Design and Analysis of Sample Surveys

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

Class 3b: Survey nonresponse

Adjusting for nonresponse

- Item nonresponse
  - Complete-case analysis
  - Available-case analysis
  - Deterministic imputation
  - Multiple (random) imputations
  - Modeling (writing down the likelihood of the observed data)
- Unit nonresponse
  - Poststratification
  - Weighting
  - Multilevel regression and poststratification (MRP)

Understanding missingness and corrections in any particular example

- Studying the missingness process:
  - Compare missing to observed items
  - Compare missing to observed people
  - Logistic regression to model Pr (missing)
- Understanding and checking imputations
  - Inspecting the distributions of imputed variables
  - Building trust via cross-validation
  - Canary variables

Summary of adjustments for *item* nonresponse

- Existing missing-data imputation packages in R and Stata
- Can sort-of check using cross-validation or simply by inspecting the distribution of imputed values
- Carry along 10 completed datasets, analyze each separately, incorporate between-imputation variance only at the final step of the analysis
- Complete-case and available-case analysis can be convenient, sometimes using partial imputation to fill in the holes

Summary of adjustments for *unit* nonresponse

- Correcting for known differences between sample and population
- Weighted estimates of population averages and totals
- Regression controlling for all variables that predict inclusion in the sample ("inclusion" = sampling & response)
- *Multilevel* regression to adjust for a large number of predictors
- *Poststratification* to sum up to larger groups in the population
- Many practical and statistical challenges

Nonresponse in different survey modes

- Consider the sampling frame for each mode:
  - Mail
  - Face-to-face
  - Telephone
  - Internet