What are Designed Data?
By James Wagner

What are Designed Data?

- **Designed data:**

  - Designed for an analyst or researcher

  - Data collected from persons or establishments

  - Asking **scientifically-designed** questions

  - Interviewer (in-person, video, or telephone), mail, SMS, or web

Designed Data Process

Data Collection Design	Data Collection Process	Data processing and Analysis
Type: • Qualitative • Quantitative • Mix	**Recruitment:** • Focus groups • Cognitive Interviews • Surveys • Experimental Study	**Processing:** • Coding Open Ends • Data Editing/Coding • Consistency Checks • Missing Values • Imputation • Synthetic versions of data for privacy
Method: • Cognitive Interviewing • Focus groups • Surveys • Experiment	**Sampling:** • Probability Based (ABS, RDD, Other) • Nonprobability Based • Sample Sizes • Sampling Units • Randomization	**Weighting:** • Based Sample Weight • Nonresponse Adjustments • Other Adjustments
Measurement: • Cognitive Questions • Survey items; Open & Closed ended **Data Elements** • Survey Responses • Focus Group Summaries • Cognitive Interviews • Paradata • Experimental Data	**Data Collection:** • Frequency (Cross-sectional, Panel, Repeated) • Self-Administrated • Computer Assisted • Interviewer Mediated • Observed/Measured	**Calibration:** • Population Targets • Known Benchmarks **Analysis** • Design-Based Framework • Model-Based Framework • Hybrid/Other Framework

Designed Data… or Not? (1)

- **Administrative records** are designed… but not in the sense defined here

  - Records are designed for purposes other than analysis

    - For example, unemployment insurance payments are important for administration of unemployment payments

  - May not measure the concept of interest to the analyst

    - For example, unemployment insurance payments record reported income, but not other income

Designed Data… or Not? (2)

- Machine learning often **labels** gathered data… are labels "designed"?

- Yes, but…

  - Gathered data are not designed for analysts

  - Labels may be designed for analysts, but are grafted onto data that are not

- Nevertheless, principles of design often apply to the labelling process

  - For example, human-centric labels often include "measurement error"

Designed Data: Summary

- Designed data

  - Intended for analysis

  - Most frequently collected via a survey

  - Data generation is specified by a research plan

  - Follow scientific principles in their design

  - The principles of design are also relevant for gathered data

What's Next?

- In the next lecture, you will have a chance to **create your own survey**!

UNIVERSITY OF MICHIGAN