Stratified Sampling

The idea:

divide heterogeneous population into stratas should be homogenous within, heterogenous between stratas.

N - pop. size
k - # of stratas
N_i - # obs in each strata $\qquad \sum N_i = N$
n_i - # subsample size $\qquad \sum n_i = n$

Procedure:

1) $N = \sum_{i=1}^{k} N_i$

2) Draw n_i from N_i

SRS(WOR)

Cluster sampling: heterogenous within
homogenous between

Stratification (variance reduction)

$E(y)$: $\bar{y} = \frac{1}{n} \sum y_i$ $\bar{y}$ – unbiased

$var(\bar{y}) = \frac{var(y)}{n}$

K strata : w_k – prob. to be in k-th strata

$n_k = W_k \cdot n$

$\hat{y}_{strat} = \sum W_k \cdot \bar{y}_k$

$\bar{y}_k$ – s. mean for k-th strata

$$E(\bar{y}_{srs}) = \mu$$

$$E(\hat{y}_{strat}) = \sum W_k \cdot E(\bar{y}_k) = \sum W_k \mu_k = \mu$$

$$Var(\bar{y}_{srs}) = \sigma^2/n$$

$$Var(\hat{y}_{strat}) = \sum W_k^2 \, Var(\bar{y}_k) =$$

$$= \sum \frac{n_k^{\cancel{2}}}{n^2} \cdot \frac{\sigma_k^2}{\cancel{n_k}} = \frac{1}{n} \sum W_k \cdot \sigma_k^2$$

$$Var(\bar{y}_{srs}) \overset{?}{=} Var(\hat{y}_{strat})$$ z – stratification covariate

$$Var(y) = E(Var(y|z)) + Var(E(y|z)) =$$

$$= E\left(\sum \sigma_k^2 \cdot I(z=k)\right) + Var\left(\sum \mu_k \cdot I(z=k)\right) =$$

$$= \sum \sigma_k^2 \cdot E(I(z=k)) + E\left(\left(\sum \mu_k \cdot I(z=k)\right)^2\right) -$$

$$\left(E\left(\sum \mu_k \cdot I(z=k)\right)\right)^2 = \sum \sigma_k^2 \cdot W_k +$$

$$+ \sum \mu_k^2 \cdot W_k - \mu^2 = \sum \sigma_k^2 \cdot W_k + \sum W_k (\mu_k - \mu)^2$$

$$Var(\bar{y}_{srs}) = \underbrace{\frac{1}{n} \sum \sigma_k^2 \cdot W_k}_{Var(\hat{y}_{strat})} + \frac{1}{n} \sum W_k (\mu_k - \mu)^2$$

within var

between var

$n_1, n_2, \dots, n_k$?

(I) min cost of survey (given precision)
(II) max precision (given cost)

1) Equal allocation:

$$n_i = \frac{n}{K}$$

2) Proportional allocation:

$$n_i \propto N_i$$

$$n_i = \delta N_i$$

$$n = \delta . N \Rightarrow \delta = n/N$$

$$n_i = \frac{n}{N} \cdot N_i$$

- Precision PSS $\geq$ Precision SRS

$\Leftrightarrow$ proportion / strata sizes are the same

3) Optimum (Neyman) allocation

$n_i \propto N_i S_i$ $\qquad S_i$ - unb. est. of var for i-th strata

$n_i = \delta^* N_i S_i$

$$n = \delta^* \cdot \Sigma N_i S_i$$

$$\delta^* = \frac{n}{\Sigma N_i S_i}$$

$$n_i = \frac{N_i S_i}{\Sigma N_i S_i}$$

- Precision OSS $\geqslant$ Precision PSS

  $=$ if var within each strata are equal

Cost:

$$C = C_0 + \sum C_i n_i \quad \swarrow \text{cost per unit for i-th strata}$$

$$TC = FC + VC$$

$\min_{n_i} Var(\bar{y})$ s.t. budget constraint

$$\mathcal{L} = Var(\bar{y}) + \lambda^2 (c - c_0) =$$

$$= \sum W_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \cdot S_i^2 + \lambda^2 \sum C_i n_i =$$

$$= \sum \frac{W_i^2 S_i^2}{n_i} + \lambda^2 \sum C_i n_i - \sum \frac{W_i^2 S_i^2}{N_i} =$$

$$= \sum \left(\frac{W_i S_i}{\sqrt{n_i}} - \lambda \sqrt{C_i n_i} \right)^2 + B$$

min $Var(\bar{y})$ at $\dfrac{W_i S_i}{\sqrt{n_i}} = \lambda \sqrt{C_i n_i}$

$$n_i = \frac{1}{\lambda} \cdot \frac{W_i S_i}{\sqrt{C_i}}$$

$$\lambda \begin{cases} \nearrow \text{min Var , fixed cost} & \text{(i)} \\ \searrow \text{min cost , fixed var} & \text{(ii)} \end{cases}$$

(i) $\quad C = C_0^*$

$$\sum c_i n_i = C_0^*$$

$$\sum c_i \cdot \frac{W_i S_i}{\lambda \sqrt{c_i}} = C_0^*$$

$$\lambda_{FC} = \frac{\sum \sqrt{c_i} \cdot W_i S_i}{C_0^*}$$

$$n_i^{FC} = \frac{1}{\lambda_{FC}} \cdot \frac{W_i S_i}{\sqrt{c_i}}$$

(ii) $\quad V = V_0^*$

$$\sum \left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 S_i^2 = V_0^*$$

$$\sum \frac{\lambda \cdot \sqrt{c_i}}{W_i S_i} \cdot W_i^2 \cdot S_i^2 = V_0^* + \sum \frac{W_i^2 \cdot S_i^2}{N_i}$$

$$\lambda_{FV} = \frac{V_0^* + \frac{\sum W_i^2 S_i^2}{N_i}}{\sum W_i S_i \cdot \sqrt{c_i}}$$

$$n_i^{FV} = \frac{1}{\lambda_{FV}} \cdot \frac{W_i S_i}{\sqrt{c_i}}$$

min sample sizes	$\bar{X}$	$\hat{p}$	e- MF
SRS	$\frac{z^2_{\alpha/2} \cdot \sigma^2}{e^2}$	$\frac{z^2\, p \cdot (1-p)}{e^2}$	
PSS	$\frac{z^2_{\alpha/2} \cdot \Sigma W_k \cdot \sigma_k^2}{e^2}$	$\frac{z^2_{\alpha/2} \cdot \Sigma W_k \cdot p_k(1-p_k)}{e^2}$	
OSS	$\frac{z^2_{\alpha/2} \cdot (\Sigma W_k \cdot \sigma_k)^2}{e^2}$	$\frac{z^2_{\alpha/2} \cdot \left(\Sigma W_k \cdot \sqrt{p_k \cdot (1-p_k)}\right)^2}{e^2}$	

$\sigma_i = \sigma_j = \sigma \quad i \neq j$

$W_k = 1/k$

$$\frac{(\Sigma W_k \cdot \sigma_k)^2}{e^2} = \frac{(\Sigma \, 1/k \cdot \sigma)^2}{e^2} = \frac{\sigma^2}{e^2}$$