Causal diagram
Hypotheses
Research design

Lecture 2.b

21 Nov 2023

Learning objectives module 2

- Define and explain difference between
  - Concepts and variables
  - Unit of analysis and unit of observation
  - Propositions and hypotheses
  - Theories, research models and measurement models
- Formulate a simple causal diagram that links a few concepts

Learning objectives module 3

- Select a proper research design for a given research question or causal diagram
- Identify disturbing effects that threaten internal validity for a given research design
- Reflect on how research design allows claims on causality, internal and external validity
- Select a proper data collection method for a given research design

What is a causal diagram?

- Set of related hypotheses
  - Should answer the research questions
  - Should be motivated, preferably informed by literature review / meta-analysis
  - Should be on a specific unit of observation

What is a hypothesis?

- A formal statement of some unproven supposition that tentatively explains uncertain phenomena
  - Causal links: not `similar to' or `is part of'
  - Typically linking two or three variables

What is a unit of analysis?

Unit of analysis	Unit of observation
The major entity being studied	The unit described in the dataset
What the research question is about	What each data point is about
Examples: Individual, team, group, division, organization, alliance, industry, country	Examples: Individual, team, group, division, organization, alliance, industry, country

Causal diagram & hypotheses

Data sovereignty is important for data providers. Yet, its impact on the broader data economy is still unclear. Based on a survey from a UK representative sample (n=404), this study explores the implication of data sovereignty to four components of data economy. The findings reveal a direct positive relationship between data sovereignty and providers' willingness to share data. Perceived risk mediates this relationship, while trust in platform operators and consumers surprisingly does not.

- Make a list of key concepts of the study
- Define the unit of analysis
- Draw a causal diagram that links the key concepts

(Abbas et al., in preparation)

What is a hypothesis?

- A formal statement of some unproven supposition that tentatively explains uncertain phenomena
  - Causal links: not `similar to' or `is part of'
  - Typically linking two or three variables

Causal diagram & hypotheses

Data sovereignty is important for data providers. Yet, its impact on the broader data economy is still unclear. Based on a survey from a UK representative sample (n=404), this study explores the implication of data sovereignty to four components of data economy. The findings reveal a direct positive relationship between data sovereignty and providers' willingness to share data. Perceived risk mediates this relationship, while trust in platform operators and consumers surprisingly does not.

- Write down four hypotheses that link the key concepts
- For each hypothesis, write down if it is directional or non-directional
- For each concept, write down if they are independent variable, dependent variable, moderator or mediator

(Abbas et al., in preparation)

Validity: internal vs. external

- Internal = to what extent does the research design permit us to say that an independent variable *X* causes a change in dependent variable *Y*?

- External = To what extent can the results be generalized to a broader population or to a broader (organizational) setting?

Research design

Data sovereignty is important for data providers. Yet, its impact on the broader data economy is still unclear. Based on a survey from a UK representative sample (n=404), this study explores the implication of data sovereignty to four components of data economy. The findings reveal a direct positive relationship between data sovereignty and providers' willingness to share data. Perceived risk mediates this relationship, while trust in platform operators and consumers surprisingly does not.

- What research approach is selected in this paper?
- What would be threats to internal validity (=causal claims) in this study?
- What would be threats to external validity (=generalizability) in this study?

(Abbas et al., in preparation)

Causal diagram/hypotheses

Meta-platforms amplify data sovereignty concerns: the inability of data providers to control the shared data. Control mechanisms (i.e., certification, smart contract) help maintain data sovereignty. Nevertheless, the complex offerings of meta-platforms (e.g., enabling data flow from an IoT platform to others) make us question the efficacy of these control mechanisms. This study aims to evaluate the perceived efficacy of control mechanisms to maintain data sovereignty in meta-platforms. The findings from a survey study (n=93) indicate that respondents perceive high data sovereignty. Smart contracts enable providers to control the shared data; certification signals meta-platforms' responsibility to deliver secure data-sharing infrastructure.

- Draw a causal diagram that summarizes the findings of the study

(Abbas et al., 2023)

Research design

This study has some limitations. First, we used a one-sample t-test, which only compares results to the middle point of a rating scale. To improve the validity of the findings, we will continue this research by conducting a between-subject 2x2 experiment. In doing so, we can compare the effect of the presence of these control mechanisms and identify potential interaction effects to confirm the proposed H1 and H2. Second, while our study focuses on the most critical dimension of data sovereignty, we are aware of the potential significance of other dimensions (e.g., justice). To account for this, we considered the justice dimension as a control variable in the prototype development by suggesting appropriate data pricing to ensure fair revenue distributions. Finally, the technical aspects of smart contracts and certification are beyond the scope of our work.

- Discuss the limitations of the paper: Which of the hallmarks of scientific research are affected?
- Imagine one reviewer is positivist while another is constructionist. What would be their main point of disagreement?
- Suggest an alternative research strategy and explain how it would resolve the limitations

Follow-up study

To what extent do data providers perceive that the availability of control mechanisms (i.e., smart contracts and certification) enhance data sovereignty for data sharing in a meta-platform?

H1: Availability of smart contracts positively influence data sovereignty by enabling a) data ownership and b) control over data for data providers.

H2: Availability of certification positively influences data sovereignty by a) ensuring security, b) clarifying responsibility, and c) adhering to relevant compliance.

Follow-up study

Setting: Controlled experiment of 1 hour

Question: How to set up an experiment to test this? What to measure? When to measure?

Menti.com →

Disturbing effects in experiments

Disturbing factor	
History	Unpredictable factors happening during experiment (e.g. political event)
Maturation	Changes due to passing of time (e.g. patients recover over time)
Testing	First test affects outcomes second test
Instrumentation	Changing instrument or observers
Regression to the mean	Selecting extreme subjects based on measurements with errors
Selection	Groups differed from the start
Mortality	Loss of subjects during experiment (e.g. because lack of motivation
Interaction	Factors may interact…

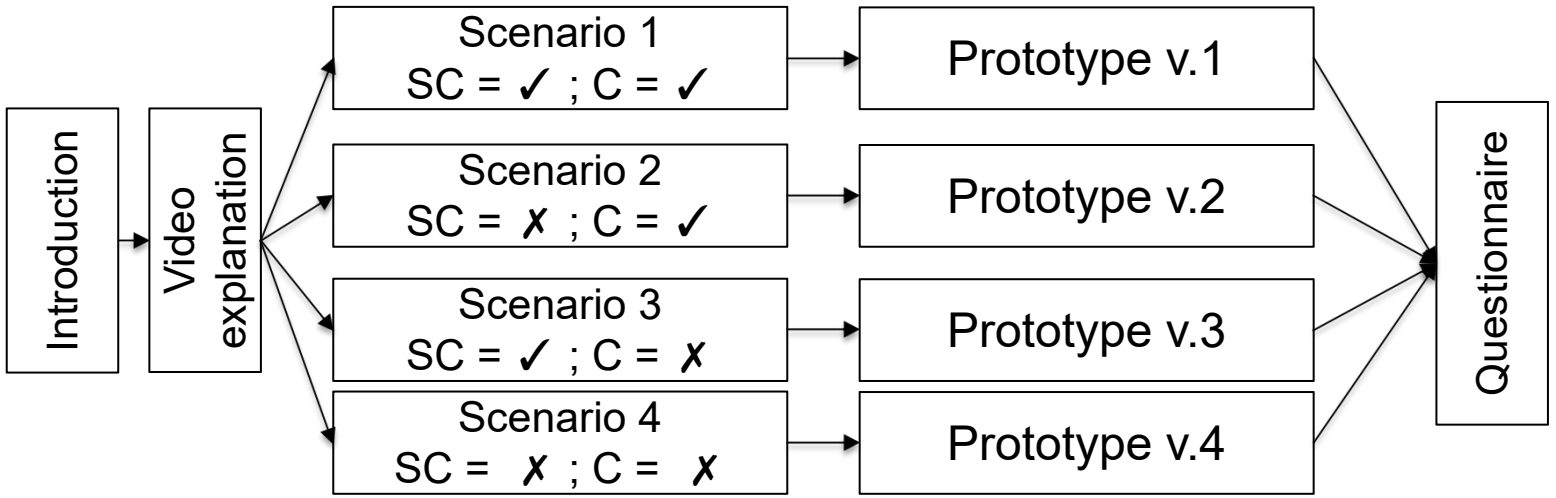
Example case

To what extent do data providers perceive that the availability of control mechanisms (i.e., smart contracts and certification) enhance data sovereignty for data sharing in a meta-platform?

H1: *Availability of smart contracts positively influence data sovereignty by enabling a) data ownership and b) control over data for data providers.*

H2: *Availability of certification positively influences data sovereignty by a) ensuring security, b) clarifying responsibility, and c) adhering to relevant compliance.*

- Experimental design:
  - Video explanation
  - Prototype exploration, depends on the scenario
  - Questionnaire (e.g., I believe the meta-platform enables sovereignty for the sensitive data that I would share.
- What would be threats to internal validity (=causal claims) in this experiment?



How to deal with disturbing factors that are external?

- E.g. experience with data sharing
- Make disturbing factors constant
  - e.g. limit to novice users
  - → limited external validity
- Systematic variation of disturbing factors
  - disturbing factor as a variable; sampling quota
  - → insight in interaction between independent variables

To do before Monday 27 Nov

- Read S&B: Ch 5,6,11
- Weekly assignment (deadline 18 Nov)

- Monday: Wrap-up module 2 & 3
  - Q&A
  - Generic feedback
  - Quizzes