Sampling Foundations

Brian Morrow, MA.

Applied Sciences Branch

Division of Reproductive Health

Centers for Disease Control and Prevention

Atlanta, GA 30341

Basic Sampling Concepts

- Advantages of Sampling
- Sampling and Non-Sampling Errors
- Factors Affecting Sampling Error Size

Advantages of Sampling

- Time
- Cost
- Resources
- Knowledge of Sampling Error

Sampling and Non-Sampling Errors

- Non-sampling errors:
- - Non-response bias
- - Non-coverage bias
- - Recall bias
- - Mode bias

Factors Affecting Sampling Error Size

- Sample Design
- Population Variability
- Sample Size

Probability Sampling Methods

- Simple Random Sampling
- Systematic Sampling
- Stratified Sampling
- Cluster Sampling
- Complex Sample Design

Simple Random Sampling – Definition

- Construct a list of every element in the population
- Use a random process to select n units from the population
- Every possible sample of size n from the population has an equal chance of being selected

Simple Random Sampling -- Advantages

- Simple
- Provides a baseline to which other methods can be compared
- Sample can be analyzed with nearly any data analysis software

Simple Random Sampling -- Disadvantages

- Must have list of entire population
- Sampled units may be far away from each other
- Certain subgroups may be oversampled or undersampled by chance

Systematic Sampling -- Definition

- Sort population by desired characteristics
- Choose a random start
- Sample every nth unit

Systematic Sampling -- Advantages

- Sometimes possible where a simple random sample cannot be selected
- Sample more likely to be uniformly spread over the population
- May be analyzed using same methods as those for simple random sample

Systematic Sampling -- Disadvantages

- Must have list of entire population
- Sampled units may be far away from each other
- Biased results may occur if sampling interval is related to length of a cyclical sampling frame

Stratified Sampling -- Definition

- Divide population into mutually exclusive segments (strata)
- Sample separately from every strata
- - Equal allocation
- - Proportional allocation
- Best results when strata related to outcome of interest

Stratified Sampling -- Advantages

- Increased precision compared to SRS if strata are related to outcome of interest
- Control sample size of each strata
- Can be easier to select compared to SRS for administrative/logistical reasons

Stratified Sampling -- Disadvantages

- Must have list of entire population
- Sampled units may be far away from each other
- Equal allocation may lead to reduced precision
- Proportional allocation may prevent adequate analysis of strata of interest

Cluster Sampling -- Definition

- Divide population into mutually exclusive segments (clusters)
- Choose a sample of clusters
- Choose sub-samples within each sampled cluster
- Best results when clusters are not related to outcome of interest

Cluster Sampling -- Advantages

- Detailed sampling frames need only be constructed for clusters selected at next-to-last stage
- Sampled units closer together

Cluster Sampling -- Disadvantages

- Less precision than other sampling techniques
- Specialized software required to properly compute standard errors, variances, and confidence intervals

Complex Sample Design

- Any sample taken using stratified and/or cluster sampling
- Requires special analysis software to properly compute standard errors and confidence intervals
- SUDAAN, Stata, SAS Complex Survey procedures, SPSS Complex Sample modules

Design Effect

- Ratio of variance using cluster/stratified sample to the variance with simple random sampling of the same sample size
- Design effect depends upon type of sample design and variability of population

Examples of CDC Surveys Using Complex Sample Designs

- National Hospital Discharge Survey (NHDS)
- Youth Risk Behavior Surveillance System (YRBSS)
- Pregnancy Risk Assessment Monitoring System (PRAMS)

National Hospital Discharge Survey (NHDS)

- Three types of sampling conducted within the program
- - Large hospitals selected with certainty; discharges selected at first stage
- - Some PSUs selected with certainty; discharges selected at second stage with hospitals as first stage

CDC
Centers for Disease Control and Prevention

National Hospital Discharge Survey (NHDS)

- - PSUs sampled at first stage; hospitals selected at second stage; discharges selected at third stage

Youth Risk Behavior Surveillance System (YRBSS)

- Three-stage, cluster sample design
- PSUs defined by large counties or groups of small counties
- PSUs stratified by MSA and percentage of Black/Hispanic students
- Stages: PSUs; schools; classes

Pregnancy Risk Assessment Monitoring System (PRAMS)

- Sampling frame is electronic birth certificate file, mode of collection is mail and phone
- No need for clustering
- Population of mothers stratified by characteristics of interest (birth weight, race/ethnicity, region)

Size of Design Effect

- 2 Populations of one million people
- 1000 clusters of 1000 people
- Overall population behavior = 50%
- Pop. A: all clusters = 50%
- Pop. B: 500 clusters 90%; 500 clusters 10%
- Sample size of 400

Size of Design Effect

- Simple Random Sample of each of 2 populations: SE = 2.5%

Size of Design Effect

- Cluster sample; 200 clusters, 2/cluster, Pop. A: SE = 2.5%
- Cluster sample; 4 clusters, 100/cluster, Pop. A: SE = 2.5%
- Cluster sample; 200 clusters, 2/cluster, Pop. B: SE = 3.2%
- Cluster sample; 4 clusters, 100/cluster, Pop. B: SE = 19.5%