Lecture Note 3: Unequal Prob. Sampling

SRS: $\pi_i = \frac{n}{N}$ ← sample, ← population

Unequal prob samples:

① stratified random sample
② survey non-response

} π_i → survey weight **or** sampling weight

Types of weights:

① design weights
② nonresponse weights **or** poststratification weights

Finite population: $i = 1, \cdots, N$

Sample : $i = 1, \cdots, n$

Horvitz-Thompson

Total: $Y = \sum_{i=1}^{N} y_i$

Estimator: $\hat{Y} = \sum_{i=1}^{n} w_i y_i \rightarrow$ unbiased...

$$E[\hat{Y}] = Y$$

$$E\left[\sum_{i=1}^{n} w_i y_i\right] = Y$$

$$E\left[\sum_{i=1}^{N} 1[i \in S] w_i y_i\right] = Y$$

$$\sum_{i=1}^{N} E\left[1[i \in S] w_i y_i\right] = \sum_{i=1}^{N} y_i$$

$$\sum_{i=1}^{N} \pi_i w_i y_i = \sum_{i=1}^{N} y_i$$

need to set $w_i = \frac{1}{\pi_i}$ ↙ # people in pop represented by i

Weighted average

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i y_i}{\underbrace{\sum_{i=1}^{n} w_i}_{=N}} = \sum_{i=1}^{n} \left(\frac{w_i}{\sum_{i=1}^{n} w_i} \right) y_i = \frac{\hat{Y}}{N}$$

$\rightarrow$ unbiased & consistent for $\mu = \frac{Y}{N}$

Works for:

$\rightarrow$ $w_i = \frac{1}{\pi_i}$

$\rightarrow$ $w_i = \frac{n}{N} \frac{1}{\pi_i}$

$\rightarrow$ $w_i = 1m \times \frac{1}{\pi_i}$

Suppose we have $\bar{y}^s$ in stratum $s \in \{u, r\}$

$$\hat{\mu} = \frac{N^u}{N}\bar{y}^u + \frac{N^r}{N}\bar{y}^r$$

Contrast with:

$$\hat{\mu} = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

$$\pi^u = \frac{n^u}{N^u}, \quad \pi^r = \frac{n^r}{N^r}$$

$$w^u = \frac{1}{\pi^u}, \quad w^r = \frac{1}{\pi^r}$$

Weighted least squares (WLS)

$$\min_{\hat{b}_0^{WLS},\, \hat{b}_1^{WLS}} \sum_{i=1}^{n} w_i\left(y - \hat{b}_0^{WLS} - \hat{b}_1^{WLS} x_i\right)^2$$

$$\hat{\beta}_1^{WLS} = \frac{\sum_{i=1}^{n} w_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} w_i (x_i - \bar{x})^2}$$

$$\beta_1^{POP} = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$\hat{\beta}_1^{WLS}$ is unbiased and consistent for β_1^{POP}

In this sense, $\hat{\beta}_1^{WLS}$ is representative

But efficiency cost?

Structural equation:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

constant, homogeneous

$$y_i = \beta_{0i} + \beta_{1i} x_i + u_i$$

heterogeneous

$$\operatorname{cov}(X, W+Y+Z)$$
$$= \operatorname{cov}(X, W) + \operatorname{cov}(X, Y) + \operatorname{cov}(X, Z)$$

$$\hat{\beta}_1^{WLS} \neq \bar{\beta}_1 = \frac{\sum_{i=1}^{N} \beta_{1i}}{N}$$

$$\beta_1^{POP} \neq \bar{\beta}_1$$

$$\beta_1^{POP} = \frac{\operatorname{cov}(x_i, y_i)}{V[x_i]} = \frac{\operatorname{cov}(x_i, \beta_{0i} + \beta_{1i} x_i + u_i)}{V[x_i]} = \frac{\operatorname{cov}(x_i, \beta_{0i})}{V[x_i]} + \frac{\operatorname{cov}(x_i, \beta_{1i} x_i)}{V[x_i]}$$

$= \bar{\beta}_1$ if x_i randomly assigned