Sampling with unqual probabilities of selection (WOR) - theory

Week 9 (6.4)

Stat 260, St. Clair

Horvitz-Thompson Estimator

$$\hat{t}_{HT} = \sum_{i=1}^{n} w_i t_i = \sum \frac{t_i}{\pi_i}$$

Derive the variance of $\hat{t}_{HT}$:

theory

$$Var(\hat{t}_{HT}) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} t_i^2 + 2 \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i<k}}^{N} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} t_i t_k$$

ch. 2

$$Z_i = \begin{cases} 1, & \text{if } i \text{ in sample} \\ 0, & \text{not in sample} \end{cases} \qquad Z_i \sim \text{Bern}(\pi_i)$$

$$P(Z_i = 1) = P(\text{unit } i \text{ in sample}) = \pi_i$$

$$E(Z_i) = \pi_i \qquad Var(Z_i) = \pi_i(1-\pi_i)$$

units i + k in sample

$$E[Z_i Z_k] = 1 \times P(Z_i = 1 \,\&\, Z_k = 1) + 0 \cdot \_\_ = \pi_{ik}$$

$$\hat{t}_{HT} = \sum_{i=1}^{n} \frac{t_i}{\pi_i} \underset{\delta}{=} \sum_{i=1}^{N} Z_i \cdot \frac{t_i}{\pi_i}$$

$$\mathrm{Var}(\hat{t}_{HT}) = \mathrm{Var}\left(\sum_{i=1}^{N} Z_i \frac{t_i}{\pi_i}\right) \xrightarrow{\text{ch.2 thy}} \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \underbrace{2\,\mathrm{Cov}(X,Y)}_{\text{all pairs}}$$

$$= \sum_{i=1}^{N} \mathrm{Var}\left(Z_i \frac{t_i}{\pi_i}\right) + \underbrace{2 \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i<k}}^{N} \mathrm{Cov}\left(Z_i \frac{t_i}{\pi_i}, Z_k \frac{t_k}{\pi_k}\right)}_{\text{all pairs } i,k}$$

$$= \sum_{i=1}^{N} \left(\frac{t_i}{\pi_i}\right)^2 \mathrm{Var}(Z_i) + 2 \sum_{\substack{\text{all} \\ \text{pairs}}} \sum \frac{t_i}{\pi_i} \frac{t_k}{\pi_k} \mathrm{Cov}(Z_i, Z_k)$$

$$\mathrm{Cov}(Z_i, Z_k) = E[Z_i Z_k] - E[Z_i]\,E[Z_k] = \pi_{ik} - \pi_i \pi_k$$

$$\mathrm{Var}(\hat{t}_{HT}) = \sum_{i=1}^{N} \left(\frac{t_i}{\pi_i}\right)^2 \underbrace{\pi_i(1-\pi_i)}_{\text{Var.}} + 2 \underset{\text{pairs}(i,k)}{\sum_i \sum_k} \frac{t_i}{\pi_i}\frac{t_k}{\pi_k}\left(\pi_{ik} - \pi_i\pi_k\right)$$

$$= \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} t_i^2 + 2 \underset{\text{pairs}(i,k)}{\sum_i \sum_k} \frac{\pi_{ik} - \pi_i\pi_k}{\pi_i \pi_k} t_i t_k$$

Horvitz-Thompson Estimator

- **SRS:** We measure $t_i = y_i$ for each unit and $\hat{t}_{HT} = N\bar{y}$.

$$\pi_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N} \qquad \pi_{ik} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

- **SRS:** variance is then

$$Var(\hat{t}_{HT}) = \sum_{i=1}^{N} \frac{1-\frac{n}{N}}{\frac{n}{N}} y_i^2 + 2\sum_{i=1}^{N}\sum_{\substack{k=1 \\ i<k}}^{N} \frac{\frac{\cancel{n}(n-1)}{\cancel{N}(N-1)} - \frac{n}{N}\frac{n}{\cancel{N}}}{\frac{n}{N}\frac{n}{\cancel{N}}} y_i y_k$$

$$\text{....lots of algebra....}$$

$$= N^2\left(1-\frac{n}{N}\right)\frac{S^2}{n}$$

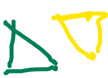
Fill in the missing work... $\longrightarrow$ ch. 2

$$Var(\hat{t}_{HT}) = \frac{N}{n}\left(1-\frac{n}{N}\right)\sum_{i=1}^{N} y_i^2 + \underbrace{\frac{N}{n}\left[\frac{n-1}{N-1} - \frac{n}{N}\right]}_{B}\underbrace{2\sum_{\substack{i \\ \text{pairs}}}\sum_{k} y_i y_k}_{A}$$

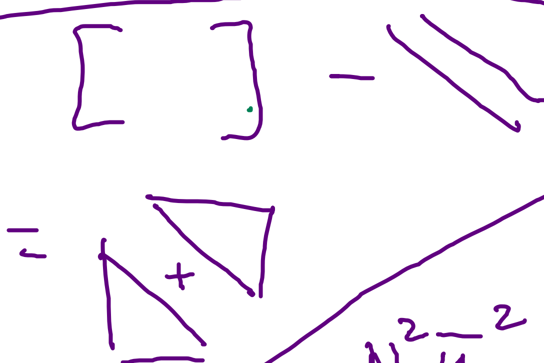
$$B \quad \frac{N}{n}\left[\frac{n-1}{N-1}\cdot\frac{N}{N} - \frac{n}{N}\cdot\frac{N-1}{N-1}\right] = \frac{N}{n}\left[\frac{N(n-1) - n(N-1)}{N(N-1)}\right]$$

$$= \frac{N}{n}\left[\frac{Nn - N - nN + n}{N(N-1)}\right] = \frac{N}{n}\left[\frac{n-N}{N(N-1)}\right] = \frac{N}{n}\,\frac{-1}{N-1}\left(1-\frac{n}{N}\right)$$

$$A \quad \boxed{2\sum_{\substack{i \\ \text{pairs}}}\sum_{k} y_i y_k} = \frac{N^2}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} y_i y_k - \sum_{i=1}^{N} y_i^2 = N^2\,\frac{\sum_{i=1}^{N} y_i}{N}\,\frac{\sum_{i=1}^{N} y_k}{N}$$



$$\begin{array}{c|c}
 & \begin{matrix} y_1 & y_2 & \cdots & y_N \end{matrix} \\
\hline
\begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{matrix} & \begin{bmatrix} y_1^2 & y_1 y_2 & \cdots & y_1 y_N \\ & y_2^2 & & \\ & & \ddots & \\ & & & y_N^2 \end{bmatrix}
\end{array}$$



$$- \sum y_i^2$$

$$= N^2 \bar{y}_u^2 - \sum_{i=1}^{N} y_i^2$$

SRS

$\downarrow$

$$\mathrm{Var}(\hat{t}_{HT})$$

A

$\downarrow$

$$= \frac{N}{n}\left(1-\frac{n}{N}\right)\left[\sum_{i=1}^{N} y_i^2 \times \frac{N-1}{N-1} - \frac{1}{N-1}\left(N^2\bar{y}_U^2 - \sum y_i^2\right)\right]$$

$$= \frac{N}{n}\left(1-\frac{n}{N}\right)\frac{1}{N-1}\left[N\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} y_i^2 - N^2\bar{y}_U^2 + \sum_{i=1}^{N} y_i^2\right]$$

$$= \frac{N}{n}\left(1-\frac{n}{N}\right)\left(\frac{N}{N-1}\right)\underbrace{\left[\sum y_i^2 - N\bar{y}_U^2\right]}_{\text{ch.2 thy } = \sum_{i=1}^{N}(y_i-\bar{y}_U)^2}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y}_U)^2$$

$$Var(\hat{t}_{HT}) = Var(N\bar{y})$$

$$= N^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

✓

Estimating HT variance

A

B

$$Var(\hat{t}_{HT}) = \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i} t_i^2 \right) + 2 \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i<k}}^{N} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} t_i t_k$$

Derive the HT-estimator of $Var(\hat{t}_{HT})$

proposed way

$1/\pi_i$

$$\frac{1}{\pi_{ik}} = \frac{1}{E(Z_i Z_k)}$$

$$\widehat{Var}_{HT}(\hat{t}) = \sum_{i=1}^{N} w_i Z_i \, A + 2 \sum_{\text{pairs}} \sum w_{ik} Z_i Z_k \, B$$

$$= \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i \cdot \pi_i} t_i^2 + 2 \sum_{\substack{i \\ \text{obs. pairs}}} \sum_{k} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}} \cdot \frac{t_i}{\pi_i} \cdot \frac{t_k}{\pi_k}$$

HT- variance est.

Horvitz-Thompson Estimator

Prove that

$$\sum_{i=1}^{N} \pi_i = n$$

$\rightarrow$ use fact $\quad E[z_i] = \pi_i$

$$\underline{\underline{\sum_{i=1}^{N} \pi_i}} = \sum_{i=1}^{N} E[z_i] = E\Big[\underbrace{\sum_{i=1}^{N} z_i}_{\text{must equal } n!}\Big] = E[n] = \underline{\underline{n}}$$

Horvitz-Thompson Estimator

$$\frac{\sum w_i t_i}{\sum w_i}$$

Prove that if there are M observation units in the population,

$$E\Big(\sum_{\text{all sampled obs. units}} w_i\Big) = N$$

$$E\Big[\sum_{\substack{\text{all}\\ \text{sampled}\\ \text{obs. units}}} w_i\Big] = E\Big[\sum_{\substack{\text{all pop.}\\ \text{obs.}\\ \text{unit}}} Z_i \cdot w_i\Big] = \sum_{\substack{\text{all pop.}\\ \text{obs. units}}} \overbrace{w_i}^{\frac{1}{\pi_i}} E[Z_i]$$

$$= \sum_{\substack{\text{all pop.}\\ \text{obs. units}}} \Big(\frac{1}{\pi_i}\Big)\pi_i = \sum_{\substack{\text{all pop.}\\ \text{obs. unit}}} 1 = \Big(\underbrace{1 + 1 + \dots + 1}_{M \text{ obs. units}}\Big) = M$$