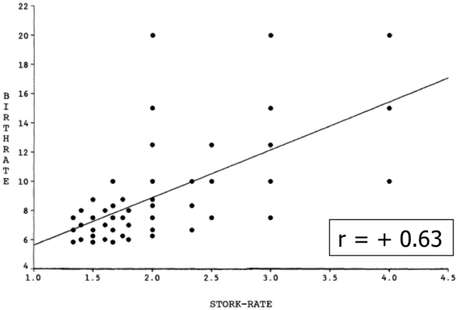
Storks and Babies

- Presence of storks (stork rate) correlates with fertility levels (birth rate). Correlation = 63%

A long contested issue:

do storks bring babies?

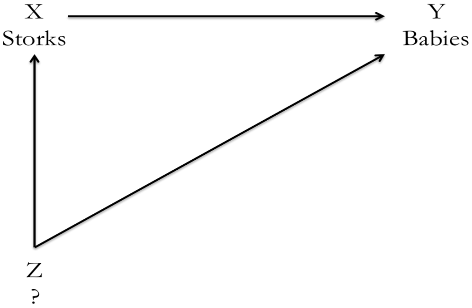
Neyman (1952) & Kronmal (1993)



This graph shows the correlation between birth rates and stork rates in different geographical areas.

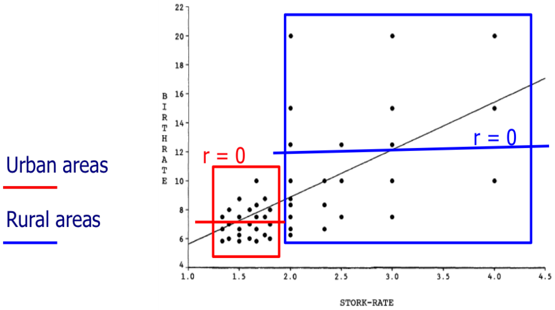
Do Storks (X) Bring Babies (Y)?

- The higher the **presence of storks** the higher the **birth rate**
- We are tempted to conclude that storks bring babies and *cause* the increase of the fertility rate.
- Is there **an alternative explanation (Z)** to the observed relationship?



What can Z be?

- If we look at the association between storks and babies, but splitting urban and rural areas, the correlation is now 0
- There is an external factor Z that influences both X and Y
- We cannot claim that there is a causal relationship



This graph shows the correlation between birth rates and stork rates in different geographical areas, separating urban (red) and rural (blue) areas.

Experimental Research Design

- How do we go about testing whether X causes Y?
- Conducting an experiment is – usually – the best way to study causality
- Randomized experiment: 'It is a research design in which the researcher both CONTROLS and RANDOMLY ASSIGNS values of the independent variable (X) to the participants'
- *Control*: the value of X cannot be determined by the participants
- *Randomly Assign*: the values of X are assigned to participants randomly

Internal and external validity of research designs

- Internal validity
  - The extent to which we are confident that results of a study establish causality between X and Y
  - Does the design cross hurdles 2 and 4?
    - Does the design eliminate reverse causality?
    - Does the design eliminate all/most confounders (Z)?
- External validity
  - The extent to which you can translate your knowledge to other settings
  - Whether the results can be generalized to other situations, people, countries, etc.
- Usually (although not necessarily) there is a trade-off between internal and external validity

Internal Validity

Suppose we want to test whether a vaccine reduces Chickenpox.

- Design 1: find those who received the vaccine and compare with those who did not
- Design 2: vaccinate all 2nd graders, compare them with 1st and 3rd graders
- Design 3: divide students randomly into a control (placebo) and a treatment (vaccine) group, compare them
  - Internal Validity is about confidence in causal inferences drawn from a design
  - Which design has the highest internal validity?

Threats to Internal Validity

1. Selection
2. Confounding
3. Design contamination (diffusion)
4. Experimenter bias
5. Mortality (differential drop-out/attrition)
6. Others: history, maturation, repeated testing, demand effect, instrumentality...

Threats to Internal Validity

1. Selection
    - Are there pre-treatment differences between comparison groups which may be related to Y?
    - Think about selection issues in Design 1, Design 2, and Design 3
    - Can there be a selection bias in Design 1 and 2 (hint: think about children's age)?

Threats to Internal Validity

2. Confounding
   - Is there another variable Z that can influence both X and Y?
   - Think about some confounders in Design 1, Design 2, and Design 3

Threats to Internal Validity

3. Design contamination
   - Did the control group know (or find out) about the treatment group?
   - Can treatment somehow diffuse to the control group?
   - Think about potential design contamination in Design 2 and 3

Threats to Internal Validity

4. Experimenter bias
   - Experimenter (subconsciously) behaving in different ways to control and treatment groups
   - Can there be an experimenter bias in Design 3 (hint: think about double-blind procedure)?

Threats to Internal Validity

5. Mortality/Differential drop-out/Attrition
   - Is dropping out systematically related to a feature in the design?
   - Can there be a mortality/differential drop-out/attrition problem in Design 3?

Threats to Internal Validity

- There are several threats to Internal Validity
- Randomization of treatment and control solves some of the problems (but not all)
- An experimental design with good internal validity takes those threats into account so that we have more confidence in a causal association between X and Y
- However experiments may have disadvantages:
  - External validity
  - Not every cause can be manipulated
  - Experiments can be ethically problematic
  - A causal effect established by an experiment may not be the most important one

External Validity

- Randomization of treatment/control versus random sample
- Randomization of treatment/control: internal validity
- Random sample: external validity
- Random sampling: every member of the population has an equal probability of being selected to be part of the study/experiment
- External Validity assures that if we replicate the same experiment on a different sample the results are the same, i.e. the results of an empirical investigation can be generalized to and across individuals, settings, and times.

External Validity of the Sample

- It is not the case that experiments require a random sample of the target population, so it is very rare
- e.g. in drug-trial experiments that offer some form of compensation, people who see and respond to the advertisement are not a random sample of the entire population (i.e. all potential recipients of the drug)
- The participant pool is called **sample of convenience**
- The way to know if the results apply to a broader population is **replication**: the same experiment is implemented repeatedly to see if the results hold

Threats to External Validity

1. **Differences in Populations (Population Validity)**
   - How representative is the sample of the population from which it was selected? The more representative, the more confident we can be in generalizing from the sample to the population.
   - How widely does the finding apply? Generalizing across populations occurs when a particular research finding works across many different kinds of people, even those not represented in the sample.
   - e.g. How different would be the results of our vaccination study had we implemented it in other countries (how about other species)?

Threats to External Validity

2. **Differences in Settings (Ecological Validity)**
   - The extent to which the characteristics of the setting or context of the research study are representative of the setting/context to which the results have to be generalized
   - e.g. computer lab vs. home

Take-home points

- Association does not imply causation!
- Association only makes you cross the 3rd hurdle
- Each research design has a level of internal and external validity
- Internal validity: how confident we are about the causal inference derived from a design
- External validity: to what extent we can translate knowledge obtained from a study to other settings
- Randomization in experiments alleviates many threats to internal validity but not all
- Experiments may have high internal but low external validity