Statistics 522: Sampling and Survey Techniques

Topic 7a

Topic Overview

This topic will cover

- Variance Estimation on Complex Surveys

Variance Calculation for Complex Surveys

First Step: Recognize Structure of Survey

Example 7.1 (page 223): Bed nets in The Gambia

1. *Stratify* – region (3)

2. *Sample PSU's* – districts (3)

3. *Stratify* – PHC vs. not in villages

    - Could move up, since stratified on same variable within PSU's and strata, but ruins independence of sampling

4. *Sample SSU's* – villages (2)

5. *SRS* – compounds

Section 7.6 (page 242): NCVS

1. *Stratify* – by location, demographics, crime rates to have $> 650,000$

2. *Sample PSU's* – county/group of counties/metropolitan area (84)

    - Large ($> 550,000$) PSU's sampled with probability 1.
    - Otherwise, proportional to size

3. *Sample SSU's* – enumeration districts.

    - About 300-400 households

4. *Sample TSU's* – clusters of houses

5. etc.

Question 5.16 (page 174) 46 school districts with 311 elementary schools with 15086 teachers

1. *Stratify* – by size of school district (245 schools from large districts, 66 from small/medium districts)

2. *Sample PSU's* – schools

3. *Sample SSU's* – teachers within schools

Second Step: Understand Building Blocks

Stratification

- Sample assumed independent within strata

- Take weighted average to get estimate/variance

    - Weight associated with variance is the squared relative size of stratum

- Decrease in variance comes from fixed weighting (or less random)

Sampling (Primary/Secondary/Tertiary) Units

- Typically sampled WOR

- Use Horvitz-Thompson on totals

    - Includes term for variance *within* totals
    - For mean, take weighted average scaled by probability of selection
    - To estimate variance of mean, use relative weights to adjust H-T variance estimator.

Design Effects

- Compare variance calculated from SRS

- To get SRS variance, get mean value and mean of value squared

- Could sample WOR with weights according to design and take variance (ignore weights of pairs of observations).

Ignoring...

- Nonresponse

- Issues of coverage

- Other adjustments to weights.

- Comparisons of groups

Example (Question 5.16)

- $y_{i,j,k} = i$th teacher, jth school in kth stratum

- N_k, $N_{j,k}$ – number of schools/teachers in kth stratum/jth school

- n_k, $n_{j,k}$ – number of schools/teachers *sampled* in kth stratum/jth school

Unbiased estimation of mean of *hrwork*

- Estimate of total in school j: $\hat{t}_{j,k} = \frac{N_{j,k}}{n_{j,k}} \sum_i y_{i,j,k}$

- Estimate of total in stratum k: $\hat{t}_k = \frac{N_k}{n_k} \sum_j \hat{t}_{j,k}$

- Estimate of total: $\hat{t} = \sum_k \hat{t}_k = \sum_k \frac{N_k}{n_k} \sum_j \frac{N_{j,k}}{n_{j,k}} \sum_i y_{i,j,k}$

- Estimate of mean: $\hat{\bar{y}} = \hat{t}/15086 = 17.5$

- Mean of observations: 34.924, Standard deviation: 3.325

- Estimate of number of teachers: 7729

- Estimate of mean with estimated number of teachers: 34.093

Ratio estimation of mean of *hrwork*

- Estimate of total in school j: $\hat{t}_{j,k} = \frac{N_{j,k}}{n_{j,k}} \sum_i y_{i,j,k}$ (same as in unbiased)

- Estimate of mean in stratum k: $\hat{\bar{y}}_k = \frac{1}{\sum_j N_{j,k}} \sum_j \hat{t}_{j,k}$

- Estimate of mean: $\hat{\bar{y}} = \frac{\sum_k N_k \hat{\bar{y}}_k}{\sum_k N_k} = \sum_k N_k \frac{\sum_j N_{j,k} \frac{\sum_i y_{i,j,k}}{n_{j,k}}}{\sum_j N_{j,k}} = 34.249$

Estimate of Variance of Unbiased Estimator

- *At stratum level:* use equation (5.25) on page 147

$$\hat{V}(\hat{t}_k) = \{5780, 69539\}$$

- Estimated standard error of total: $\sum_k \hat{V}(\hat{t}_k) = 75319.52$

- Divide by number squared for SE of mean: $\{0.0182, 0.0355\}$

Estimate of Variance of Ratio Estimator

- *At stratum level:* use equation (5.29) on page 148

$$\hat{V}(\hat{\bar{y}}_k) = \{0.27, 0.033\}$$

- Take weighted average of variances:

$$\hat{Var}(\hat{\bar{y}}) = \sum_k \frac{N_k^2}{(\sum_k N_k)^2} \hat{V}(\hat{\bar{y}}_k)$$

- Estimated standard error: 0.1809