

Surveys

Alan Zaslavsky

Dept of Health Care Policy

Harvard Medical School

Three designs for research

- Experiments
 - Treatment is applied under control of experimenter
- Surveys
 - Emphasis on *representation* of a population
- Observational studies (quasi-experiments)
 - Analyze survey/administrative data like an experiment

Conducting a survey

- Instrument design
- Design of survey administration
- Sample design
- Data collection
- Analysis
 - Estimates of parameters
 - Estimates of standard errors

Instrument design

- Item and instrument construction have many pitfalls
 - Double-barreled
 - Unclear definitions
 - Open- vs. closed-ended responses
 - Effects of response scales
 - Reference periods and recall
 - Context effects

Instrument design

- Testing/validation is crucial
 - Focus groups for development
 - Cognitive interviews
 - Behavioral coding
 - Field tests
 - Validation against “objective” controls

Instrument design

- Psychometric validation
 - Factor analysis, internal consistency (alpha)
 - Item response theory approaches

Survey administration methods

- Considerations
 - Expense
 - Response rates
 - Accuracy of administration
 - Control
 - Confidentiality, sensitivity

Survey administration methods

- Modes
 - Mail: cheap, low response, private
 - Telephone: more expensive, better response, CATI for control of complex instrument
 - Personal interview: “tough cases”, non-telephone, complex information/observation, CAPI
 - Mixed mode

“Representative sample”

- Historical (non-random) views
 - “Most typical” representative of population (family with 2.3 children ...)
 - Purposively selected units designed to represent parts of population
- Random (probability) sampling
 - Choice of units is driven by random process
 - Known probability distribution of possible samples
 - Every unit has positive probability of selection

Populations and samples -- Getting from here to there!

- Inferential population
- Target population
- Frame population
- Survey population
- Sample
- Respondents
- Completed items

Frames

- Explicit
 - List
 - Area
- Implicit
 - RDD (Random Digit Dialing)

Sample designs

- Must have well-defined probability rules
- Tradeoffs of cost versus precision
- Tradeoffs of precision for different objectives
- Accommodate/exploit structure of the population

Common designs

- Simple random sampling
- Stratified simple random sampling
- Systematic sampling
- Cluster sampling
- Multistage sampling

Analysis of surveys

- Basic principles and objectives differ from those of nonsurvey statistics
 - Finite population inference
 - Nonparametric approach
 - Sampling distribution
 - Complex designs

Estimation under SRS

- $\text{Var}(\text{mean}) = [(1-f)/n] S^2$
- n = sample size
- N = population size
- S^2 = population variance
- f = sampling fraction = n/N
- $(1-f)$ = finite population correction (not relevant for most analytical studies)

Corrections for complex designs

- Point estimates

- Ordinary estimates may not be accurate
- Weighted analysis
 - Weight = $1 / (\text{Probability of selection})$
 - Changes definitions of point estimators
 - E.g. $\sum w_i y_i / \sum w_i$ instead of $\sum y_i / n$
- Analogous analyses for more complex estimators

Corrections for complex designs

- Variances

- Design effect = $D_{eff} = \frac{\text{Var}(\text{complex design})}{\text{Var}(\text{SRS})}$
with equivalent sample size
 - Measure of efficiency/inefficiency of design
- Misspecification effect = $M_{eff} = \frac{\text{Var}(\text{complex design})}{\text{Var.Est.}(\text{SRS})}$
for the sample at hand
 - Measure of how wrong the SRS analysis would be (for variances)

Analysis of complex surveys

- Special analytical methods required
- Special software
 - SUDAAN, Stata, WesVar, PCCARP, EpiInfo, VPLEX, (SAS 6.13?)
- Attention to analytical objectives
- More difficult with combination of complex sample design, complex estimand
- Alternative approaches -- hierarchical models