## Sampling Scheme: Advantages & Disadvantages:

Though cluster sampling is economical under circumstances, it is generally less efficient than sampling of individual units directly A compare between direct sampling and cluster sampling nuits can be achieved by selecting a sample of clusters and surveying only a sample of units within each sample cluster instead of complete enum-erating all the units in the sample cluster instead of complete enumas two-stage sampling, since the units is selected in two stages. Here clusters are turned as first stage unit (fsn) and the ultimate observational unit are turned as second stage unit (ssn). It may be noted that this procedure can be generalized to multistage sampling sampling where the sampling units at each stage are the cluste of units of the next stage.

This procedure being a compromise between with or direct sampling and cluster sampling is expected to be

- i) more efficient turn uni-stage sampling, less efficient than cluster sampling from consideration of operational convenience and cost.
- ii) less efficient than uni-stage sampling, more efficient than cluster sampling from the view point of variability, when the sample size in turns of the number of unit is fixed.

In practice, it wouldy happens that we have more information for group of sampling units from individual units. Hence, the groups are taken as itsu, the information available for them can be used in effecting good stratification or arrangement and in selection of the sample of isu. Further since the ssn's are selected only from sample isus, it would be practical to collect information about the ssu at the time of listing them and use the information for obtaining a better sample of ssu, Because a this, it may be possible that a multistage design where the information available at each stage is properly utilised is more efficient than unique unistage campling even from the point of view of sampling variety and applied.

Notations:

N: Number of FSUS

"Mi: Number of 85Us within it FSU (1=1(1)N)

Tij: jth gov within the ith Fou of the population (i=1(1)N, j=1(1)Mi)

n: Number of sample FSUs

mi: Number of ssus selected from its sample FSU (1=1(1)n)

Yij: jth selected SSU from the ith sample FSU (i=1(1)n)

If the total values of the selected FSUs where are known, it can be seen that it could be possible to get an estimator of the population total Y with the help of the probability scheme at the first stage as in cluster sampling. But in two stage sampling the actual total of the selected FSUs are not known, and hence they also have to be estimated on the basis of the selected SSUS using the probability scheme adopted in selecting them.

·  $\forall i = \text{total}$  of the ith sample FSU, i = 1(1)n. In case of cluster sampling, Y is estimated by  $\sum_{i=1}^{n} a_i y_i$ , where  $a_i = \text{inflation}$  factor.

But in case of two stage sampling, yi's are also unknown and they also have to be estimated. There Their estimate is of the form

$$\hat{y}_{i} = \sum_{j=1}^{m} a_{ij} y_{ij}$$
, where  $a_{ij}$  is the inflation factor.

Ultimately, 
$$\hat{Y} = \sum_{i=1}^{n} ai\hat{y}_{i}$$
 =  $\sum_{i=1}^{n} ai \left[ \sum_{j=1}^{m_i} a_{ij} y_{ij} \right]$ 

 $\overline{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} = Mean of the sampled units of the its selected FSU (i=1(1)n)$ 

$$\frac{y_i}{M_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i} = Actual mean of all observations in the jtm selected FSU (i=1(1)n)$$

Under SRS,  $\frac{y_i}{M_i}$  is unbiasedly estimated by  $\frac{1}{m_i} \sum_{j=1}^{m_i} \forall i = \overline{\forall} i$ 

Thus,  $y_i$  is unbiasedly estimated by  $\frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij} = M_i y_i = \hat{y}_i$ , i = 1(1)n.

Estimated true mean of all units of the sample FSUs is  $\frac{1}{n}\sum_{i=1}^{n}\hat{y_i}$ .

$$\therefore \hat{Y} = \frac{N}{N} \sum_{i=1}^{N} \hat{y}_i = \frac{N}{N} \sum_{i=1}^{N} \frac{m_i}{m_i} \sum_{j=1}^{m_i} \hat{y}_j = \frac{N}{N} \sum_{i=1}^{N} \frac{m_i}{m_i} \hat{y}_j$$

## Expectation and Variance:

$$E(\hat{Y}) = E_1E_2(\hat{Y})$$

where, E2 k V2 : Conditional expectation/variance over the SSUs
for given sample FSU

EI & VI: Unconditional mean/variance of the FSUs.

Notation:  $M = \frac{1}{N} \sum_{i=1}^{N} M_i = Average number of ssus per FSUs.$ 

$$S_b^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{M_i \overline{Y_i}}{M} - \overline{Y} \right)^2 = Between, FSU variance$$

$$S_{N_{i}}^{2} = \frac{1}{M_{i-1}} \sum_{j=1}^{M_{i}} (Y_{ij} - \overline{Y_{i}})^{2} = i^{4n} FSU Variance (i=1(1)N)$$

Now.

$$E(\hat{Y}) = E_1 E_2 \left( \hat{Y} \mid FSU \right) = E_1 \left[ E_2 \left( \frac{N}{n} \sum_{i=1}^{n} M_i \bar{y}_i \right) \mid FSU \right]$$

$$= E_1 \left[ \frac{N}{n} \sum_{i=1}^{n} M_i E_2(\bar{y}_i \mid FSU) \right]$$

$$= E_1 \left[ \frac{N}{n} \sum_{i=1}^{n} M_i \overline{Y}_i \right] = E_1 \left[ \frac{N}{n} \sum_{i=1}^{n} M_i \frac{y_i}{M_i} \right]$$

$$= E_1 \left[ \frac{N}{n} \sum_{i=1}^{n} y_i \right]$$

$$= Actual Mean of the im Sample FSU = 1(1)n$$

$$= N \quad E_1 \left[ \frac{1}{N} \sum_{i=1}^{N} y_i \right] = N \cdot \frac{1}{N} \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} Y_i = Y$$

$$= N \quad Y \quad \left[ \frac{1}{N} \sum_{i=1}^{N} y_i \right] = \frac{1}{N} \sum_{i=1}^{N} Y_i = Y = \frac{Y}{N} \right]$$

$$= Y$$

Hence,  $\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} M_i \bar{y}_i$  is an unbiased estimator of Y.

$$F_2\left(\hat{Y}\mid FSU\right) = \frac{N}{n} \sum_{i=1}^{n} \forall i$$

$$V_{1} F_{2} \left(\hat{Y} \mid FSU\right) = V_{1} \left(\frac{N}{n} \sum_{i=1}^{n} y_{i}\right) = N^{2} V_{1} \left(\frac{1}{n} \sum_{i=1}^{n} y_{i}\right)$$

$$= \frac{N^{2} \cdot \frac{1}{n} Var(y_{i}) \left(\frac{N-n}{N-1}\right)}{N^{2} \cdot \frac{1}{n} Var(y_{i}) \left(\frac{N-n}{N-1}\right)}$$

$$E(y_i) = \frac{1}{N} \sum_{i=1}^{N} Y_i = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij} = \frac{NM \overline{Y}}{N} = M \overline{Y} \qquad \left[ :: M = \frac{1}{N} \sum_{j=1}^{N} M_i \right]$$

$$Var(y_i) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - M \overline{Y})^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left( M_i \overline{Y}_i - M \overline{Y} \right)^2$$

$$= \frac{M^2}{N} \sum_{i=1}^{N} \left( \frac{M_i \overline{Y}_i}{M} - \overline{Y} \right)^2$$

$$= \frac{M^2}{N} (N-1) S_b^2$$

$$V_{1}E_{2} \left(\hat{Y} \mid FSU\right) = \frac{N^{2}}{n} \frac{(N-n)}{(N-1)} \cdot \frac{M^{2}}{N} (N-1) S_{b}^{2}$$

$$= N^{2}M^{2} \left(1-\frac{1}{T}\right) \frac{S_{b}^{2}}{n} \quad \text{where} \quad \frac{1}{T} = \frac{n}{N}$$

$$V_{2}\left(\hat{Y}\mid FSU\right) = V_{2}\left[\begin{array}{cccc} \frac{N}{n} & \sum\limits_{i=1}^{n} \frac{M_{i}}{m_{i}} & \sum\limits_{i=1}^{m} \frac{M_{i}^{2}}{m_{i}^{2}} & \sum\limits_{i=1}^{m} \frac{M_{i}^{2}}{$$

Note: Unbiased Estimator of 
$$\overline{Y}$$
 is  $\hat{\overline{Y}} = \frac{\hat{Y}}{NM} = \frac{1}{Mn} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$ 

$$= \frac{1}{n N M^2} \sum_{i=1}^{N} \frac{M_i^2 S_{W_i}^2}{m_i} (1-h_i) + (1-h_i) \frac{S_b^2}{N^2}$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{M_{i}^{2}}{m_{i}}S_{N_{i}}^{2}(i-f_{i}) \text{ is estimated unbiasedly by } \frac{1}{n}\sum_{i=1}^{n}\frac{M_{i}^{2}}{m_{i}}J_{N_{i}}^{2}(i-f_{i})$$

$$\frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2 S_{W_i}^2}{m_i} (i-\frac{1}{n}) \text{ is estimated unbiasedly by } \frac{N^2}{n^2} \sum_{i=1}^{n} \frac{M_i^2 S_{W_i}^2}{m_i} (i-\frac{1}{n})$$

Fstimation of  $S_b^2$ :

$$\mathcal{Q}_{b}^{2} = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{M_{i} \overline{Y_{i}}}{M} - \overline{Y} \right)^{2}$$

$$= \frac{1}{M^2 (N-1)} \sum_{i=1}^{N} (M_i \overline{Y_i} - M_i \overline{Y})^2$$

$$= \frac{1}{M^{2}(N-1)} \sum_{i=1}^{N} \left[ M_{i}^{2} \overline{Y}_{i}^{2} + M^{2} \overline{Y}^{2} - 2M \overline{Y}_{i} M_{i} \overline{Y}_{i} \right]$$

$$= \frac{1}{M^{2}(N-1)} \left[ \sum_{i=1}^{N} M_{i}^{2} \overline{Y_{i}}^{2} + N M^{2} \overline{Y}^{2} - 2 M \overline{Y} \sum_{i=1}^{N} M_{i} \overline{Y_{i}} \right]$$

$$= \frac{1}{M^{2}(N-1)} \left[ \sum_{i=1}^{N} M_{i}^{2} \overline{Y_{i}}^{2} + N M^{2} \overline{Y}^{2} - 2 M^{2} \overline{Y}^{2} \right]$$

$$= \frac{1}{M^2(N-1)} \left[ \sum_{i=1}^{N} M_i^2 \overline{Y}_i^2 - NM^2 \overline{Y}^2 \right]$$

$$= \frac{1}{M^2(N-1)} \left[ \sum_{i=1}^{N} M_i^2 \overline{Y}_i^2 - NM^2 \left( \frac{Y}{NM} \right)^2 \right]$$

$$= \frac{1}{M^{2}(N-1)} \cdot \left[ \sum_{i=1}^{N} M_{i}^{2} \overline{\gamma_{i}}^{2} - \frac{Y^{2}}{N} \right]$$

Need to estimate \sum\_{Mi^2\subseteq i^2}^N Mi^2\subseteq 2 \ \text{Separately}.

Estimation of 
$$Y^2$$
:  $E(\hat{Y}) = Y$ 

$$Var\left(\hat{Y}\right) = E\left(\hat{Y}^{2}\right) - \left\{E\left(\hat{Y}\right)\right\}^{2} = E\left(\hat{Y}^{2}\right) - Y^{2}$$

ie. 
$$Y^2 = E(\hat{Y}^2) - Var(\hat{Y})$$

Suppose, 4 be an unbiased estimator of voca (i) ie.

E(v) = vor (i)

$$Y^2 = E(\hat{Y}^2 - v)$$

$$\frac{Y^2}{N}$$
 is unbiasedly estimated by  $\frac{\hat{Y}^2 - v}{N}$ 

Estimation of IMi272  $\frac{1}{N}\sum_{i=1}^{N}M_i^2\bar{Y}_i^2$  is unbiasedly estimated by  $\frac{1}{n}\sum_{i=1}^{n}M_i^2\bar{Y}_i^2$ We have,  $E\left(\frac{1}{m_i}\sum_{i=1}^{m_i}y_{ij}\right) = Y_i$ ie. E (yī) = Yī , 1 = 1(1) n  $k = Var(\bar{y}_i) = \frac{Sw_i^2}{m!} (1-t_i) y_i^2 = 1(1)n$ :  $E(\overline{y_i}^2) = -\{E(\overline{y_i})\}^2 = \frac{S_{W_i}^2}{m_i} (1-\frac{1}{4})$ , i = 1(1) n  $E(\overline{y_i}^2) - \overline{Y_i}^2 = \frac{s_{n_i}^2}{m_i} (i-h)$  $\overline{Y_i}^2 = E(\overline{y_i}^2) - \frac{S_{N_i}^2}{m_i}(1-E)$  , i = 1(1)n = E ( \( \varphi^2 \) - E ( \( \frac{\lambda\_{W\_1}^2 \lambda\_{1}^2 + \text{\$\frac{1-\text{\$\frac{1}{2}}}{m\_1}} \right)  $\therefore \overline{Y_i}^2 = E \left[ \overline{y_i}^2 - \frac{\lambda_{w_i}^2}{m!} (i-\frac{1}{4}i) \right] , i=10n$  $\frac{A}{Y_i}^2 = \frac{1}{y_i^2} - \frac{2w_i}{m!} (i-\frac{1}{x_i})$  ,  $i = \frac{1}{x_i}$ hence I N Mi2 Fi 2 is unbiasedly estimated by  $\frac{1}{n} \sum_{i=1}^{n} M_{i}^{2} \left( \overline{y_{i}}^{2} - \frac{g_{w_{i}}^{2}}{m_{i}} (-h) \right) = \frac{1}{n} \sum_{i=1}^{n} M_{i}^{2} \overline{y_{i}}^{2} - \frac{1}{n} \sum_{i=1}^{n} \frac{M_{i}^{2} g_{w_{i}}^{2}}{m_{i}} (-h)$ :.  $8b^2 = \frac{1}{M^2(N+1)} \left[ \sum_{i=1}^{N} M_i^2 \overline{Y_i}^2 - \frac{y^2}{N} \right]$  is unbiasedly estimated by  $8b^{2} = \frac{1}{M^{2}(N-1)} \left[ \frac{N}{n} \sum_{i=1}^{n} M_{i}^{2} g_{i}^{2} - \frac{N}{n} \sum_{i=1}^{n} \frac{M_{i}^{2} \Delta h_{i}^{2}}{m_{i}} (i-\frac{1}{n}) - \frac{\hat{Y}^{2} - 1}{N} \right]$ :.  $Var(\hat{Y}) = N^2M^2 (1-\hat{Y}) = \frac{S_b^2}{n} + \frac{N}{n} = \frac{N}{m_i} \frac{M_i^2 S_{m_i}^2}{m_i} (1-\hat{Y})$ is unbiasedly estimated by  $v = \frac{N^2 M^2 (1-1)}{n} s_b^2 + \frac{N^2 N}{n^2} \frac{Mi^2 s_b^2}{mi} (i-1i)$ 

$$ie. \ \ U = \frac{N^2 M^2 (1-\frac{1}{4})}{n} \cdot \frac{1}{M^2 (N-1)} \left[ \frac{N}{n} \sum_{i=1}^{n} M_i^2 \bar{y}_i^2 - \frac{N}{n} \sum_{i=1}^{n} \frac{M_i^2 J_{Ni}^2}{m_i} (1-\frac{1}{4}i) - \frac{\hat{Y}^2 - u}{N} \right] + \left[ \frac{N^2}{n^2} \sum_{i=1}^{n} \frac{M_i^2 J_{Ni}^2}{m_i} (1-\frac{1}{4}i) \right]$$

$$\frac{N^{2}(1-\frac{1}{r})}{n(N-1)} \cdot \frac{N}{n} \sum_{i=1}^{n} M_{i}^{2} \overline{y}_{i}^{2} - \frac{N^{2}(1-\frac{1}{r})}{n(N-1)} \frac{N}{n} \sum_{i=1}^{n} \frac{M_{i}^{2} \lambda_{w_{i}}^{2}}{m_{i}} (1-\frac{1}{r}) + \left(-\frac{\hat{Y}^{2}}{N} + \frac{U}{N}\right) \frac{N^{2}(1-\frac{1}{r})}{n(N-1)} + \frac{N^{2}}{n^{2}} \sum_{i=1}^{n} \frac{M_{i}^{2} \lambda_{w_{i}}^{2}}{m_{i}} (1-\frac{1}{r})$$

$$\frac{9\left[1-\frac{N(1-\frac{1}{4})}{n(N-1)}\right]}{n(N-1)} = \frac{N^{2}(N-n)}{n^{2}(N-1)} \sum_{i=1}^{n} M_{i}^{2} \overline{y_{i}^{2}} - \frac{\hat{Y}^{2}}{N}, \frac{N^{2}(1-\frac{1}{4})}{n(N-1)} + \left(\frac{N^{2}}{n^{2}} \sum_{i=1}^{n} \frac{M_{i}^{2} \mathcal{S}_{w_{i}}^{2}}{m_{i}^{2}} (1-\frac{1}{4}i)\right) \left(1-\frac{N(1-\frac{1}{4})}{N-1}\right)$$

$$\frac{1}{n} \left[ \frac{mN - y'_{1} - N + y'_{1}}{n (N-1)} \right] = \frac{N^{2} (N-n)}{n^{2} (N-1)} \sum_{i=1}^{n} M_{i}^{2} \overline{y_{i}}^{2} - \hat{Y}^{2} \underbrace{N-n}_{n (N-1)} \right] + \left( \frac{N^{2}}{n^{2}} \sum_{i=1}^{n} \frac{M_{i}^{2} S_{w_{i}}^{2}}{m_{i}^{2}} (1-\frac{1}{n}) \right) \left( \frac{y'_{1} - y'_{1} + n}{N-1} \right)$$

$$\frac{N(n-1)}{n(N-1)} = \frac{N^{2}(N-n)}{n^{2}(N-1)} \sum_{i=1}^{n} M_{i}^{2} \overline{y_{i}^{2}} - \hat{Y}^{2} \left( \frac{N-n}{n(N-1)} \right) + \frac{N^{2}(n-1)}{n^{2}(N-1)} \sum_{i=1}^{n} \frac{M_{i}^{2} J_{M_{i}}^{2}}{m_{i}^{2}} (-\frac{1}{2}i)$$

$$0 = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^{n} M_{i} \overline{y_{i}^{2}} - Y^{2} \frac{(N-n)}{N(n-1)} + \frac{N}{n} \sum_{i=1}^{n} \frac{M_{i}^{2} S_{w_{i}}^{2}}{m_{i}^{2}} (1-\frac{1}{n})$$

$$Var(\hat{Y}) = U = \frac{M_{2}^{2}(\frac{1}{N})}{(\frac{N-\eta}{N-1})} \left[ \frac{N}{n} \sum_{i=1}^{N} M_{i}^{2} \overline{y}_{i}^{2} - \frac{\hat{Y}^{2}}{N} \right] + \frac{N}{\eta} \sum_{i=1}^{N} \frac{M_{i}^{2} J_{w_{i}}^{2}}{m_{i}} (1-1)$$

We minimize Vor(\$) subject to a given cost when Mi=M 4i=1(1) N and mi=m 4i=1(1)n.  $Var\left(\frac{\hat{Y}}{Y}\right) = \frac{Sb^2}{n}\left(1 - \frac{n}{N}\right) + \frac{1}{N}\sum_{i=1}^{N}\frac{Sw_i}{mn}\left(1 - \frac{m}{M}\right)$  $= \frac{Sb^2}{n} - \frac{Sb^2}{N} + \frac{Sw}{mn} \left(1 - \frac{m}{M}\right) \left[\text{where, } Sw = \frac{1}{N} \sum_{i=1}^{N} Sw_i^2\right]$  $= \frac{S_b^2}{n} - \frac{S_b^2}{N} + \frac{S_W^2}{mn} - \frac{S_W^2}{nM}$  $Var(\hat{Y}) = A_0 + \frac{A_1}{m} + \frac{A_2}{mn} (say)$ Where,  $A_0 = -\frac{S_b^2}{N}$ ,  $A_1 = S_b^2 - \frac{S_h^2}{M}$ ,  $A_2 = S_h^2$ Consider the following cost function C = a + Cin + Cimn a = overhead cost, a = cost of selecting a FSU, cz = per noit east of selecting a SSU We will minimize Var(\$\hat{\hat{\gamma}}) subject to C=Co Consider the Lagrangian function  $Z = Var(\widehat{Y}) + \lambda(c-c_0)$ where, d = the lagrangian multiplier.  $\therefore \ \ Z = \left(A_0 + \frac{A_1}{n} + \frac{A_2}{mn}\right) + \lambda \left(a + c_1 n + c_2 m n - c_0\right)$ Minimize 2 with respect to m and on,  $\frac{\partial 2}{\partial m} = 0 \Rightarrow -\frac{A_2}{m^2n} + \lambda c_2 n = 0 \Rightarrow \frac{A_2}{c_2 m^2} = \lambda m^2 - 0$  $\frac{1}{7} - \frac{A_1}{n^2} - \frac{A_2}{mn^2} + \lambda (C_1 + C_2 m) = 0$ 

9

?

-3

-3

1 ~

3

3

-3

7

$$\left(\begin{array}{c} A_{1} + \frac{A_{2}}{m} \right) = \lambda n^{2} \left(c_{1} + c_{2}m\right)$$

$$\Rightarrow \left(\begin{array}{c} A_{1} + \frac{A_{2}}{m} \right) = \frac{A_{2}}{c_{2}m^{2}} \left(c_{1} + c_{2}m\right) \left[\begin{array}{c} ming \\ \end{array}\right]$$

$$\Rightarrow A_{1} + \frac{A_{2}}{m} = \frac{A_{2} C_{1}}{c_{2}m^{2}} + \frac{A_{2}}{m}$$

$$\Rightarrow m^{2} = \frac{A_{2} C_{1}}{n}$$

ie. 
$$M = \frac{c_0 - a}{c_1 + c_2 m} = \frac{c_0 - a}{\frac{c_1 + c_2 \sqrt{\frac{A_2/c_2}{A_1/c_1}}}{\frac{A_1/c_1}{a_1/c_1}}}$$

$$m = \frac{c_0 - a}{c_1 + c_2 \sqrt{\frac{A_2/c_2}{A_1/c_1}}} = \left(\frac{c_0 - a}{\sqrt{A_1c_1} + \sqrt{A_2c_2}}\right) \sqrt{\frac{A_1}{c_1}}$$

Assignment > mi's are not equal, then what is the optional allocation.