Introduction to Survey Design
The Total Survey Error Approach

Dino P. Christenson
Department of Political Science
Hariri Institute for Computational Science & Engineering

What You Need
1. Questionnaire - what to ask
2. Sample - who gets it
3. Implementation - how to collect it
4. Statistics - how to understand it

Survey
- What do you want to know?
  - Do we already know it?
    - Or, have we already asked it?
      - Second hand data can be nice!
  - If not, then, ask it:
    - What's your question?
- Regardless, requires in-depth knowledge of the lit

Implementation
- Survey houses (academic & private)
  - E.g., Gallup, YouGov, ISR at UofM, SSI
- Academic surveys
  - E.g., CCES, ANES, GSS, TOPS
- DIY
  - Paper and pencil
    - Door-to-door or mail or telephone (modes)
  - Online software (e.g., Qualtrics, SurveyGizmo)

Sample
- Representative of the population!
- Which population?
- Convenience samples
- Friends and family (not recommended)
- Students
- Opt-in online

Statistics
- For some good designs...
  - Just need the basics
- PO 502 or some introduction to applied social science statistics
Total Survey Error Approach

- Systematic way to consider tradeoffs in conducting a survey
- Where to expend resources
  - Time
  - Money
  - Ethics
- In order to minimize survey error

Good’ish News

- Lots of potential errors
- But no perfect study!
  - Think about likely errors for your study
- TSE helps to minimize specific errors by focusing on tradeoffs

Types of Survey Error

- Sampling Error
- Coverage Error
- Unit Level Nonresponse Error
- Item Level Nonresponse Error
- Respondent Measurement Error
- Interviewer Measurement Error
- Post-Survey Error
- Mode Effects
- Equivalence Error

Respondent Measurement Error

- Response accuracy problem
- Respondent lacks motivation to answer correctly
- Unclear question wording
- Temporal issues
- Double-barreled
- Overly sophisticated
- Biased question wording

Minimize Respondent Measurement Error

- Use vetted survey questions
- Conduct pre-tests of the questionnaire
- Use “think-aloud protocols”
- Random half-samples of question wording
- Simplify each “stage of survey response”
  1. Comprehension
  2. Retrieval
  3. Judgment
  4. Reporting
The High & Low Roads

- Minimize “satisficing”
- Responding in order to move on rather than responding after carefully thinking through the question
- Potential problems
  - E.g., very short answers to open ended questions
  - E.g., long batteries with same response options

Solutions
- Time survey responses
- Encourage respondent engagement
- Mix up the direction of response options
- Break up questions

Minimize Interviewer Measurement Error

- Interviewer objective:
  - Facilitate interview
  - Obtain accurate answers
  - But they can also introduce error
  - Random
    - E.g., wrongly records an answer
  - Systematic
    - E.g., always mispronounces a word

Solutions
- Time survey responses
- Encourage respondent engagement
- Mix up the direction of response options
- Break up questions

Modes Matter

- Survey modes
  - Face-to-face
  - Mail
  - Telephone
  - Internet
- Interviewer error vanishes
  - Mail
  - Internet
- Costs shrink too!
- Tradeoff
  - Response accuracy may decline, especially on open ended questions

Item Nonresponse

- Nonresponse on particular survey questions
  - E.g., refusals, skipped questions, inadequate response options
  - Results biased when those who answer are different than those who don’t
  - E.g., study of income on vote choice, but if higher income vote more conservatively but don’t report income than relationship understated

Minimize Item Nonresponse

- Require answering question
  - Tradeoff: Respondent drops out
  - Skilled interviewer can encourage answers
  - Tradeoff: cost in training interviewer
  - Multiple imputation
  - Create values for missing values via predicted values from regressions with random error term
  - Requires a lot of data and missing at random
## Unit Nonresponse

- Respondents in sample do not take survey
- Cannot be contacted
- Refuse to take it
- Can bias the sample if those who participate are systematically different than those who do not
- Refusal rate increasing
- Some conservative pundits discourage participation
  - Could result in underestimate of Rep vote

## Minimize Unit Nonresponse

- Tailor interview request as valuable to the respondent
- Pay respondents to participate
  - E.g., $1-$5 can yield 2%-12% increase with diminishing returns
  - Unless very large
    - E.g., ANES 2012 $25-$100 yields 38% pre- and 94% post

## Coverage Error

- Discrepancy between list of population and actual population
  - E.g., sampling from a telephone book which misses all those with unlisted telephone numbers
  - In 2012 Republican pollsters overstated Romney’s chances because cell phone numbers were not sampled

## Minimize Coverage Error

- Address-based sampling uses addresses instead of telephone numbers
- Internet surveys initially high coverage error but decreasing steadily with greater Internet access
- Use multiple sampling frames - and weight those with greater probability of falling into sample

## Sampling Error

- Any time we interview only a sample of the population
- By chance our estimates will be off from the population
- With probability sampling we therefore provide a margin of error
  - Conventional confidence interval is 95%
  - Estimate is within 2.5% of true population estimate

## Minimize Sampling Error

- Increase sample size
  - Tradeoff: can be costly
    - Less so for internet surveys
      - But more is not always better
    - 1936 Literary Digest poll of 2m
**Minimize Sampling Error**

- **Stratified sample**
- Take proportions from subcategories, e.g., regions

- **Cluster sample**
- Sample within known clusters, e.g., city blocks

---

**Sampling & the Internet**

- Internet surveys
- Difficult to conduct probability sampling
- Email list of the population of interest?
  - Option: probability sample via telephone or mail requesting they take an online survey

- Most use opt-in polls
- Sampling errors cannot be validly computed
- Increased risk of selection bias
- Similar to coverage error and nonresponse biases
- Solution: weight respondents
  - E.g., poststratification adjustment, sample matching, propensity score weights

---

**Convenience Samples**

- Crowd source your sample!
- Not always appropriate
- See previous slide
- Mechanical Turk okay for
  - Experiments
  - When other approaches difficult
  - Tight panel waves around developing events

---

**Survey Mode Effects**

- How the survey is conducted
  - Face-to-face & telephone
    - Interviewer effects, esp on sensitive questions
    - Social desirability bias, appear likable to interviewer
  - Solution: phrasing of questions to legitimate all responses
  - Solution: use interviewer-less modes

---

**Symbolic Racism Scale**

1. It’s really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites

2. Irish, Italian, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same...

Henry, P.J., & Sears, D.O. 2002
Survey Mode Costs

- **Money**
  1. Face-to-face
  2. Telephone
  3. Mail
  4. Internet

- **Time**
  1. Face-to-face
  2. Mail (awaiting response)
  3. Telephone
  4. Internet

- **Response**
  1. Internet
  2. Mail
  3. Telephone
  4. Face-to-face

Junk Mail

Spam

Interaction Ritual

Post-Survey Error

- Error during the processing and analysis of survey data
  - E.g., coding open-ended questions
  - Solution: create comprehensive coding schemes
  - Solution: calculate inter-coder reliability

Equivalence Error

- Lack of equivalence of surveys measuring same concepts
  - House effects, survey organization regularly attaining more of one response than another
  - Different countries, interpretations differ by culture
  - Different times, real world conditions change
    - E.g., “liberal” and “conservative” different today than in the past

Solution: tread carefully when comparing surveys across:
  - time
  - countries
  - survey organizations

Types of Survey Error

- Sampling Error
- Coverage Error
- Unit Level Nonresponse Error
- State Level Nonresponse Error
- Respondent Measurement Error
- Interviewer Measurement Error
- Post-Survey Error
- Mode Effects
  - Telephone
  - Internet

Weisberg 2005

Resources

Books


Articles

Resources

POQ Special Edition


Thank You

- I welcome your questions
- Also via email: dinopc@bu.edu