

36-303: Sampling, Surveys and Society

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

16 February 2010

1

Handouts

- These Lecture Notes
- Formula Sheet for Exam
- HW02 solutions (on course website)
- Midsemester Course Evaluation (at end of class)

16 February 2010

2

Outline

- Review For Midterm Exam
 - Thurs Feb 18, 2008
 - Closed book, closed notes
 - Formula sheet provided; calculator encouraged!
- Team Project Status
 - Where everyone is now
 - Review requirements 'till Spring Break

16 February 2010

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? Not representative!"
- Measurement (e.g. Survey Questions)
- **Nonresponse**
 - Response Rate
 - (Self-)Selection bias
 - Following up nonrespondents
- Coding, Editing, Analyzing, Reporting

16 February 2010

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**

16 February 2010

5

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)

16 February 2010

6

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%

16 February 2010

7

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

16 February 2010

8

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

16 February 2010

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

16 February 2010

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

16 February 2010

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

16 February 2010

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

16 February 2010

13

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?)

16 February 2010

14

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

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

16 February 2010

15

Project Topics Chosen (1)

- Perceptions of academic integrity & old course material archives
- Satisfaction with CMU meal plan
- Use and satisfaction with UC athletic facilities
- Career plans of graduating seniors, and influences on seniors here at CMU
- The attributes of innovation needed to meet career challenges in the current economy

16 February 2010

16

Project Topics Chosen (2)

- Knowledge and use of safe walk and escort services on campus
- Perceptions of use, fairness, appropriateness of mandatory student fee at CMU
- Consumer survey: where do students order pizza and why?
- Bike usage survey

16 February 2010

17

Team Project Status: Going Forward

- Everyone chose on-campus projects; that's fine!
- All are moderately interesting/creative; 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**
 - **Clear plan for nonresponse followup**
 - Well-designed and pre-tested survey questions
 - Clear statistical analysis
 - Clear, thoughtful scientific writeup

16 February 2010

18

Team Assignments – So Far

- Revised Project Schedule at <http://www.stat.cmu.edu/~brian/303>
- I.0 Teams Formed (*Tue Jan 19*)
- I.1 Propose Two Topics (*Tue Jan 26*)
- I.2 Revise proposals (*Thu Feb 4*)
- I.3 Choose Topic (*Thu Feb 11*)
- I.4 Target Pop, Sampling Frame, Sampling Plan, Nonresponse Plan (*Tue Feb 16*)
- Team Working Agreements (*Tue Feb 23*)

16 February 2010

19

Team Assignments – “Till Spr. Break

- Team Working Agreements (*Tue Feb 23*)
- II.5a. Sampling Scheme & Question Design (*Thu Feb 25*)
 - Items K, L, M on “designing a sample survey” handout
- II.5b. Pretest & Revise Questions (*Tue Mar 2*)
 - Items N, O on “designing a sample survey” handout
- II.6. Final IRB & Project Plan (*Thu Mar 4*);
 - Final, full project proposal (items A-M on the “designing a sample survey” handout).
 - IRB forms filled out completely.
 - A draft informed consent statement for your survey. Turn in all three to me, not CMU IRB.
- EACH ASSIGNMENT ABOVE TURN IN BY EMAIL

16 February 2010

20

Review

- Review For Midterm Exam
 - Thurs Feb 18, 2008
 - Closed book, closed notes
 - Formula sheet provided; calculator encouraged!
- Team Project Status
 - Where everyone is now
 - Review of Requirements ‘till Spring Break

16 February 2010

21