

## 36-303: Sampling, Surveys and Society

Data Collection Methods and Ethics  
Brian Junker  
132E Baker Hall  
brian@stat.cmu.edu

26 January 2010

1

## Handouts

- Office Hours Survey
- Article packet
  - Interview with Fred Roper
  - Report on Netflix Challenge and Confidentiality
- Lecture Notes

26 January 2010

2

## Outline

- Office Hours Query
- Assignments and Readings
- Team Assignments
- Methods of Data Collection (Groves Ch 5)
  - Evolution of methods
  - Interviewer effects
  - Response Rates
  - Sampling Frame and Coverage
- Ethics (First part of Ch 11: pp 345-358)
  - Fabrication, Falsehood, Plagiarism
  - Dealing with Clients and the Public
  - IRB: Legal and Ethical Obligations to Respondents
  - Confidentiality and Statistical Disclosure

26 January 2010

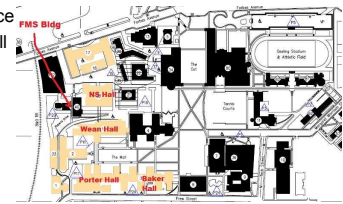
3

## Office Hours Query

- We have not scheduled Jionglin Wu's office hours:

Please fill out office hours survey; we'll choose 2 times that most people can meet.

HW due Thu's...



26 January 2010

4

## Assignments & Readings

- Things to read:
  - Done already: Groves Ch 1, 2
  - This week:
    - Groves, Ch 5
    - Groves, Ch 11 (pp 345-358)
    - Groves, Ch 4 (pp 93-102)
  - Next reading will be a handout on probability for surveys
- An individual homework will be assigned Thursday

26 January 2010

2

## Team Assignments

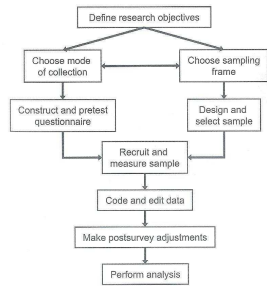
- Assignment due today – sorry for confusion!  
Turn in either of the following for your two topic proposals:
  - I.1 A, B, C (short outline from, week01); OR
  - Items a,b,c,d,e (p 4, longer outline from week02)
- Assignment due next Tues – Complete answers (for both proposed topics):
  - I.1 A, B, C (short outline from week01)
  - I.2 D, E, F, G (short outline from week 01)

26 January 2010

6

## Methods of Data Collection

- **Mode** or **method** of data collection: how to get data from each respondent?
- Must answer two questions:
  - What method is most appropriate for your research question(s)?
  - How does the method impact errors and costs?



26 January 2010

7

## Some Typical Methods of Data Collection

- 1930's-1950's:
  - Face to Face Interviews
  - Mail Surveys
- 1960's:
  - Telephone Surveys (vs face-to-face)
- 1980's-1990's:
  - Disk by mail (vs paper & pencil mail surveys)
- Today:
  - Computer assisted methods
  - Email and Web surveys

26 January 2010

8

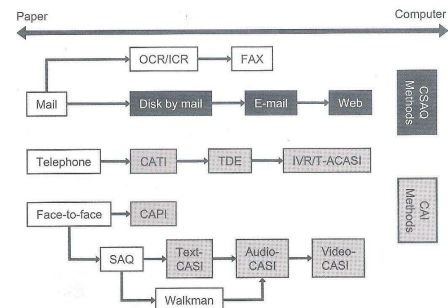
## Technology-Assisted Methods

- Computer-assisted personal interview (CAPI)
  - Interviewer reads questions aloud from laptop, enters respondent answers on laptop
- Computer-assisted telephone interview (CATI)
  - CAPI with a phone
- Computer-assisted self-interview (CASI)
  - Respondent operates the laptop
- Audio questions, keypress or voice-recog response
  - ACASI: Audio questions CASI (text/video too)
  - T-ACASI or IVR (Interactive Voice Response)
- Self-Administered Questionnaires (SAQ)
  - P&P mail surveys, email surveys
  - Computerized SAQ (CSAQ), Web surveys

26 January 2010

9

## Evolution of Survey Technology



26 January 2010

10

## Degree of Interviewer Involvement

- Can be deep or non-existent
  - Interviewer reads questions, records responses
  - Mail-in questionnaire (no interviewer at all)
- Can blend the two approaches
  - SAQ as part of an interview
  - (A)CASI for sensitive questions (NSDUH drug use questions)
  - Look at this video and then I will ask you questions about your reaction...

26 January 2010

11

## Interviewer Benefits and Costs

- **Benefits**
  - Interviewers can be effective recruiters – higher response rates
  - High interaction with respondent → more raw data per respondent → more control over measurement process
  - Answering "metaquestions" about the survey can increase respondent cooperation
  - Assist in clarifying, probing and motivating respondents (reduces item nonresponse → higher proportion of answered questions/respondent)
  - Even when no "interview" is involved, trained observers can solve data collection problems in the field
    - Gas prices survey used trained observers "at the pump", not telephone reports
    - NAEP periodically collects academic transcripts and other administrative information from schools
    - Seek personal permission from farmer to do crop measurements

26 January 2010

12

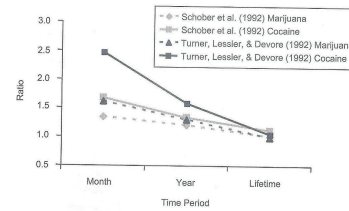
## Interviewer Benefits and Costs

- Costs
  - Some sensitive questions show large interviewer effects
    - Race/gender/personal affect of interviewer
    - Personal questions, social desirability questions
  - Privacy issues
    - Interviewer presence or proximity
    - Proximity of other respondents
    - Randomized response methods
  - Accessibility/comfort level of locations, interviewer fraud, ...
  - Requires a corps of trained, equipped, motivated interviewers
    - Interviewers require supervision, monitoring and support
    - Large portion of survey cost and survey organizational effort
  - Costs of transcribing and coding complex interviewer interactions

26 January 2010

13

## Interviewer Effects for Sensitive Questions



- Ratio of Proportion of respondents reporting illicit drug use in self-administered questionnaire vs interviewer-administered questionnaire (Groves Ch 5)

26 January 2010

14

## As We Remove the Interviewer from the Process

- Channels of Communication
  - Figures and Text
  - Audio
  - Video
- Social Presence
  - Is respondent "socially focused" on the survey, or multitasking?
- Order effects
  - Primacy: the first choice is preferred
  - Recency: the last choice is preferred
- First tries
  - The first answer may be "socially desirable" or not thought out
  - Giving respondents the ability to backtrack/review/revise answers reduces "first try" effects

26 January 2010

15

## Response Rates

- Hager et al. (2003, *Nonprofit and Voluntary Sector Quarterly*, pp 252-267):

What constitutes an acceptable return rate? Research methods textbooks differ in their guidelines, although most argue strongly for securing a high rate of return as a means of minimizing nonresponse bias. Babbie (1990) contended that a return of **50%** is adequate, although Bailey (1987) set the adequacy bar at **75%**. More recently, Schutt (1999) instructed students that "a response rate below **60 percent** is a disaster" and concluded that "it is hard to justify the representativeness of the sample **if more than a third failed to respond**" (pp. 254-255).

- The key is to be able to justify representativeness. This is much easier if the response rate is higher.
- **Nonresponse Bias**: How different are the responders from the nonresponders?
  - *Big issue for Web surveys and other self-selected samples!*

26 January 2010

16

## Response Rates: FTF, Telephone, Mail

- Generally speaking
  - Face to Face Interviews have the highest response rates
  - Telephone Interviews have the next highest response rates
    - O'Neil et al (1979)
      - Response rates ranging from 50% to 70% (seems high)
        - Pew (2008 Religious Landscape Survey) reports 24%
      - Most refusals in the first few minutes, so careful how you start!
    - Mail-in surveys have the lowest response rates
- Key issues
  - Motivation of respondents
  - Advance notification, number of contact attempts, followup of nonresponders

26 January 2010

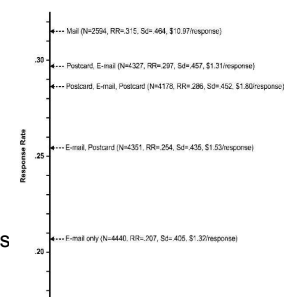
17

## Response Rates: Email vs Snail Mail

- Kaplowitz et al (2004, *Public Opinion Quarterly*, pp 94-101)

- Compared **mail-in** vs **email** surveys for Michigan State undergraduates with excellent email access

- This is a "best case" scenario!



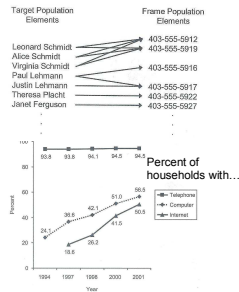
26 January 2010

18

## Sampling Frame and Coverage Issues

### Mode affects frame and coverage

- Face to face – need accessible locations
- Telephone – how does phone book affect coverage?
- Land line vs cellphone
- Email
  - Availability, self-selection
- Web Survey
  - Availability, self-selection



26 January 2010

19

## Choosing the Mode

### Face to Face interviews

- Expensive, seen most often in large scale federally funded surveys
- Can be appropriate in small surveys if you really need to control or problem-solve the measurement process

### Comparing modes (FTF, Phone, Mail, Email, Web, ...)

- Cost
- Respondent Attention
- Sensitive Questions
- Response Rate
- Nonresponse bias & Coverage bias

26 January 2010

20

## Ethics

- Survey researchers, like all scientific researchers, are held to high ethical standards
- <http://www.aapor.org> lists a Code of Ethics and acceptable behaviors for survey researchers
- Federal Department of Health and Human Services funds most human subjects research and enforces ethics through its Office of Research Integrity

26 January 2010

21

## Some Obvious Ethical Issues

- **Fabrication** – making up data or results and recording or reporting them
- **Falsification** – Manipulating equipment or materials, or miscoding/changing/omitting results so that the reported research does not reflect the raw research data.
- **Plagiarism** – theft, misappropriation, unauthorized use of intellectual work. *Does not include* well-marked, credited quotation.

26 January 2010

22

## Ethical Issues

- Fabrication, Falsification, Plagiarism are obvious issues for the Researcher
- They are also issues for Interviewer training and quality control!

Survey	Pct of Interviewers Falsifying
Current Population Survey	0.4%
National Crime Victimization Survey	0.4%
New York City Housing Survey	6.5%

(Source: Schreiner, Pennie & Newbrough, 1988, as reported in Groves Ch 11)

26 January 2010

23

## Standards for Dealing with Clients

- Undertake only research that can reasonably be carried out in the given time & budget
- Report fully the conditions, and limitations, of your study
- If you discover serious errors in methodology, disclose, and if possible, correct them
  - Roper poll for American Jewish Committee
    - "Does it seem possible... that the Nazi extermination of the Jews never happened?" 22% agreed.
    - Redid survey at own expense, reworded question, now only 1% agreed.

26 January 2010

24

## Standards for Dealing with the Public

Table 11.3. Elements of Minimal Disclosure (AAPOR Code)

1. Who sponsored the survey, and who conducted it
2. The exact wording of questions asked, including any preceding instruction or explanation that might reasonably be expected to affect the response
3. A definition of the population under study and a description of the sampling frame
4. A description of the sample selection procedure
5. Size of sample and, if applicable, completion rates and information on eligibility criteria and screening procedures
6. The precision of the findings, including, if appropriate, estimates of sampling error and a description of any weighting or estimating procedure used
7. Which results, if any, are based on parts of the sample rather than the entire sample
8. Method, location, and dates of data collection

Source: <http://www.aapor.org>

26 January 2010

25

## Standards for Dealing with Respondents

### ■ Legal Obligations

- Institutional Review Board (IRB)
  - Ensure that the possible **benefits** of the research are **balanced against risks** to research subjects.
  - Ensure that research subjects have opportunity to provide **informed consent** to be studied.
- Risks are obvious in medical studies
  - New treatment/placebo for AIDS, cancer, etc.
  - Tuskegee Study: placebo for syphilis w/o informed consent
- Risks less obvious but still present in social research
  - Milgram "obedience" experiments – subjects were told to "shock" fake patients who acted out the pain.
    - The psychological effects on the subjects persisted long after the experiment.

26 January 2010

26

## Standards for Dealing with Respondents

### ■ Ethical Obligations

- **Beneficence**: Protecting Respondents from Harm
- **Justice**: Balance between those who bear the burdens of research vs. those who benefit from the research.
- **Respect for persons**: The human right to self-determination (life, liberty, pursuit of happiness, other significant decisions, ...)
- **Informed consent**: Each respondent should be fully informed about the nature of the study, and have an unencumbered opportunity to consent—or refuse—to be studied.
- These issues may need to be revisited throughout the life of a survey or other research study

26 January 2010

27

## Standards for Dealing with Respondents

Table 11.4. Essential Elements of Informed Consent

1. A statement that the study involves research, and explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures, and identification of any procedures that are experimental
2. A description of any foreseeable risks or discomfort
3. A description of any benefits to the subject or others that may reasonably be expected
4. A disclosure of appropriate alternative procedures or courses of treatment
5. A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained
6. For research involving more than minimal risk, an explanation of whether and what kind of compensation or treatment are available if injury occurs
7. An explanation of whom to contact with further questions about the research, subjects' rights, and research-related injury
8. A statement that participation is voluntary and the subject may discontinue participation at any time without penalty or loss of benefits.

26 January 2010

28

## Confidentiality and Statistical Disclosure

- Most research situations, surveys included, include a commitment to maintain confidentiality of results
  - This is part of respect for persons
  - Confidentiality can also help with sensitive questions
- Threats to Confidentiality
  - Carelessness & Negligence
  - Legal Demands for Identified Data
    - Freedom of Information Act; exceptions for sensitive research
    - 2002 "Confidential Information and Statistical Efficiency Act"
    - Homeland Security, USA PATRIOT Act
  - Statistical Disclosure
    - Using matching between data bases together with statistical modeling to de-anonymize "anonymous" data bases

26 January 2010

29

## Statistical Disclosure: Netflix Database

- 2007: Netflix released anonymized data base of movie rentals as public challenge for better *recommendation* or *collaborative filtering* systems.
- Researchers immediately found ways to "hack" the database to reveal identities (and rental habits) of individual Netflix users
  - One method: Cross-matching with signed interviews on IMDb
  - More generally: after you eliminate approximately the top 100 most-watched movies, our viewing habits are highly individual!
- Similar with other data releases (AOL, US Census, ...)
- See Crypto-Gram handout today.

26 January 2010

30

## IRB Approval in 36-303

- Historically IRB has been more focused on medical research than social research
- In recent years, liability concerns (risk/benefit, confidentiality, etc.) have spread IRB review to most social and survey style research
  - Studies conducted for research must undergo IRB review
  - Studies done for commercial clients, done in the process of consulting, or done for class credit, often do not require IRB approval
- In this class I will require you to go through the IRB process for your projects; I will act on behalf of IRB

26 January 2010

31

## Review

- Please Turn In First Team Assignment today:  
*Two Possible Projects*
  - I will email you feedback this week
  - Next Tuesday, Team Assignment 1.2 is due:  
*Complete items A,B,C,D,E,F,G from shorter project outline (from week01)*
- Methods of Data Collection (Groves Ch 5, & Ch 4 pp 93-102)
- Ethics (First part of Ch 11: pp 345-358)

26 January 2010

32