

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2012, The Johns Hopkins University and Freya Sonenstein. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

Issues in Survey Research Design

Coverage, Sampling Error and
Non-response

Steps in Sampling

Population of Inference

Target Population (finite group of units to which inference will be made)

Sampling Frame (list/procedure to identify all elements of target population)

Sample (units selected for study with known probability)

Respondents



Error Associated with Sample Strategy

Population of Inference

Target Population

Coverage Error

Sampling Frame

Sampling Error

Sample

Non Response Error

Respondents

Probability Sampling: Sampling Error

Every element in the population has a known non-zero probability of being selected

Simple random sample SRS—assigns equal probability of selection to each frame element

Sampling Error-

- Unavoidable variance around point estimates from particular sample

Sampling Variance-reduced with big samples

Modifications of Simple Random Sample

- Systematic sample
- Stratified sample
- Multistage sample
- Cluster sample

Systematic Sampling

- Sample of size n from population of size N in which the sample units are obtained by a selection interval applied to list
- Random start with every k th selected in both directions
- Variance of estimate is biased and not straightforward

Stratified Sample

- Strata are mutually exclusive groups of elements in a frame (15-19 & 20-24 yr olds)
- Separate samples drawn from each group
 - Sampling can be done to accomplish a self weighting sample (based on proportion of elements in population)
 - Sampling can be done to adjust size of subsamples but requires adjustment for population estimate (based on desired subsample size-but probabilities of selection still known)
- Reduces variance/increases precision

Cluster Sampling

- A sample in which the sampling units are clusters of samples
- Equal sized or unequal sized clusters
- Often carried out for convenience rather than more precision
- Household surveys usually multi-stage cluster designs

Geographic Area Household Surveys

- Involves
 - Selection of Geographic Units (eg Counties and smaller units within)
 - Enumeration of households within smallest units and selection of households
 - Clustering of households
 - Deduction from household to individual

Design Effect (Deff)

- The ratio of the actual variance of a statistic obtained in a modified sampling design (eg clustered sample) to that for the estimator from simple random sample (srs) *with* the same number of n elements.
- Aiming for close to 1 (srs design effect)

Implications of the Design Effect

For stratified sample $Deff$ often < 1 (more precision than srs)

For cluster samples $Deff > 1$ (less precision than srs)

$Deff$ depends on the intraclass correlation coefficient within clusters

Complete homogeneity within clusters $Deff > 1$

– Extreme heterogeneity $Deff < 1$

– Random distribution within clusters $Deff = 1$

Implications for power analyses to determine sample size:
needs to adjust for $Deff$

Examples

- Small Community Based Study
- NYC Hanes
- National Longitudinal Study of Adolescent Health
- National Health Interview Survey + SSA supplement
- Vaccine Coverage study in Ethiopia
- Sexual Behavior in the U.S.
- Head Start Evaluation

Error Associated with Sample Strategy

Population of Inference

Target Population

Coverage Error

Sampling Frame

Sampling Error

Sample

Non Response Error

Respondents

Population of Inference

Target Population



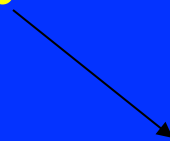
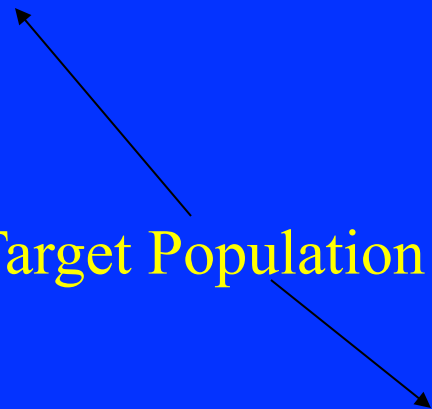
Sampling Frame



Sample



Respondents



Sampling Error

- How do we control?
 - Sample size
 - Stratification
 - Clustering

Non-Response Error

- Theoretically, when do we especially worry about non-response?
 - Ignorable
 - Non-ignorable—related to key statistic of the survey
- Practically— since difficult to know effect of non-response— researchers try to increase response rates

Components of Non Response

- Non-contact-inability to contact sampled person
- Non cooperation – refusal to be a study participant or denied participation by another
- Unable to respond-different language, blind, deaf, disabled

Calculating Response Rate

- Divide Number of Completed Interviews by
 - (Number of Refused and Broken Off Interviews + Number of Non-contacts—adjusted or not adjusted for estimate of eligibles)
- Those unable to respond usually removed from denominator unless related to key survey measures
- Calculating response rate in follow-up survey (Singer article)

Tools for Reducing Unit Nonresponse Rates



Adapted by CTLT from Groves, R.M., et al. (2009) *Survey Methodology*. Wiley-Interscience.

Incentives?

- Readings:
 - Vangeest: Monetary Incentives
 - Singer: Incentives and Expectations

Example?

- Sample Frame-Coverage?
- Sampling plan-Sampling error?
- Response -Non- response error?