Sample Survey

Inclusion Probability

Suppose π_α is the probability the α th unit of the popln will be included. Then we can show

$$\pi_{\alpha\beta} = \pi_\alpha + \pi_\beta + \pi_{\alpha^c\beta^c} - 1$$

$E(n(s)) = \sum_{\alpha=1}^{N} \pi_\alpha$ where $n(s) = \sum_{\alpha=1}^{N} I_\alpha(s)$; $I_\alpha(s) = \begin{cases} 1 & \text{if } s \ni \alpha \\ 0 & \text{o.w.} \end{cases}$

If $P[n(s) = n] = 1 \Rightarrow E(n(s)) = n = \sum_{\alpha=1}^{N} \pi_\alpha$

An estimator of population total is,

$$T_{HT} = \sum_{\alpha \in s} \frac{Y_\alpha}{\pi_\alpha} = \sum_{\alpha=1}^{N} \frac{Y_\alpha}{\pi_\alpha} I_\alpha(s)$$

This is known as Horvitz-Thompson estimator.

$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$, $\bar{Y} = \frac{1}{N} \sum_{\alpha=1}^{N} Y_\alpha$

Then $E(\bar{y}) = \bar{Y}$ [Both SRSWOR and SRSWR]

$$V(\bar{y}) = \begin{cases} \frac{\sigma^2}{n} , & \text{for SRSWR} \\ \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) & \text{for SRSWOR} \end{cases}$$

$s'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$, $S^2 = \frac{1}{N-1} \sum_{\alpha=1}^{N} (Y_\alpha - \bar{Y})^2$

$$\text{Then } E(s'^2) = \begin{cases} \sigma^2 & \text{for SRSWR} \\ S^2 & \text{" SRSWOR} \end{cases}$$

Estimation of population setup under SRSWR two set-up.

Suppose a finite popl^n of size N has N_1 units which possess A.

So $P = \frac{N_1}{N}$

$$Y_\alpha = \begin{cases} 1 & \text{if } \alpha\text{th unit possesses } A \\ 0 & \text{o.w.} \end{cases}$$

$\bar{Y} = \frac{N_1}{N} = P$

For a r.s. of size n of which n_1 unit possess A we define

$$y_i = \begin{cases} 1 & \text{if } i\text{th chosen unit possesses } A \\ 0 & \text{o.w.} \end{cases}$$

$\sum y_i = n_1$

So, $\bar{y} = \frac{n_1}{n} = p$

Now, as $E(\bar{y}) = \bar{Y} \Rightarrow E(p) = P \Rightarrow \hat{P} = p$

$\sigma^2 = \frac{1}{N} \sum Y_\alpha^2 - \bar{Y}^2 = P - P^2 = P(1-P)$

$V_{WR}(\bar{y}) = \frac{\sigma^2}{n} = \frac{P(1-P)}{n}$

$(n-1) s'^2 = \sum y_i^2 - n\bar{y}^2$

$= np(1-p)$

For SRSWR, $E(s'^2) = \sigma^2$

So, $\hat{V}_{WR}(\hat{p}) = \frac{pq}{n-1}$

$$V_{WOR}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) s^2$$

$$\Rightarrow \hat{V}_{WOR}(\hat{p}) = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{np(1-p)}{n-1}$$

Stratified Random Sampling

Define $\bar{Y} = \sum_i W_h \bar{Y}_h$, $W_h = \frac{N_h}{N}$

$$\bar{y}_{st} = \sum W_h \bar{y}_h$$

$$\Rightarrow E(\bar{y}_{st}) = \bar{Y}$$

If Y = popl^n total

$$= \sum_h Y_h$$

$$\hat{=} \sum N_h \bar{y}_h$$

$$V_{WR}(\bar{y}_{st}) = \sum W_h^2 \frac{\sigma_h^2}{n_h} \qquad \left[\because cov(\bar{y}_h, \bar{y}_{h'}) = 0\right]$$

$$V_{WOR}(\bar{y}_{st}) = \sum W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2$$

$$s_h'^2 = \frac{1}{n_h - 1} \sum (y_{hi} - \bar{y}_h)^2$$

$$\Rightarrow E(s_h'^2) = \begin{cases} \sigma_h^2 & \text{for SRSWR} \\ S_h^2 & \text{'' SRSWOR} \end{cases}$$

$$\hat{V}_{WR}(\bar{y}_{st}) = \sum W_h^2 \frac{s_h'^2}{n_h}$$

$$\hat{V}_{WOR}(\bar{y}_{st}) = \sum W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_h'^2$$

→ For proportional allocation,

$$n_h = \frac{n}{N} N_h = n w_h$$

$$V_{wr}(\bar{y}_{st})_{prop} = \frac{1}{n} \sum w_h \sigma_h^2 \doteq \frac{1}{n} \sum w_h s_h'^2$$

$$V_{wor}(\bar{y}_{st})_{prop} = \sum w_h \left(\frac{1}{n} - \frac{1}{N}\right) \sum w_h S_h^2 \doteq \left(\frac{1}{n} - \frac{1}{N}\right) \sum w_h s_h'^2$$

→ To minimize the variance of the estimator when cost is fixed at c' and SRSWOR is used for drawing sample from each stratum.

Then $n_h \propto \frac{w_h S_h}{\sqrt{C_h}}$ [Similar will be the case when cost is being minimized]

→ Gain due to stratification

$$V_{un} - V_{prop} = \frac{1}{n} \sum w_h (\bar{Y}_h - \bar{Y})^2 \geq 0$$

'=' holds when $\bar{Y}_h = \bar{Y} \quad \forall h$

⇒ Stratification isn't required

$$V_{prop} - V_{opt} = \frac{1}{n} \sum w_h \left(\sigma_h - \sum w_h \sigma_h\right)^2 \geq 0 \quad [SRSWR]$$

'=' holds when $\sigma_h = \sum w_h \sigma_h$

i.e. no difference in stratum variability