estimand

Study design

Dr Gianluca Campanella

5th May 2016

Where we are

1. Define the **research question**
2. **Get** the data
3. **Explore** the data
    - (Re)format, clean, merge, stratify...
    - Identify trends and outliers
4. **Model** the data
    - Select and build model(s)
    - Evaluate and refine model(s)
5. **Summarise** the results
    - Summarise findings
    - Describe assumptions and limitations
    - Identify follow-up research questions

Contents

Research question

Why do we need a good question?

'Far better an approximate answer to the right *question, which is often vague, than an* exact *answer to the wrong question, which can always be made precise'*

— JW Tukey (1962), Ann Math Stat ***33****(1)*

What makes a good question?

Specific

- What exactly do you want to accomplish?
- Why?

→ Dataset and key variables

What makes a good question?

Measurable

- How will you demonstrate attainment of the goal?
- How will you evaluate it?

→ Type of analysis and major assumptions

What makes a good question?

Attainable

- Can the question be answered with the data at hand?
- Are conclusions likely to be biased?

→ Limitations and need for further information

What makes a good question?

Reproducible

- Could another person understand what you did?
- Could another person reproduce your results?

→ Documentation and standardisation

What makes a good question?

Time-bound

- What are the relevant time period and population?
- Can results be extrapolated?

→ Applicability and generalisability

Exercise: come up with a research question

1. Divide into groups
2. Go to `https://www.kaggle.com/datasets` and **choose a dataset** that looks interesting
3. **Identify** a research question
    - What type of data are available?
    - What is your outcome?
    - What are the SMART aims for these data?
4. **Share** your question with the class

Study designs

Study designs

Observational

The researcher studies, but **does not alter**, what occurs

Experimental

The researcher **intervenes to change reality**, then observes what happens

Observational studies

Sampling based on...

Exposure	Outcome	Neither
↓	↓	↓
Cohort	**Case-control**	**Cross-sectional**

Cohort studies

- Sampling based on **exposure**
- **Prospective**: exposure before outcome
- Lengthy → **attrition**

Example

- Select two groups:
  1. Smokers
  2. Non-smokers
- Follow-up for lung cancer after 10 years

Case-control studies

- Sampling based on **outcome**
- **Retrospective**: outcome before exposure
- May be biased by imperfect recall

Example

- Select two groups:
    1. Lung cancer patients
    2. Cancer-free 'controls'
- Ask them whether they have ever smoked

Cross-sectional

- All data are collected **at the same time**
- No distinction between exposure and outcome

Strengths

- Often population-based
- Less expensive than other designs

Weaknesses

- No direction of causality
- Over-representation of cases with longer durations

Experimental studies

Treatment randomly assigned?

Yes

↓

Randomised Controlled Trial

No

↓

Quasi-experiment

Randomised Controlled Trials (RCTs)

- Control for all main forms of bias
- **Ethical** concerns

Example

- Divide patients in two groups:
  1. Those who take the drug
  2. Those who take the placebo
- Evaluate influence of drug on disease course

Quasi-experiments

- More practical than RCTs (**natural experiments**)
- **Allocation bias**

Example (1854 Broad Street cholera outbreak)

- Public water pumps supplied by:
  1. Southwark and Vauxhall Waterworks Company
  2. Lambeth Waterworks Company
- High disease rate in districts supplied by 1
- Water obtained downstream from sewage discharge

Exercise: identify hypothesis and design

1. Divide into groups
2. **Read** "Worsening depression 'may predict dementia risk'" at `http://www.bbc.co.uk/news/health-36170259`
3. **Identify** the research question and the study design
4. **Discuss**