

36-303: Sampling, Surveys and Society

Midterm Review
Brian W. Junker
132E Baker Hall
brian@stat.cmu.edu

28 Feb 2012

1

Handouts, Etc.

- Handouts:
 - These Lecture Notes
 - Formula Sheet for Exam
- HW solutions are finally up on website
 - Sorry for the delays!
- Turn in I.4 on Fri, not Thu
 - Feedback on I.3 went out Monday
 - Feedback on I.4 ASAP

28 Feb 2012

2

Outline

- Review For Midterm Exam
 - Thurs Mar 1, 2012
 - Closed book, closed notes
 - Formula sheet provided; calculator encouraged!
- Team Project Status
 - Where everyone is now
 - Review requirements 'till Spring Break

28 Feb 2012

3

Review: Major Components of a Survey

- Research Objectives
 - Research Questions
 - Constructs
- **Target Population**
- Mode of Data Collection
 - FTF, Phone, Mail, Email/Web, ...
- **Sampling Frame**
- **Random Sample** "Not random? Where's the Bias?"
- Measurement (e.g. Survey Questions)
- **Nonresponse**
 - Response Rate
 - (Self-)Selection bias
 - Following up nonrespondents
- Coding, Editing, Analyzing, Reporting

28 Feb 2012

4

Measurement Quality

- Validity (low bias) – Are the answers giving us information about our research question?
- Reliability (low variability) – Do two people with the same status (opinion, income, etc.) give the same answer?
- Question Design – see below
- Processing & Coding Errors

Representation Quality

- Coverage Error How well does the Sampling Frame cover the Target Population?
- Sampling Error
 - Bias – Reduce with random sample, high response rate r/n
 - Variability – Reduce with larger sampling fraction n/N
- Nonresponse Error
 - Response Rate
 - (Self-)Selection bias
 - Following up nonrespondents
- Adjustment Error (weights)

Methods of Data Collection

- **For human surveys:**
 - *Affordable? Believable Results? Coverage? Response Rates?*
 - Face to face (FTF)
 - Telephone
 - Mail
 - Email/Web
- **Response Rates r/n (number of responses / number asked):**
 - *Suggest at least 50% (more like 70%) response rate, to make “representativeness” argument easy*
 - FTF 70% or greater
 - Telephone 20-70%
 - Mail 30%
 - Email/Web 20-30%

Questions and Answers

1. Define what you want to measure
 - Make sure research question is well focused
 2. Design the questions around that
 3. Pretest every revision
 - Does respondent understand question?
 - Can respondent recall relevant information?
 - Can respondent combine, edit relevant info?
 - Does respondent accurately report answer?
- Experts, Cog Interviews & Focus Gps, Field Tests

Questions and Answers – Some Pointers

- Simple Language
- Common Concepts
- Manageable Tasks (shared definitions, recall, hypotheticals)
- Widespread Information
- Specific vs General Questions
- Question Order
- Open vs Closed Questions
- Likert (agree/disagree) vs Forced-Choice
- Question Wording; Loading
- Pleasing the Interviewer (socially desirable answers)
- Pretest, pretest, pretest

28 Feb 2012

9

Ethics

- Fabrication, Falsification, Plagiarism
- Responsibilities to clients – manageable projects; report & correct errors
- Reporting to the public
 - Who sponsored it, who carried it out
 - The exact wording of questions
 - Target population, sampling frame, sampling method, response rates, nonresponse followup
 - Sample size, precision (SE) of estimates, which results are based on only part of sample
 - Method, location, dates of data collection

28 Feb 2012

10

Ethics (2)

- Legal obligations to respondents – IRB
 - Risk/Benefit tradeoffs
 - Informed Consent
- Ethical obligations to respondents
 - Beneficence
 - Justice
 - Respect for Persons
 - Informed Consent
- Informed Consent
 - Purpose
 - Risks/Benefits
 - Confidentiality
 - Compensation for harm
 - Contact info for any questions
 - Participation is voluntary
- Confidentiality
 - Respect for persons; Sensitive information
 - Threats to confidentiality
 - Carelessness
 - Open gov't laws
 - Statistical disclosure

28 Feb 2012

11

Statistics for Surveys

- Review:
 - Discrete RV's
 - Expected Value, Mean, Variance
 - Covariance and Independence
 - Linear Combinations
 - SRS with replacement:
 - CLT, Confidence Interval, Sample Size ...
 - Conditioning

28 Feb 2012

12

Statistics for Surveys (2)

■ Urn Models

- SRS with replacement (*elementary statistics*: the urn never changes)
- SRS w/o replacement (*survey sampling*: the urn changes after every draw)
 - SE's are smaller than for SRS with replacement
 - CLT doesn't work for all "large enough" sample sizes
 - $n > 20$ or so seems to be important, as usual
 - $n/N > 0.8$ or 0.9 and things start getting bad again

Statistics for Surveys (3)

■ Finite Population Correction (FPC)

- Data y_i are fixed;
- Sampling indicators Z_i are random
- Leads to FPC:

$$SE_{(SRS \text{ w/o repl})} = \sqrt{1 - f} \times SE_{(SRS \text{ with repl})}$$

where

$$f = n/N$$

(what would we do with this?)

Statistics for Surveys (4)

■ Sample size calculation, SRS with replacement

$$n \geq n_0, \text{ where } n_0 = \frac{z_{\alpha/2}^2 (SD)^2}{(ME)^2}$$

■ Sample size calculation, SRS without replacement

$$n \geq \frac{N n_0}{N + n_0}, \text{ where } n_0 = \frac{z_{\alpha/2}^2 (SD)^2}{(ME)^2}$$

Project Topics Chosen

- A political survey of the CMU community
- Music/movie internet piracy at CMU
- How to improve our on-campus parking system?
- Political Attitudes and Academic Major at CMU
- Analysis of the Off-Campus Housing Search for CMU Students
- Parking Meters at CMU
- Spatial and Analytic Study of Student Housing at CMU

Team Project Status: Going Forward

- All projects on-campus
 - 2 surveys of “objects” (parking meters, student records)
 - 2 surveys involve faculty/staff in addition to students
 - 3 surveys are standard student surveys
- Topics range from moderately to very interesting; all are doable; most are “actionable”
- Going forward: Very good methods and execution.
 - Well-focused research question(s)
 - **Well-defined target population, sampling frame**
 - **Clear plan for random sampling or equivalent**
 - **Clear plan for nonresponse followup**
 - Well-designed and pre-tested survey questions
 - Clear statistical analysis
 - Clear, thoughtful scientific writeup

28 Feb 2012

17

Team Assignments – Through Spring Break

- **UPDATED PROJECT SCHEDULE** at <http://www.stat.cmu.edu/~brian/303>

28 Feb 2012

18

Review

- Review For Midterm Exam
 - Thurs Mar 1, 2012
 - Closed book, closed notes
 - Formula sheet provided; calculator encouraged!
- Team Project Status
 - Where everyone is now
 - Review/Update of Requirements ‘till Spring Break

28 Feb 2012

19