Lecture Note 3: Unequal Probability Sampling

Simple random sample: $\pi_i = \frac{n}{N}$ ← sample size, ← pop

Unequal prob samples:

① Stratified random sample
② survey non-response

$\Big\}$ $\pi_i \rightarrow$ survey weight

or

sampling weight

Two types of weights:

① Design weight
② Nonresponse weights or post-stratification weights

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

Sample (S): $i = 1, \dots, n$

<u>Horvitz-Thompson</u>

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

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

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

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

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

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

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

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

Weighted average

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

$\underbrace{\sum_{i=1}^{n} w_i}_{N}$

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

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

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

Weighted least squares

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

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

as an estimator for

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

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

"representative"

BUT efficiency cost

Structural equation:

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

constant, homogeneous coefs

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

heterogeneous coefs

$\hat{\beta}_1^{WLS}$ not unbiased or consistent estimator of $\bar{\beta}_1 = \frac{\sum_{i=1}^{N} \beta_{1i}}{N}$

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

only = if x_i randomly assigned