DOUBLE SAMPLING

A number of sampling techniques depend on the possession of advanced information about an auxilliary variable X. Ratio and regression estimators require a knowledge of the population mean $\bar{X}$. If it is desired to stratify the population according to the values of X, their frequency distribution must be known.

When such information is lacking, it is sometimes relatively cheap to take the large preliminary sample in which X alone is measured. The purpose of this sample is to furnish a good estimate of $\bar{X}$ or of the frequency distribution of X. In a survey whose function is to make estimates for some other variable Y, it may pay to devote part of the resources to this preliminary sample, although this means that the size of the sample in the main survey on Y must be decreased. This technique is known as Double Sampling or Two Phase Sampling. This technique is profitable only if the gain in precision from the ratio or regression estimates or stratification more than offsets. The loss in precision due to the reduction in the size of the main circle.

Double sampling may be appropriate when the information about x is on file-cards that have not be tabulated. For instance in survey of the German Civilian Population in 1945, the sample from any town was usually drawn from rationing registration list. In addition to geographic stratification within the town for which the data were already available, stratification by age and sex ~~way~~ was proposed. Since the sample had to be drawn in a hurry and since the list were in constant use, tabulation of the complete age-sex distribution was not feasible. A moderately large systematic sample however could be drawn quickly. Each person drawn was classified into the appropriate age-sex class. From this data the much smaller list of persons to be interviewed was selected.

Double Sampling for stratification :

Population is classified into K-strata.

1st sample is of size n'.

$W_h = \frac{N_h}{N}$ = ~~Proportion of~~ Proportion of population individuals falling in the stratum h.

$w_h = \frac{n'_h}{n'}$ = Proportion of 1st sample individuals falling in the stratum h.

w_h is an u.e. of W_h.

The 2nd sample is of size n in which y s are measured.

n_h units are drawn from h^{th} stratum.

Usually the second sample of size n_h from stratum h is a subsample of the n'_h units of the stratum.

Objective of the 1st sample >> To estimate the stratum weights W_h.

Objective of the 2nd sample >> To estimate the stratum means $\bar{Y}_h$.

Population Mean: $\bar{Y} = \sum_{h=1}^{K} W_h \bar{Y}_h$

As an estimator, we use the following $\bar{y}_{st}^{d} = \sum_{h=1}^{K} w_h \bar{y}_h$

Ultimate Objective >> To choose n' & n_h is such a way that

$Var(\bar{y}_{st}^{d})$ is minimized subject to a given cost.

Result: 1 $\quad \bar{y}_{st}^{d}$ is unbiased for $\bar{Y}$.

$$E(\bar{y}_{st}^{d}) = E_1 E_2 (\bar{y}_{st}^{d} \mid w_h)$$

E_2: Conditional Expectation over the samples for which w_h is fixed.

E_1: Unconditional Expectation over the first sample.

$$E_2(\bar{y}_{st}^{d} \mid w_h) = E_2\left(\sum_{h=1}^{K} w_h \bar{y}_h \mid w_h\right)$$

$$= \sum_{h=1}^{K} w_h \, E_2(\bar{y}_h \mid w_h)$$

$$= \sum_{h=1}^{K} w_h \bar{Y}_h$$

$$E_1 E_2(\bar{y}_{st}^{d} \mid w_h) = E_1\left[\sum_{h=1}^{K} w_h \bar{Y}_h\right] = \sum_{h=1}^{K} E_1(w_h)\bar{Y}_h = \sum_{h=1}^{K} W_h \bar{Y}_h = \bar{Y}$$

$$\therefore \quad E(\bar{y}_{st}^{d}) = \bar{Y}$$

Result: 2 $\quad$ If the 1st sample is random and of size n', the second sample is a random subsample of the first, i.e. $n_h = n_h' \upsilon_h$, where $0 \le \upsilon_h \le 1$ and if the υ_h's are fixed, then

$$Var(\bar{y}_{st}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_{h=1}^{K} \frac{W_h S_h^2}{n'}\left(\frac{1}{\upsilon_h} - 1\right)$$

Proof:

Suppose that Y values are measured for all of the n_h' first stage units in ~~the~~ stratum h, not just for the random subsample of size n_h from n_h'.

We have $\quad \bar{y}_{st}^{d} = \sum_{h=1}^{K} w_h \bar{y}_h$, where $\bar{y}_h = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj}$, $w_h = \frac{n_h'}{n'}$

Define $\quad \bar{y}_h' = \frac{1}{n_h'}\sum_{j=1}^{n_h'} y_{hj}$ & $\bar{y}' = \sum_{h=1}^{K} w_h \bar{y}_h'$

$$Var(\bar{y}_{st}^{d}) = V_1 E_2(\bar{y}_{st}^{d} | w_h) + E_1 V_2(\bar{y}_{st}^{d} | w_h)$$

Now, $$E_2(\bar{y}_{st}^{d} | w_h) = \sum_{h=1}^{K} w_h E_2(\bar{y}_h | w_h)$$

$$= \sum_{h=1}^{K} w_h \bar{y}_h'$$

$$= \bar{y}'$$

$$V_1 E_2(\bar{y}_{st}^{d} | w_h) = V_1(\bar{y}') = S^2\left(\frac{1}{n'} - \frac{1}{N}\right)$$

$$V_2(\bar{y}_{st}^{d} | w_h) = V_2\left(\sum_{h=1}^{K} w_h \bar{y}_h \,|\, w_h\right)$$

$$= \sum_{h=1}^{K} w_h^2 V_2(\bar{y}_h | w_h)$$

$$= \sum_{h=1}^{K} w_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{n_h'}\right)$$

$$= \sum_{h=1}^{K} w_h^2 S_h^2 \left(\frac{1}{n_h' v_h} - \frac{1}{n_h'}\right)$$

$$= \sum_{h=1}^{K} \frac{w_h^2 S_h^2}{n_h'} \left(\frac{1}{v_h} - 1\right)$$

$$= \sum_{h=1}^{K} \frac{w_h S_h^2}{n'} \left(\frac{1}{v_h} - 1\right) \quad [\because n_h' = n' w_h]$$

$\therefore$ ~~Var~~

$$E_1 V_2(\bar{y}_{st}^{d} | w_h) = \sum_{h=1}^{K} E_1(w_h) \frac{S_h^2}{n'} \left(\frac{1}{v_h} - 1\right)$$

$$= \sum_{h=1}^{K} \frac{W_h S_h^2}{n'} \left(\frac{1}{v_h} - 1\right)$$

Thus,

i.e. $$Var(\bar{y}_{st}^{d}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_{h=1}^{K} \frac{W_h S_h^2}{n'} \left(\frac{1}{v_h} - 1\right)$$

Corollary:

The total variance can be partitioned as

$$(N-1)\, S^2 = \sum_{h=1}^{K} (N_h - 1) S_h^2 + \sum_{h=1}^{K} N_h (\bar{Y}_h - \bar{Y})^2.$$

Define, $g' = \dfrac{N-n'}{N-1}$

Multiplying both sides by $\dfrac{g'}{n'N}$

$$\frac{g'}{n'N} (N-1)\, S^2 = \frac{g'}{n'N} \sum_{h=1}^{K} (N_h - 1) S_h^2 + \frac{g'}{n'N} \sum_{h=1}^{K} N_h (\bar{Y}_h - \bar{Y})^2$$

or, $$S^2 \left(\frac{1}{n'} - \frac{1}{N}\right) = \frac{g'}{n'} \sum_{h=1}^{K} \left(W_h - \frac{1}{N}\right) S_h^2 + \frac{g'}{n'} \sum_{h=1}^{K} W_h (\bar{Y}_h - \bar{Y})^2$$

$$\therefore \; Var(\bar{y}_{st}^{d}) = \frac{g'}{n'} \sum_{h=1}^{K} \left(W_h - \frac{1}{N}\right) S_h^2 + \frac{g'}{n'} \sum_{h=1}^{K} W_h (\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{K} \frac{W_h S_h^2}{n'} \left(\frac{1}{v_h} - 1\right)$$

$$g' = \frac{N-n'}{N-1}$$

$$\frac{g'}{n'} = \frac{\frac{N}{n'} - 1}{N-1}$$

$$\frac{g'}{n'} - \frac{1}{n'} = \frac{\frac{N}{n'} - 1}{N-1} - \frac{1}{n'} = \frac{\frac{N}{n'} - 1 - \frac{N}{n'} + \frac{1}{n'}}{N-1} = \frac{\frac{1}{n'} - 1}{N-1}$$

$$= \frac{1}{Nn'} \, \frac{Nn'\left(\frac{1}{n'} - 1\right)}{N-1} = \frac{N - Nn'}{Nn'(N-1)}.$$

$$= \frac{1}{N} \left(\frac{N - n' + n' - Nn'}{n'(N-1)} \right)$$

$$= \frac{1}{N} \left(\frac{N-n'}{n'(N-1)} - 1 \right)$$

$$= \frac{1}{N} \left(\frac{g'}{n'} - 1 \right)$$

Scanned with CamScanner

$$Var(\bar{y}_{st}^{d}) = \frac{g'}{n'}\sum_{h=1}^{K}\left(W_h - \frac{1}{N}\right)S_h^2 + \frac{g'}{n'}\sum_{h=1}^{K} W_h(\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{K}\frac{W_h S_h^2}{n'}\left(\frac{1}{\upsilon_h} - 1\right)$$

$$= \sum_{h=1}^{K} W_h S_h^2 \left(\frac{g'}{n'} - \frac{1}{n'} + \frac{1}{n'\upsilon_h}\right) - \frac{1}{N}\frac{g'}{n'}\sum_{h=1}^{K} S_h^2 + \frac{g'}{n'}\sum_{h=1}^{K} W_h(\bar{Y}_h - \bar{Y})^2$$

$$= \sum_{h=1}^{K} W_h S_h^2 \left(\frac{1}{N}\left(\frac{g'}{n'} - 1\right) + \frac{1}{n'\upsilon_h}\right) - \frac{1}{N}\frac{g'}{n'}\sum_{h=1}^{K} S_h^2 + \frac{g'}{n'}\sum_{h=1}^{K} W_h(\bar{Y}_h - \bar{Y})^2$$

$$= \frac{g'}{Nn'}\sum_{h=1}^{K}(W_h - 1)S_h^2 + \sum_{h=1}^{K} W_h S_h^2\left(\frac{1}{n'\upsilon_h} - \frac{1}{N}\right) + \frac{g'}{n'}\sum_{h=1}^{K} W_h(\bar{Y}_h - \bar{Y})^2 \quad \text{---} (*)$$

For most of the applications the factor $\frac{g'}{Nn'}$ is negligable.

$Var(\bar{y}_{st}^{d})$ is then simplified as the sum of 2nd and 3rd term.

Result: 3

$V = Var(\bar{y}_{st}^{d})$

Then, $n'\left(V + \frac{S^2}{N}\right) = \left(S^2 - \sum_{h=1}^{K} W_h S_h^2\right) + \sum_{h=1}^{K}\frac{W_h S_h^2}{\upsilon_h}$

$$\left[V = \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{n' \vartheta_h} - \frac{1}{N} \right) + \frac{g'}{N n'} \sum_{h=1}^{k} (W_h - 1) S_h^2 \right.$$

$$+ \frac{g'}{n'} \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2$$

$$n' V = \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{\vartheta_h} - \frac{n'}{N} \right) + \frac{g'}{N} \sum_{h=1}^{k} (W_h - 1) S_h^2 + g' \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2$$

Again,

$$(N-1) S^2 = \sum_{h=1}^{k} (N_h - 1) S_h^2 + \sum_{h=1}^{k} N_h (\bar{Y}_h - \bar{Y})^2$$

$$\therefore \frac{n' S^2}{N} = \frac{n'}{N-1} \left[\sum_{h=1}^{k} \left(\frac{N_h}{N} - \frac{1}{N} \right) S_h^2 + \sum_{h=1}^{k} \frac{N_h}{N} (\bar{Y}_h - \bar{Y})^2 \right]$$

$$= \frac{n'}{N-1} \left[\sum_{h=1}^{k} \left(W_h - \frac{1}{N} \right) S_h^2 + \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2 \right]$$

$$V + \frac{S^2}{N} = \frac{S^2}{n'} + \sum_{h=1}^{k} \frac{W_h S_h^2}{n'} \left(\frac{1}{\vartheta_h} - 1 \right)$$

$$n' \left(V + \frac{S^2}{N} \right) = \left(S^2 - \sum_{h=1}^{k} W_h S_h^2 \right) + \sum_{h=1}^{k} \frac{W_h S_h^2}{\vartheta_h} \quad \text{(proved)}$$

---

term.

Proof of (*). $\mathrm{Var}(\bar{y}_{st}^{d}) = \left(\frac{N-n'}{n'N} \right) \frac{1}{(N-1)} \left[\sum_{h=1}^{k} (N_h - 1) S_h^2 + \sum_{h=1}^{k} N_h (\bar{Y}_h - \bar{Y})^2 \right] + \sum_{h=1}^{k} \frac{W_h S_h^2}{n'} \left(\frac{1}{\vartheta_h} - 1 \right)$

$$= \left(\frac{N-n'}{n'N} \right) \left(\frac{1}{N-1} \right) \sum_{h=1}^{k} (N_h - 1) S_h^2 + \frac{g'}{n'} \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{n' \vartheta_h} - \frac{1}{N} \right)$$

$$- \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{n'} - \frac{1}{N} \right)$$

$$= \frac{g'}{n'} \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{n' \vartheta_h} - \frac{1}{N} \right) +$$

$$\left(\frac{N-n'}{n'N} \right) \frac{1}{(N-1)} \sum_{h=1}^{k} \left\{ (N_h - 1) - (N-1) W_h \right\} S_h^2$$

$$= \frac{g'}{n'} \sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{k} W_h S_h^2 \left(\frac{1}{n' \vartheta_h} - \frac{1}{N} \right) + \frac{g'}{n'N} \sum_{h=1}^{k} (W_h - 1) S_h^2$$

$$\left[\because N W_h = N_h \ \& \ g' = \frac{N-n'}{N-1} \right]$$

Optimal Allocation Problem:

Objective is to choose n' and ϑ_h such that $Var(\bar{y}^d_{st})$ is minimum ~~such~~ subject to a given cost.

$c' =$ Per unit cost of classification

$c_h =$ Per unit cost of measurement in stratum h.

Total Cost : $$C = c'n' + \sum_{h=1}^{K} c_h n_h$$

$$= c'n' + \sum_{h=1}^{K} c_h \vartheta_h n_h'$$

$$= c'n' + \sum_{h=1}^{K} c_h \vartheta_h \underbrace{w_h}_{\text{Random}} n'$$

Expected Cost

$$C^* = E(C) = n'\left[c' + \sum_{h=1}^{K} c_h \vartheta_h E(w_h)\right]$$

$$= n'\left[c' + \sum_{h=1}^{K} c_h \vartheta_h W_h\right]$$

$C^*\left(V + \frac{S^2}{N}\right)$ is free of n'

$$C^*\left(V + \frac{S^2}{N}\right) = \left[c' + \sum_{h=1}^{K} c_h \vartheta_h W_h\right]\left[\left(S^2 - \sum_{h=1}^{K} W_h S_h^2\right) + \sum_{h=1}^{K} \frac{W_h S_h^2}{\vartheta_h}\right]$$

Define,

$$a_0 = c'$$

$$a_h = c_h \vartheta_h W_h \quad , h = 1(1)K$$

$$b_0 = S^2 - \sum_{h=1}^{K} W_h S_h^2$$

$$b_h = \frac{W_h S_h^2}{\vartheta_h} \quad , h = 1(1)K$$

$$\therefore\ c^*\left(v+\frac{S^2}{N}\right)=\left(\sum_{h=0}^{K}a_h\right)\left(\sum_{h=0}^{K}b_h\right)$$

$$\geq\left(\sum_{h=0}^{K}\sqrt{a_h b_h}\right)^2 \text{, from C-S-inequality.}$$

sign of equality holds if $a_h \propto b_h\ \forall h$

$$\therefore\ \frac{a_0}{b_0}=\frac{a_1}{b_1}=\frac{a_2}{b_2}=\cdots=\frac{a_K}{b_K}$$

$$\frac{c'}{S^2-\sum_{h=1}^{K}W_h S_h^2}=\frac{c_h v_h W_h}{W_h S_h^2/v_h}=\frac{c_h v_h^2}{S_h^2}\quad \forall h=1(1)K$$

$$v_h=\frac{\sqrt{c'}\,S_h}{\sqrt{c_h\left(S^2-\sum_{h=1}^{K}W_h S_h^2\right)}}\ ;\ h=1(1)K$$

~~Now~~ Again,

$$c^*=n'\left[c'+\sum_{h=1}^{K}c_h v_h W_h\right]$$

$$=n'\left[c'+\sum_{h=1}^{K}c_h W_h\frac{\sqrt{c'}\,S_h}{\sqrt{c_h\left(S^2-\sum_{h=1}^{K}W_h S_h^2\right)}}\right]$$

$$=n'\left[c'+\sum_{h=1}^{K}\frac{\sqrt{c'}\,W_h S_h\sqrt{c_h}}{\sqrt{S^2-\sum_{h=1}^{K}W_h S_h^2}}\right]$$

$$\therefore\ n'=\frac{c^*}{\left[c'+\sum_{h=1}^{K}\frac{\sqrt{c'}\,W_h S_h\sqrt{c_h}}{\sqrt{S^2-\sum_{h=1}^{K}W_h S_h^2}}\right]}$$

$$n'=\frac{c^*}{c'}\left[\frac{1}{1+\sum_{h=1}^{K}\left(\sqrt{\frac{c_h}{c'}}\frac{W_h S_h}{\sqrt{S^2-\sum_{h=1}^{K}W_h S_h^2}}\right)}\right]$$

$$\therefore\ V = s^2\left(\frac{1}{n'}-\frac{1}{N}\right)+\sum_{h=1}^{K}\frac{W_h S_h^2}{n'}\left(\frac{1}{v_h}-1\right)$$

$$\cancel{=\sum_{h=1}^{K}\frac{W_h S_h^2}{n'}\left(\frac{1}{v_h}-1+s^2\right.}$$

$$=\frac{1}{n'}\left\{s^2+\sum_{h=1}^{K}W_h S_h^2\left(\frac{1}{v_h}-1\right)\right\}-\frac{s^2}{N}$$

$$=\frac{1}{n'}\left\{\left(s^2-\sum_{h=1}^{K}W_h S_h^2\right)+\sum_{h=1}^{K}\frac{W_h S_h^2\sqrt{C_h\left(s^2-\sum_{h=1}^{K}W_h S_h^2\right)}}{\sqrt{C'}\,S_h}\right\}-\frac{s^2}{N}$$

$$=\frac{1}{n'}\left\{A_0+\sum_{h=1}^{K}W_h S_h\sqrt{\frac{C_h}{C'}A_0}\right\}-\frac{s^2}{N}$$

where, $A_0 = s^2-\sum_{h=1}^{K}W_h S_h^2$

$$=\frac{C'}{C^*}\left(1+\sum_{h=1}^{K}\sqrt{\frac{C_h}{C'}}\frac{W_h S_h}{\sqrt{A_0}}\right)\left(A_0+\sum_{h=1}^{K}W_h S_h\sqrt{\frac{C_h A_0}{C'}}\right)-\frac{s^2}{N}$$

$$=\frac{C'}{C^*}\left[A_0+2\sum_{h=1}^{K}W_h S_h\sqrt{\frac{C_h A_0}{C'}}+\frac{1}{C'}\left(\sum_{h=1}^{K}W_h S_h\sqrt{C_h}\right)^2\right]-\frac{s^2}{N}$$

$$V=\frac{C'}{C^*}\left[A_0+2\sqrt{\frac{A_0}{C'}}\sum_{h=1}^{K}W_h S_h\sqrt{C_h}\right]+\frac{1}{C^*}\left(\sum_{h=1}^{K}W_h S_h\sqrt{C_h}\right)^2-\frac{s^2}{N}$$

$$\cancel{V=\frac{C'}{C^*}\left[A_0+2\sqrt{\frac{A_0}{C'}}A_1\right]+\frac{A_1^2}{C^*}-\frac{s^2}{N}}$$

$$\boxed{\begin{aligned}&\text{where, } A_0 = s^2-\sum_{h=1}^{K}W_h S_h^2\\ &A_1=\sum_{h=1}^{K}W_h S_h\sqrt{C_h}\end{aligned}}$$

$$V=\frac{1}{C^*}\left[A_1^2+A_0C'+2A_1\sqrt{A_0C'}\right]-\frac{s^2}{N}$$

$$\boxed{V=\frac{\left(A_1+\sqrt{A_0C'}\right)^2}{C^*}-\frac{s^2}{N}}$$