PPS Sampling:

Under certain circumtances, selection of units with unequal probabilities provides more efficient estimator than equal probability sampling and this type of sampling is known as unequal or varying probability sampling. In the most commonly used varying probability sampling scheme the units are selected with probability proportional to a given measure of size, where the size measure is the value of an auxiliary variable X related to the characteristic Y under study and this sampling scheme is termed as ~~propr~~ probability proportional to size ~~sin~~ sampling (PPS). For instance the number of persons in some previous period may be taken as a measure of the size in sampling area units for a survey of socio-economic characters which are likely to be related to population. Similarly in estimating crop characteristics, the geographical area or cultivate area for a previous period may be considered as a measure of size. Similarly, in an industrial survey, the no. of workers may be taken as the size of an industrial establishment.

Since a large unit ie. an unit with a large value for the study variable Y contributes more to the population total than smaller units, it is natural to expect that a scheme of selection which gives more chance of inclusion in a sample to larger units than smaller units, would provide estimators more efficient than equal probability sampling. Such a scheme is provided by PPS sampling, size being the value of an auxiliary variable X ~~directed~~ directly related to Y. It may appear that such a selection procedure would give biased estimators as the larger units are over represented and the smaller units are under represented in the sample. This ~~no~~ would be so if the sample mean is used as an estimator of population mean ~~insted~~ instead if the sample observations are suitably weighted at the estimation stage taking into ~~considerat~~ consideration their ~~prop~~ probabilities of selection, it is possible to obtain unbiased estimators.

Let X_i and Y_i be the value of the size variable and the response variable or study variable for the ith population unit. A sample of size n is obtained in the form of small n units with or without replacement draws.

In case of 'with replacement' drawing in each draw the probability of selecting unit i, viz. P_i is proportional to the size X_i. A sample obtained in this way is called a PPSWR sample.

PPSWR

$$P_i \propto X_i$$

$$\Rightarrow P_i = K X_i \quad , i = 1(1)N$$

$$\Rightarrow \sum_{i=1}^{N} P_i = K \sum_{i=1}^{N} X_i$$

$$\Rightarrow 1 = K \sum_{i=1}^{N} X_i$$

$$\Rightarrow K = \frac{1}{\sum_{i=1}^{N} X_i} = \frac{1}{X}$$

$$\therefore \boxed{P_i = \frac{X_i}{X}} \; ; \; i = 1(1)N$$

Normalized size measure for the ith unit.

Clearly, $0 < P_i < 1$

$$\sum_{i=1}^{N} P_i = 1$$

~~PPSWOR~~

Unit	X_i	p_i
1	27	27/403
2	47	47/403
3	69	69/403
4	21	21/403
5	43	43/403 ✓
6	105	105/403
7	74	74/403
8	17	17/403

example of PPSWR

$$X = \sum_{i=1}^{8} X_i = 403$$

PPSWOR

Unit	X_i	p_i (given that 5th unit has been drawn in 1st draw)
1	27	27/360
2	47	47/360
3	69	69/360
4	21	21/360
6	105	105/360
7	74	74/360
8	17	17/360

PPSWOR gives that 5th unit has been drawn in 1st draw.

$$\sum_{\substack{i=1 \\ i \neq 5}}^{8} X_i = 360$$

Now, when PPSWOR is used, if the unit i is selected at 1st draw then the probability of any other unit being selected is

$$\frac{p_j}{1-p_i} \text{ (In 2nd draw)}$$

$$p_j = \frac{x_j}{403-43} = \frac{x_j}{360}$$

$$\frac{p_j}{1-p_i} = \frac{x_j/360}{1-(43/403)} =$$

Event	PPSWR	PPSWOR
1) Unit i in 1st draw	p_i	p_i
2) Unit j in 2nd draw / unit i in 1st draw	p_j	$\frac{p_j}{1-p_i}$
3) Unit i in 1st draw & unit j in 2nd draw	$p_i p_j$	$\frac{p_i p_j}{1-p_i}$ = P(unit i is selected in 1st draw) P(~~unit~~ unit j is selected in 2nd draw | unit i is selected in 1st draw) = $\frac{p_i p_j}{1-p_i}$

☐ Inclusion Probability in case of PPSWR (n):

$$\pi_i = \text{Prob}\{i\text{th unit is selected in atleast one draw}\}$$

$$= 1 - P\{i\text{th unit is not selected in any draw}\}$$

$$= 1 - (1-p_i)^n$$

$$\pi_{ij} = P(A_i \cap A_j) \quad [A_i : i\text{th unit selected in any draw}]$$

$$= P\{i\text{th unit is selected in } \text{~~all any~~ atleast one draw and } j\text{th unit is selected in atleast one draw}\}$$

$$= 1 - P(A_i \cap A_j)^c$$

$$= 1 - P(A_i^c \cup A_j^c)$$

been

)

Scanned with CamScanner

$$= 1 - P(A_i^c) - P(A_j^c) + P(A_i^c \cap A_j^c)$$

$$= 1 - (1-p_i)^n - (1-p_j)^n + (1-p_i-p_j)^n$$

$$\pi_i > \pi_j \implies 1-(1-p_i)^n > 1-(1-p_j)^n$$

$$\implies (1-p_i)^n < (1-p_j)^n$$

$$\implies 1-p_i < 1-p_j$$

$$\implies p_i > p_j$$

Inclusion Probability in case of PPSWOR (n):

π_i = Probability of including i in the whole sample.

= P (unit i is selected in draw 1) + P (unit i is selected in draw 2) + ... + P (unit i is selected in draw n)

$$= p_i + \sum_{j \neq i}^{N} p_j \frac{p_i}{(1-p_j)} + \sum_{j \neq k \neq i}^{N}\sum^{N} p_j \frac{p_k}{(1-p_j)} \frac{p_i}{(1-p_j-p_k)}$$

$$+ \dots + \sum^{N}\sum^{N} \dots \sum_{i_1 \neq i_2 \neq \dots \neq i_{n-1} \neq i}^{N} p_{i_1} \frac{p_{i_2}}{(1-p_{i_1})} \frac{p_{i_3}}{(1-p_{i_1}-p_{i_2})} \cdots \cdots \frac{p_i}{(1-p_{i_1}-p_{i_2}-\dots-p_{i_{n-1}})}$$

Special Case:

For $n=2$, $\quad \pi_i = p_i + \sum_{j \neq i}^{N} \frac{p_j p_i}{1-p_j}$

$$\pi_i > \pi_j$$

$$\Rightarrow \quad p_i + \sum_{k \neq i}^{N} p_k \frac{p_i}{1-p_k} > p_j + \sum_{k \neq j}^{N} p_k \frac{p_j}{1-p_k}$$

$$\Rightarrow \quad p_i \left[1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k} \right] > p_j \left[1 + \sum_{k \neq j}^{N} \frac{p_k}{1-p_k} \right]$$

$$\Rightarrow \quad \frac{p_i}{p_j} - 1 > \frac{1 + \sum_{k \neq j}^{N} \frac{p_k}{1-p_k}}{1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}} - 1$$

$$\Rightarrow \quad \frac{p_i - p_j}{p_j} > \frac{\sum_{k \neq j}^{N} \frac{p_k}{1-p_k} - \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}}{1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}}$$

$$= \cancel{\frac{\frac{p_i}{1-p_i} - p_j}{1+}}$$

$$= \frac{\frac{p_i}{1-p_i} - \frac{p_j}{1-p_j}}{1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}}$$

$$= \frac{p_i - p_i p_j - p_j + p_i p_j}{(1-p_i)(1-p_j)\left(1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}\right)}$$

$$= \frac{p_i - p_j}{(1-p_i)(1-p_j)\left(1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}\right)}$$

$$\Rightarrow \quad (p_i - p_j) \left[\frac{1}{p_j} - \frac{1}{(1-p_i)(1-p_j)\left(1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}\right)} \right] > 0$$

either $(p_i - p_j) > 0$ or $\frac{1}{p_j} - \frac{1}{(1-p_i)(1-p_j)\left(1 + \sum_{k \neq i}^{N} \frac{p_k}{1-p_k}\right)} > 0$

or both are negative.

Procedure of drawing PPS samples :

① Cumulative total method

Step-1

. Calculate the cumulative totals

$$T_1 = x_1$$

$$T_2 = x_1 + x_2$$

$$T_3 = x_1 + x_2 + x_3$$

$$\vdots$$

$$T_N = x_1 + x_2 + \dots + x_N$$

Step-2

Draw a random number R between 0 and T_N.

Step-3

If $T_{i-1} < R \leq T_i$, then select unit i in the sample.

Now,

$$P(\text{Selecting unit } i) = P(T_{i-1} < R \leq T_i)$$

$$= \frac{T_i - T_{i-1}}{T_N}$$

$$= \frac{x_i}{\sum_{i=1}^{N} x_i}$$

$$= P_i$$

Step-4

Repeating this process, we can obtain a sample of size n.

Note

- x_i s may take fractional values. In this case, they are multiplied by 10 or 100 to carry out the procedure.
- The procedure is difficult to implement if N is large

(II) LAHIRI'S METHOD

Step 1 Let, $M = \text{Max}\{x_1, x_2, \ldots, x_N\}$

Step 2 Draw a random number R from $\{1, 2, \ldots, M\}$

Step 3 Draw another random number i from $\{1, 2, \ldots, N\}$

Step 4 If $R \le x_i$, select unit i, else repeat step 2 & 3 of this procedure.

Result

In Lahiri's method, the probability of selecting the ith unit in the 1st effective draw is $\dfrac{x_i}{\sum_{i=1}^{N} x_i} = P_i$

Proof» A draw will be called ineffective if the 2nd random ~~num~~ number is i and the 1st random number R is such that $R > x_i$

Let, $\lambda = P$ (A draw will be effective)

$$= \sum_{i=1}^{N} P(\text{2nd random number is } i)\; P(R \le x_i)$$

$$= \sum_{i=1}^{N} \frac{1}{N} \cdot \frac{x_i}{M}$$

$$= \frac{\bar{X}}{M}$$

P(Unit i will be selected at the 1st effective draw)

= P(1st draw is effective & unit i is selected in 1st draw)

+ P(1st draw is ineffective, 2nd draw is effective & unit i is selected in 2nd draw) + P(1st and 2nd draw are ineffective, 3rd draw is effective & unit i is selected in 3rd draw) + ...

At any particular draw,

$$P(\text{selecting unit } i) = P(\text{2nd random no. is } i)\, P(R \leq x_i)$$

$$= \frac{1}{N} \cdot \frac{x_i}{M}, \quad i = 1(1)N$$

$\therefore$ P(Unit i will be selected at the 1st effective draw)

$$= \frac{x_i}{NM} + (1-\lambda) \frac{x_i}{NM} + (1-\lambda)^2 \frac{x_i}{NM} + \cdots$$

$$= \frac{x_i}{MN} \left[1 + (1-\lambda) + (1-\lambda)^2 + \cdots \right]$$

$$= \frac{x_i}{MN} \cdot \frac{1}{1-(1-\lambda)} \quad [\text{provided, } 0 \leq \lambda \leq 1]$$

$$= \frac{x_i}{MN\lambda} = \frac{x_i}{MN \cdot \frac{\bar{X}}{M}} = \frac{x_i}{N\bar{X}}$$

$$= \frac{x_i}{\sum_{i=1}^{N} x_i}$$

$$= P_i$$

Unbiased Estimator of Population ~~Unit~~ Total in PPSWR

Result: Let, y_i and p_i denote respectively the value of the response and the corresponding normed-size measure for the ith drawing in PPSWR $(i = 1(1)n)$. Then an unbiased estimator of the population total is given by $t_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$ (It is called Hansen and Hurwitz's estimator).

Proof» $\frac{y_i}{p_i}$ s are independent as the drawings are independent.

Each $\frac{y_i}{p_i}$ $(i=1(1)n)$ can take any of the values $\frac{Y_\alpha}{P_\alpha}$ $(\alpha = 1(1)N)$ with corresponding probability P_α.

$$E_p\left(\frac{y_i}{p_i}\right) = \sum_{\alpha=1}^{N}\frac{Y_\alpha}{P_\alpha}\cdot P_\alpha = \sum_{\alpha=1}^{N} Y_\alpha = Y \text{, for each } i = 1(1)n$$

$$\therefore \quad E_p(t_{HH}) = E_p\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}\right) \cancel{= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{p_i}\right)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E_p\left(\frac{y_i}{p_i}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y$$

$$= Y$$

$\therefore$ t_{HH} is unbiased for the population total.

Variance of the estimator

$$V_p(t_{HH}) = \frac{1}{n^2}\sum_{i=1}^{n} V_p\left(\frac{y_i}{p_i}\right)$$ [$\because$ The Covariance terms are zero as the drawings are independent]

$$V_p\left(\frac{y_i}{p_i}\right) = E_p\left(\frac{y_i^2}{p_i^2}\right) - \left[E_p\left(\frac{y_i}{p_i}\right)\right]^2$$

$$= \sum_{\alpha=1}^{N}\frac{Y_\alpha^2}{P_\alpha^2}\cdot P_\alpha - Y^2$$

$$= \sum_{\alpha=1}^{N}\frac{Y_\alpha^2}{P_\alpha} - Y^2 = V \text{ (say)}$$

$$\therefore\ V_p(t_{HH}) = \frac{V}{n}$$

Alternative forms of V

① $$V = \sum_{\alpha=1}^{N} \left(\frac{Y_\alpha}{P_\alpha} - Y\right)^2 P_\alpha$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha^2} \cdot P_\alpha - 2 \cdot Y \sum_{\alpha=1}^{N} \frac{Y_\alpha}{P_\alpha} \cdot P_\alpha + Y^2 \sum_{\alpha=1}^{N} P_\alpha$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha} - 2Y^2 + Y^2 \qquad \left[\because\ Y = \sum_{\alpha=1}^{N} Y_\alpha \text{ and } \sum_{\alpha=1}^{N} P_\alpha = 1\right]$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha} - Y^2$$

Hence, proved.

② $$V = \sum_{\alpha<\beta=1}^{N}\sum P_\alpha P_\beta \left(\frac{Y_\alpha}{P_\alpha} - \frac{Y_\beta}{P_\beta}\right)^2$$

$$= \sum_{\alpha<\beta=1}^{N}\sum P_\alpha P_\beta \frac{Y_\alpha^2}{P_\alpha^2} + \sum_{\alpha<\beta=1}^{N}\sum P_\alpha P_\beta \frac{Y_\beta^2}{P_\beta^2}$$

$$- 2 \sum_{\alpha<\beta=1}^{N}\sum P_\alpha P_\beta \frac{Y_\alpha}{P_\alpha} \cdot \frac{Y_\beta}{P_\beta}$$

$$\cancel{= \sum_{\alpha=1}^{N}\sum_{\beta>\alpha+1}^{N}} = \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha} \sum_{\beta>\alpha} P_\beta + \sum_{\beta=1}^{N} \frac{Y_\beta^2}{P_\beta} \sum_{\alpha<\beta} P_\alpha$$

$$- 2 \sum_{\alpha<\beta}\sum Y_\alpha Y_\beta$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha}\left(\sum_{\beta>\alpha} P_\beta\right) + \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha}\left(\sum_{\beta<\alpha} P_\beta\right) - \sum_{\alpha\neq\beta}\sum Y_\alpha Y_\beta$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha}\left[\sum_{\beta>\alpha} P_\beta + \sum_{\beta<\alpha} P_\beta\right] - \left[\sum_{\alpha=1}^{N} Y_\alpha \cdot \sum_{\beta=1}^{N} Y_\beta - \sum_{\alpha=1}^{N} Y_\alpha^2\right]$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha}[1-P_\alpha] - \left[\left(\sum_{\alpha=1}^{N} Y_\alpha\right)^2 - \sum_{\alpha=1}^{N} Y_\alpha^2\right]$$

$$\left[\because \sum_{\beta=1}^{N} P_\beta = 1 \Rightarrow \sum_{\beta<\alpha} P_\beta + \sum_{\beta>\alpha} P_\beta = 1 - P_\alpha\right]$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha} - \cancel{\sum_{\alpha=1}^{N} Y_\alpha^2} - Y^2 + \cancel{\sum_{\alpha=1}^{N} Y_\alpha^2}$$

$$= \sum_{\alpha=1}^{N} \frac{Y_\alpha^2}{P_\alpha} - Y^2$$

Hence proved.

If $Y_i \propto P_i$

Then $Y_i^2 = K^2 P_i^2 \quad (i = 1(1)N)$

$$\frac{Y_i^2}{P_i} = K^2 P_i$$

$$\& \sum_{i=1}^{N} \frac{Y_i^2}{P_i} = K^2 \quad \left[\because \sum_{i=1}^{N} P_i = 1\right]$$

$$Y = \sum_{i=1}^{N} Y_i = K \qquad [\text{Do}]$$

$$\therefore\ V = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2 = K^2 - K^2 = 0$$

If Y_i's are roughly proportional to P_i's then the ~~varian~~ $Var_p(t_{HH})$ is small. But $P_i = \frac{X_i}{\sum_{i=1}^{N} X_i}$, so $Y_i \propto P_i \Rightarrow Y_i \propto X_i$, i.e. if the regression of Y on X is linear, then PPS sampling works best, or it is a proper sampling.

$$\left[\sum_{\alpha=1} Y_\alpha^2\right]$$

Unbiased estimation of $Var(t_{HH})$

$$Var(t_{HH}) = \frac{V}{n} \quad \text{where} \quad V = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2$$

Lemma:

Let y_r :- value of the unit selected at r^{th} drawing.

Let $r \neq s$, then $E_p\left[\frac{y_r}{p_r} - \frac{y_s}{p_s}\right]^2 = 2V$

Proof: $E_p\left[\frac{y_r}{p_r} - \frac{y_s}{p_s}\right]^2$

$$= E_p\left[\left(\frac{y_r}{p_r} - Y\right) - \left(\frac{y_s}{p_s} - Y\right)\right]^2$$

$$= E_p\left[\frac{y_r}{p_r} - Y\right]^2 + E_p\left[\frac{y_s}{p_s} - Y\right]^2$$
$$-2E_p\left[\frac{y_r}{p_r} - Y\right]\left[\frac{y_s}{p_s} - Y\right]$$

Already proved that $E_p\left(\frac{y_r}{p_r}\right) = Y$

$$\& \quad V_p\left(\frac{y_r}{p_r}\right) = E_p\left(\frac{y_r}{p_r} - Y\right)^2 = V$$

$\forall r = 1(1)n$

Also,

$$E_p\left(\frac{y_r}{p_r} - Y\right)\left(\frac{y_s}{p_s} - Y\right) = Cov\left(\frac{y_r}{p_r}, \frac{y_s}{p_s}\right) = 0$$

because of the independence of $\frac{y_r}{p_r}$ & $\frac{y_s}{p_s}$.

Hence, $E_p\left[\frac{y_r}{p_r} - \frac{y_s}{p_s}\right]^2 = V + V = 2V$

Formation of unbiased estimator

The previous result is true for all $r \neq s \in \{1, 2, \ldots, n\}$

Total number of pairs (r, s) $(r \neq s)$ can be formed is $n(n-1)$

Hence, $E_p\left[\frac{1}{n(n-1)} \sum_{r\neq s=1}^{n}\sum^{n} \left(\frac{y_r}{p_r} - \frac{y_s}{p_s}\right)^2\right] = 2V$

i.e. $E_p\left[\frac{1}{2n^2(n-1)} \sum_{r\neq s=1}^{n}\sum^{n} \left(\frac{y_r}{p_r} - \frac{y_s}{p_s}\right)^2\right] = \frac{V}{n} = Var(t_{HH})$

$\therefore \widehat{Var}(t_{HH}) = \frac{1}{2n^2(n-1)} \sum_{r\neq s}^{n}\sum^{n} \left(\frac{y_r}{p_r} - \frac{y_s}{p_s}\right)^2$

<u>Result</u> (An alternative estimator)

$$E\left[\frac{1}{n(n-1)} \sum_{r=1}^{n} \left(\frac{y_r}{p_r} - t_{HH}\right)^2\right] = \frac{V}{n} = Var(t_{HH})$$

i.e. $\frac{1}{n(n-1)} \sum_{r=1}^{n} \left(\frac{y_r}{p_r} - t_{HH}\right)^2$ is unbiased for $Var(t_{HH})$.

<u>Proof</u>: $V = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2$

Let, v be an unbiased estimator of $Var(t_{HH})$

$$\therefore E(v) = \frac{1}{n}\left[\sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2\right] = Var(t_{HH})$$

$$= E_p(t_{HH}^2) - E_p^2(t_{HH})$$

$$= E_p(t_{HH}^2) - Y^2$$

$$\therefore Y^2 = E_p(t_{HH}^2 - v) \quad \text{—(1)}$$

Again, $E_p\left(\frac{y_r^2}{p_r^2}\right) = \sum_{i=1}^{N} \frac{Y_i^2}{P_i^2} \cdot P_i = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} \quad (\forall r = 1(1)n)$

Hence, $E_p\left[\frac{1}{n} \sum_{r=1}^{n} \frac{y_r^2}{p_r^2}\right] = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} \quad$ —(2)

$$\therefore E_p\left[\frac{1}{n} \sum_{r=1}^{n} \frac{y_r^2}{p_r^2} - (t_{HH}^2 - v)\right] = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2 = V$$

Scanned with CamScanner

Thus, $\frac{V}{n}$ is unbiasedly estimated by $\frac{1}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r^2} - \frac{(t_{HH}^2 - v)}{n}$

But, we assume ~~that~~ v to be the u.e. of $Var(t_{HH}) = \frac{V}{n}$

$$\therefore \quad v = \frac{1}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r^2} - \frac{(t_{HH}^2 - v)}{n}$$

$$\text{i.e.} \quad v\left(1-\frac{1}{n}\right) = \frac{1}{n}\left[\frac{1}{n}\sum_{r=1}^{n}\frac{y_r^2}{p_r^2} - t_{HH}^2\right]$$

$$\text{i.e.} \quad v = \frac{1}{n(n-1)}\sum_{r=1}^{n}\left(\frac{y_r}{p_r} - t_{HH}\right)^2 \quad \left[\because t_{HH} = \frac{1}{n}\sum_{r=1}^{n}\frac{y_r}{p_r}\right]$$

Estimated Gain in PPSWR compared to SRSWR:

Under SRSWR, Y is unbiasedly estimated by $\hat{Y} = N\bar{y}$ &

$$Var(\hat{Y}) = N^2 Var(\bar{y}) = N^2\frac{\sigma^2}{n}$$

$$= \frac{N^2}{n}\cdot\left[\frac{1}{N}\sum_{i=1}^{N}Y_i^2 - \bar{Y}^2\right]$$

$$= \frac{N}{n}\sum_{i=1}^{N}Y_i^2 - \frac{Y^2}{n} \quad [\because N\bar{Y} = Y]$$

$$= V^* \text{ (say)}$$

Under PPSWR, Y is unbiasedly estimated by t_{HH}

$$\text{Gain} = Var(\hat{Y}) - Var(t_{HH})$$

$$= V^* - Var(t_{HH})$$

Thus, the estimated gain $= \hat{V}^* - \widehat{Var}(t_{HH})$

(Both expressions are estimated under PPSWR)

Already obtained that, under PPSWR,

$\widehat{Var}(t_{HH}) = v$.

Need to find $\hat{V}^*$ under PPSWR.

Under PPSWR, Y^2 is unbiasedly estimated by $t_{HH}^2 - v$

& $E_p\left[\frac{y_r^2}{p_r}\right] = \sum_{i=1}^{N} \frac{Y_i^2}{P_i} \cdot P_i = \sum_{i=1}^{N} Y_i^2 \quad (\forall r = 1(1)n)$

Hence, $\sum_{i=1}^{N} Y_i^2$ is unbiasedly estimated by $\frac{1}{n}\sum_{r=1}^{n}\frac{y_r^2}{p_r}$, under PPSW

$\therefore$ V^* is unbiasedly estimated by $\frac{N}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \frac{(t_{HH}^2 - v)}{n}$,

under PPSWR.

$\therefore$ Estimated gain $= \hat{V}^* - v$

$$= \frac{N}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \frac{(t_{HH}^2 - v)}{n} - v$$

$$= \frac{N}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \frac{t_{HH}^2}{n} - \left(\frac{n-1}{n}\right)v$$

$$= \frac{N}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \frac{t_{HH}^2}{n} \cancel{\frac{(n-1)}{n}\frac{1}{(n-1)}\left[\frac{1}{n}\sum_{r=1}^{n}\frac{y_r^2}{p_r^2}\right.}$$

$$- \frac{(\cancel{n}-1)}{n} \cdot \frac{1}{n(\cancel{n}-1)}\sum_{r=1}^{n}\left(\frac{y_r}{p_r} - t_{HH}\right)^2$$

$$= \frac{N}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \cancel{\frac{t_{HH}^2}{n}} - \frac{1}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r^2} + \cancel{\frac{t_{HH}^2}{n}}$$

$$= \frac{1}{n^2}\left[N\sum_{r=1}^{n}\frac{y_r^2}{p_r} - \sum_{r=1}^{n}\frac{y_r^2}{p_r^2}\right]$$

Estimated gain $= \frac{1}{n^2}\sum_{r=1}^{n}\frac{y_r^2}{p_r}\left(N - \frac{1}{p_r}\right)$

Estimator of Population Total under PPSWOR

Desh Raj's Estimator

Define,
$$t_1 = \frac{y_1}{p_1}$$

$$t_2 = y_1 + \frac{y_2}{p_2/(1-p_1)}$$

$$t_3 = y_1 + y_2 + \frac{y_3}{p_3/(1-p_1-p_2)}$$

$$\vdots$$

$$t_n = y_1 + y_2 + \dots + y_{n-1} + \frac{y_n}{p_n/(1-p_1-p_2-\dots-p_{n-1})}$$

Desh Raj's ordered estimator is defined as
$$t_D = \frac{1}{n}\sum_{i=1}^{n} t_i$$

Result ⇒ Under PPSWOR, t_D is an unbiased estimator of Y and its variance can be estimated unbiasedly by
$$v = \frac{1}{n(n-1)}\sum_{i=1}^{n}(t_i - t_D)^2$$

Proof: $$E_p(t_D) = E_p\left[\frac{1}{n}\sum_{i=1}^{n} t_i\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E_p(t_i)$$

$$E_p(t_1) = E_p\left(\frac{y_1}{p_1}\right) \cancel{= \sum_{\alpha=1}^{N}\frac{Y_\alpha}{P_\alpha}\,P(y_1 = Y_\alpha)}$$

$$= \sum_{\alpha=1}^{N}\frac{Y_\alpha}{P_\alpha}\,P(y_1 = Y_\alpha)$$

$$= \sum_{\alpha=1}^{N}\frac{Y_\alpha}{P_\alpha}\cdot P_\alpha$$

$$= \sum_{\alpha=1}^{N} Y_\alpha = Y$$

$$E_p(t_2) = E_1 E_2 \left[t_2 \mid y_1 = Y_\alpha \right]$$

Here, E_2 denotes the conditional expectation of y_2 given y_1 & E_1 denotes the unconditional expectation over y_1.

$$E_2 \left[t_2 \mid y_1 = Y_\alpha \right] = E_2 \left[y_1 + \frac{y_2}{p_2/(1-p_1)} \,\Big|\, y_1 = Y_\alpha \right]$$

$$= E_2 \left[Y_\alpha + \frac{y_2}{p_2/(1-p_1)} \,\Big|\, y_1 = Y_\alpha \right]$$

$$= Y_\alpha + E_2 \left[\frac{y_2}{p_2/(1-p_1)} \,\Big|\, y_1 = Y_\alpha \right]$$

$$= Y_\alpha + \sum_{\beta \neq \alpha} \frac{Y_\beta}{P_\beta/(1-P_\alpha)} \cdot \frac{P_\beta}{(1-P_\alpha)}$$

$$= Y_\alpha + \sum_{\beta \neq \alpha} Y_\beta$$

$$= Y$$

$$\therefore \quad E_p(t_2) = E_1(Y) = Y.$$

$$E_p(t_r) = E_1 E_2 \left[t_r \mid y_1 = Y_{\alpha_1}, y_2 = Y_{\alpha_2}, \ldots, y_{r-1} = Y_{\alpha_{r-1}} \right]$$

Here, E_2 : conditional expectation of y_r given $y_1, \ldots, y_{r-1}$

& E_1 : unconditional expectation over $y_1, \ldots, y_{r-1}$ (after fixing the units with labels $(\alpha_1, \ldots, \alpha_{r-1})$ at the first $(r-1)$ drawings)

$$E_2 \left[t_r \mid y_1 = Y_{\alpha_1}, y_2 = Y_{\alpha_2}, \ldots, y_{r-1} = Y_{\alpha_{r-1}} \right]$$

$$= E_2 \left[(y_1 + y_2 + \ldots + y_{r-1}) + \frac{y_r}{p_r/(1-p_1-\cdots-p_{r-1})} \,\Big|\, y_1 = Y_{\alpha_1}, \ldots, y_{r-1} = Y_{\alpha_{r-1}} \right]$$

$$= (Y_{\alpha_1} + Y_{\alpha_2} + \ldots + Y_{\alpha_{r-1}}) + E_2 \left[\frac{y_r}{p_r/(1-p_1-\cdots-p_{r-1})} \,\Big|\, y_1 = Y_{\alpha_1}, \ldots, y_{r-1} = Y_{\alpha_{r-1}} \right]$$

$$= \sum_{i=1}^{r-1} Y_{\alpha_i} + \sum_{\beta \neq \{\alpha_1, \ldots, \alpha_{r-1}\}} \frac{Y_\beta}{P_\beta/(1-P_{\alpha_1}-\cdots-P_{\alpha_{r-1}})} \cdot \frac{P_\beta}{(1-P_{\alpha_1}-\cdots-P_{\alpha_{r-1}})}$$

$$= \sum^{r-1} Y_{\alpha_i} + \sum Y_\beta = Y$$

$\therefore \quad E_p(t_r) = E_1(Y) = Y.$

Thus, $E_p(t_D) = \frac{1}{n}\sum_{i=1}^{n} E_p(t_i) = Y$

i.e. t_D is unbiased for Y.

Variance of t_D:

Result: $Var(t_D) = \frac{1}{n^2}\sum_{i=1}^{n} Var(t_i)$

Proof:

$t_D = \frac{1}{n}\sum_{i=1}^{n} t_i$

$\therefore Var(t_D) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} t_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(t_i) + 2\sum\sum_{i<j} Cov(t_i, t_j)\right]$

We have to prove, $Cov(t_i, t_j) = 0$ for $i<j$

$Cov(t_i, t_j) = E_p(t_i t_j) - E_p(t_i)\, E_p(t_j)$

Now, $E_p(t_i t_j) = E_1 E_2 (t_i t_j \mid t_i)$

Given, $t_i \equiv$ Given $y_1, y_2, \dots, y_i$

$E_p[t_i t_j] = E_1[t_i\, E_2(t_j \mid t_i)] = E_1[t_i\, E_2(t_j \mid y_1, \dots, y_i)]$

$= E_1[t_i Y]$, Since, $E_2(t_j \mid y_1, \dots, y_i) = Y$

$= Y\, E_1[t_i] = Y^2$

$\therefore Cov(t_i, t_j) = Y^2 - Y \cdot Y = 0$, for $i<j$.

$\therefore \quad Var(t_D) = \frac{1}{n^2}\sum_{i=1}^{n} Var(t_i)$

Result: $\frac{1}{2n^2(n-1)}\sum\sum_{i\neq j}(t_i - t_j)^2$ is an u.e. of $Var(t_D)$.

Proof: We start with $E_p(t_i - t_j)^2$ for $i<j$

$E_p(t_i - t_j)^2 = E_p[(t_i - Y) - (t_j - Y)]^2 = E_p(t_i - Y)^2 + E_p(t_j - Y)^2 - 2E_p(t_i - Y)(t_j - Y)$

$= V_p(t_i) + V_p(t_j) - 2\,Cov_p(t_i, t_j)$

$= V_p(t_i) + V_p(t_j) \quad [\because$ for $i<j$, $Cov_p(t_i, t_j) = 0]$

Similarly for $i>j$ also, we can show,

$$E_p(t_i-t_j)^2 = V_p(t_i)+V_p(t_j)$$

Thus, for $i\neq j$, $E_p(t_i-t_j)^2 = V_p(t_i)+V_p(t_j)$

$$\therefore E_p\left[\sum\sum_{i\neq j}(t_i-t_j)^2\right] = \sum\sum_{i\neq j}V_p(t_i) + \sum\sum_{i\neq j}V_p(t_j)$$

$$= (n-1)\sum_{i=1}^{n}V_p(t_i) + (n-1)\sum_{j=1}^{n}V_p(t_j)$$

$$= 2(n-1)\sum_{i=1}^{n}V_p(t_i)$$

$$\therefore E_p\left[\frac{1}{2n^2(n-1)}\sum\sum_{i\neq j}(t_i-t_j)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}V_p(t_i) = Var(t_D)$$

Drawback of the estimator and its improvement :

- Depends on the order.
- Given, an estimator $t(s, \underset{\sim}{Y})$, we can use the improved estimator

$$t^*(s^*, \underset{\sim}{Y}) = \frac{\sum_{s\sim s^*} p(s)\, t(s, \underset{\sim}{Y})}{\sum_{s\sim s^*} p(s)}$$ such that $E_p(t) = E_p(t^*)$ & $V_p(t^*) \leq V_p(t)$

Suppose $n=2$

$s_1 = \{i_1, i_2\}$, $t_D(s_1) = \frac{1}{2}\left[\frac{y_{i_1}}{p_{i_1}} + y_{i_1} + \frac{y_{i_2}}{p_{i_2}/(1-p_{i_1})}\right]$

$s_2 = \{i_2, i_1\}$, $t_D(s_2) = \frac{1}{2}\left[\frac{y_{i_2}}{p_{i_2}} + y_{i_2} + \frac{y_{i_1}}{p_{i_1}/(1-p_{i_2})}\right]$

$s_1, s_2 \sim s^* = \{i_1, i_2\}$

The improved estimator is $t_D^* = \dfrac{p(s_1)\,t_D(s_1) + p(s_2)\,t_D(s_2)}{p(s_1)+p(s_2)}$

[symmetrized Desh Raj's estimator]

Murthy's Estimator

In PPSWOR, P[selecting unit i in draw 1] $= P_i$ & P[selecting the sample s] $= p(s)$

$p(s) = \sum_{i=1}^{N} P(\text{unit } i \text{ in draw } 1)\, P(\text{selecting the sample } s \mid \text{unit } i \text{ in draw } 1)$

$= \sum_{i=1}^{N} P_i\, p(s|i)$, where $p(s|i) =$ [Note that $\sum_{s\ni i} p(s|i) = 1$]

Murthy's estimator is defined as $t_M = \frac{1}{p(s)}\sum_{i\in s} y_i\, p(s|i)$

<u>Result</u> :

$$E_p(t_M) = Y$$

<u>Proof</u>: $E_p(t_M) = \sum_{s} p(s)\, t_M(s, \underset{\sim}{Y})$

$$= \sum_{s} \sum_{i \in s} Y_i\, p(s|i)$$

$$= \sum_{i=1}^{N} Y_i \sum_{s \ni i} p(s|i)$$

$$= \sum_{i=1}^{N} Y_i \quad \left[\because \sum_{s \ni i} p(s|i) = 1\right]$$

$$= Y.$$

Scanned with CamScanner

Rao - Hartley and Cochran's Sampling Strategy:

- Divide the entire population of size N by n subpopulations.

→ First select N_1 individuals from the total of N individuals by ~~simple~~ SRSWOR..

→ Then select N_2 units from the remaining $(N-N_1)$ units.

→ Then select N_3 units from the remaining $(N-N_1-N_2)$ units and so on.

→ Continue till n groups are formed.

- Finally select one unit from each group by PPS.

Illustration >> MSc - 2nd Sem Batch

$N=18 \quad n=4$

$N_1 = N_2 = 5 \quad , N_3 = N_4 = 4$

1st step 5 from 18 by SRSWOR.

2nd step 5 from 13 by SRSWOR.

3rd step 4 from 8 by SRSWOR.

Formation of Group 1

$\boxed{1-90}$ R.N. 00, 91-99 rejected

R.N.	Unit selected	Member
	R mod (18)	
26	8	Ritom
90	00/18	Sweety
56	02	Ankita
17	17	Susmita KV
76	4	Ipsita

2nd Group	1st group	
1	1	Abhishek
	2	Ankita
2	3	Debayan
	4	Ipsita
3	5	Krishna
4	6	Pitush
5	7	Ramesh
	8	Ritom
6	9	Sanchary
7	10	~~Satrendu~~ Saharupa
8	11	Satrendu
9	12	Sougata
10	13	Sontik
11	14	Sritama
12	15	Sudeshna
	16	Susmita B
	17	Susmita KV
	18	Sweety

Formation of Group 2

2 digit R.N. 1–91, (00, 92–99 Rejected)

R.N.	Unit Selected R mod (13)	Member
73	8	Satrendu
16	3	Krishna
28	2	Debayan
39	00/13	Susmita B
85	7	Satarupa

1 Abhishek
2 Debayan
3 Krishna
4 Pitush
5 Ramesh
6 Sanchary
7 Satarupa
8 Satrendu
9 Songata
10 Sontik
11 Sritama
12 Sudeshna
13 Susmita B

Formation of Group 3

$\boxed{1-96}$ R.N. 00, 97 – 99 Rejected

R.N.	Unit selected R mod (8)	Member
50	2	Pitush
96	00/8	Sudeshna
43	3	Ramesh
33	1	Abhishek

1 Abhishek
2 Pitush
3 Ramesh
4 Sanchary
5 Songata
6 Sontik
7 Sritama
8 Sudeshna

Group-4

{ Sanchary, Songata, Sontik, Sritama }

Notation >>

$U = \{1, 2, 3, \ldots, N\} \rightarrow$ Entire Population

$\underset{\sim}{Y} = \{Y_1, Y_2, \ldots, Y_N\} \rightarrow$ set of values of the study variable

$\underset{\sim}{P} = \{P_1, P_2, \ldots, P_N\}$ such that $0 < P_\alpha < 1 \quad \forall \alpha = 1(1)N$ &

$$\sum_{\alpha=1}^{N} P_\alpha = 1$$

→ Normed size measures.

y_{ij} :- value of the study variable for the j^{th} observation within i^{th} group $\left\{ \begin{array}{l} i = 1(1)n \\ j = 1(1)N_i \end{array} \right\}$

p_{ij} :- ~~normed~~ size measure of the j^{th} observation within i^{th} group.

$$Q_i = \sum_{j=1}^{N_i} p_{ij}$$

p_{ij} / Q_i = Normed size measure of the j^{th} observation within i^{th} group.

Illustration

Individual	P_i	Individual	P_i
1	0·08	10	0·06
2	0·12	11	0·01
3	0·02	12	0·1
4	0·1	13	0·05
5	0·1	14	0·05
6	0·02	15	0·02
7	0·05	16	0·05
8	0·10	17	0·02
9	0·04	18	0·01

Group-I $\{8, 18, 2, 17, 4\}$

$p_{11} = 0·1$, $p_{12} = 0·01$, $p_{13} = 0·12$, $p_{14} = 0·02$, $p_{15} = 0·10$

$Q_1 = 0·35$

Normed size measure group-I : $\left\{ \frac{10}{35}, \frac{1}{35}, \frac{12}{35}, \frac{2}{35}, \frac{10}{35} \right\}$

Concentrate ~~of~~ on Group - i

Let, y_i be the observation taken from group i by PPS.

Then. $\hat{Y}_i = \frac{y_i}{p_i/Q_i}$ unbiasedly estimate the total of i^{th} group ie. $Y_i = \sum_{j=1}^{N_i} Y_{ij}$

$$E(\hat{Y}_i) = \sum_{j=1}^{N_i} \frac{y_{ij}}{p_{ij}/Q_i} \cdot \frac{p_{ij}}{Q_i} = \sum_{j=1}^{N_i} y_{ij} = Y_i$$

Thus, $\left(\sum_{i=1}^{n} \hat{Y}_i\right) = \sum_{i=1}^{n} \frac{y_i}{p_i/Q_i}$ will unbiasedly estimate $\sum_{i=1}^{n} Y_i = Y$.

Thus, our estimator of population total is

$$t_{RHC} = \sum_{i=1}^{n} \frac{y_i}{p_i/Q_i} = \sum_n \frac{y_i}{p_i/Q_i}$$ ($\sum_n$ denotes the sum over all groups).

Expectation & Variance of t_{RHC}

Notation:

E_1/V_1 :- Expectation / Variance operator over the formation of groups.

E_2/V_2 :- Expectation/ Variance operator for sampling units ~~within~~ within the groups.

$$E(t_{RHC}) = E_1 E_2 (t_{RHC}) = E_1 E_2 \left[\sum_n \frac{y_i}{p_i/Q_i} \right]$$

$$= E_1 \left[\sum_n E_2 \left(\frac{y_i}{p_i/Q_i} \right) \right]$$

$$= E_1 \left[\sum_n \left[\sum_{j=1}^{N_i} \frac{y_{ij}}{p_{ij}/Q_i} \cdot \frac{p_{ij}}{Q_i} \right] \right]$$

$$= E_1 \left[\sum_n \left[\sum_{j=1}^{N_i} y_{ij} \right] \right] = E_1 \left[\sum_n Y_i \right]$$

$$= E_1 [Y] = Y.$$

$$Var(t_{RHC}) = E_1 V_2 (t_{RHC}) + V_1 E_2 (t_{RHC})$$

Since, $E_2(t_{RHC}) = Y$

$$\therefore V_1 E_2 (t_{RHC}) = V_1(Y) = 0$$

$$V_2(t_{RHC}) = V_2\left(\sum_n \frac{y_i}{p_i/Q_i}\right)$$

$$= \sum_n V_2\left(\frac{y_i}{p_i/Q_i}\right) \quad [\text{Since, } y_i\text{'s are independent}]$$

$$= \sum_n \left[\sum_{j_1<j_2=1}^{N_i}\sum \frac{p_{ij_1}}{Q_i} \cdot \frac{p_{ij_2}}{Q_i} \left(\frac{y_{ij_1}}{p_{ij_1}/Q_i} - \frac{y_{ij_2}}{p_{ij_2}/Q_i}\right)^2 \right]$$

$$= \sum_n \left[\sum_{j_1<j_2=1}^{N_i}\sum p_{ij_1} p_{ij_2} \left(\frac{y_{ij_1}}{p_{ij_1}} - \frac{y_{ij_2}}{p_{ij_2}}\right)^2 \right]$$

$$E_1 V_2 (t_{RHC}) = E_1 \left[\sum_n \left[\sum_{j_1<j_2=1}^{N_i}\sum p_{ij_1} p_{ij_2} \left(\frac{y_{ij_1}}{p_{ij_1}} - \frac{y_{ij_2}}{p_{ij_2}}\right)^2 \right] \right]$$

$$= \sum_n E_1 \left[\sum_{j_1<j_2}^{N_i}\sum^{N_i} p_{ij_1} p_{ij_2} \left(\frac{y_{ij_1}}{p_{ij_1}} - \frac{y_{ij_2}}{p_{ij_2}}\right)^2 \right]$$

$$= \sum_n \left[\sum_{j_1<j_2}^{N}\sum^{N} p_{0j_1} p_{0j_2} \left(\frac{y_{0j_1}}{p_{0j_1}} - \frac{y_{0j_2}}{p_{0j_2}}\right)^2 \frac{N_i(N_i-1)}{N(N-1)} \right]$$

$$= \sum_n \frac{N_i(N_i-1)}{N(N-1)} V$$

where, $V = \sum_{j_1<j_2}^{N}\sum^{N} p_{0j_1} p_{0j_2} \left(\frac{y_{0j_1}}{p_{0j_1}} - \frac{y_{0j_2}}{p_{0j_2}}\right)^2$

Hence, $Var(t_{RHC}) = \dfrac{\sum_n N_i^2 - N}{N(N-1)} V$

Now from C-S ~~inequality~~ inequality,

$$\left(\Sigma_n N_i^2\right)\left(\Sigma_n 1^2\right) \geq \left(\Sigma_n N_i\right)^2$$

$$\therefore \quad n \Sigma_n N_i^2 \geq N^2$$

or, $$\Sigma_n N_i^2 \geq \frac{N^2}{n}$$

Equality holds if $N_i \propto 1 \quad \forall i = 1(1)n$

$$N_i = K \quad \forall i = 1(1)n$$

$$\sum_{i=1}^{n} N_i = N = nK$$

$$\Rightarrow \quad K = \frac{N}{n} \quad \Rightarrow N_i = \frac{N}{n} \quad \text{—} (*)$$

For this value of N_i,

$$Var(t_{RHC}) = \frac{\Sigma_n N_i^2 - N}{N(N-1)} V$$

$$= \frac{\Sigma_n \frac{N^2}{n^2} - N}{N(N-1)} V \quad [\text{using } (*)]$$

$$= \frac{\frac{N^2}{n} - N}{N(N-1)} V$$

$$= \frac{\frac{N}{n} - 1}{(N-1)} V = \frac{N-n}{N-1} \cdot \frac{V}{n}$$

$$\leq \frac{V}{n} \quad \left[\because \frac{N-n}{N-1} \leq 1\right]$$

$$= Var(t_{HH})$$

☐ There is a gain w.r.to the usual PPSWR.

Scanned with CamScanner

Unbiased Estimation of Variance

Result:

Define $a = \sum_n Q_i \left(\frac{y_i}{p_i} - t_{RHC} \right)^2$. Then $\left(\frac{\sum_n N_i^2 - N}{N^2 - \sum_n N_i^2} \right) a$

is the u.e. of $Var(t_{RHC})$.

Proof.

$$a = \sum_n Q_i \left(\frac{y_i^2}{p_i^2} + t_{RHC}^2 - 2 t_{RHC} \frac{y_i}{p_i} \right)$$

$$= \sum_n \frac{y_i^2}{p_i^2/Q_i} + t_{RHC}^2 \sum_n Q_i - 2 t_{RHC} \sum_n \frac{y_i}{p_i/Q_i}$$

$$= \sum_n \frac{y_i^2}{p_i^2/Q_i} + t_{RHC}^2 - 2 t_{RHC}^2 \quad \left[\because \sum_n Q_i = 1 \ \& \ \sum_n \frac{y_i}{p_i/Q_i} = t_{RHC} \right]$$

$$= \sum_n \frac{y_i^2}{p_i^2/Q_i} - t_{RHC}^2$$

$$E(a) = E\left[\sum_n \frac{y_i^2}{p_i^2/Q_i} \right] - E\left[t_{RHC}^2 \right]$$

$$= E_1 E_2 \left[\sum_n \frac{y_i^2}{p_i^2/Q_i} \right] - Var(t_{RHC}) - Y^2 \quad \left[\because E(t_{RHC}) = Y \right]$$

$$= E_1 \left[\sum_n E_2 \left(\frac{y_i^2}{p_i^2/Q_i} \right) \right] - Var(t_{RHC}) - Y^2$$

$$= E_1 \left[\sum_n \left[\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}^2/Q_i} \cdot \frac{p_{ij}}{Q_i} \right] \right] - Var(t_{RHC}) - Y^2$$

$$= E_1 \left[\sum_n \left[\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}} \right] \right] - Var(t_{RHC}) - Y^2$$

$$= \sum_n E_1 \left[\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}} \right] - Var(t_{RHC}) - Y^2$$

$$= \sum_n \left[\frac{N_i}{N} \sum_{j=1}^{N} \frac{Y_j^2}{p_j} \right] - Var(t_{RHC}) - Y^2$$

$$= \sum_n N_i \left[\frac{1}{N} \sum_{j=1}^{N} \frac{Y_j^2}{p_j} \right] - Var(t_{RHC}) - Y^2$$

$$= \left[\sum_{j=1}^{N} \frac{Y_j^2}{p_j} - Y^2 \right] - Var(t_{RHC})$$

$$= V - V \left(\frac{\sum_n N_i^2 - N}{N(N-1)} \right)$$

$$= V \left[\frac{N^2 - N - \sum_n N_i^2 + N}{N(N-1)} \right]$$

$$= V \left[\frac{N^2 - \sum_n N_i^2}{N(N-1)} \right]$$

$$\therefore E \left[\frac{N(N-1)}{N^2 - \sum_n N_i^2} a \right] = V$$

$$E \left[\left(\frac{\sum_n N_i^2 - N}{N(N-1)} \right) \cdot \left(\frac{N(N-1)\, a}{N^2 - \sum_n N_i^2} \right) \right] = \left(\frac{\sum_n N_i^2 - N}{N(N-1)} \right) V$$

i.e. $$E \left[\left(\frac{\sum_n N_i^2 - N}{N^2 - \sum_n N_i^2} \right) a \right] = Var(t_{RHC})$$

■ Some results regarding MSE of HLUE :

Consider a homogeneously linear unbiased estimator (HLUE) of Y, say $t = \sum_{i=1}^{N} b_{si} I_{si}$

$E_p(t) = Y$

$$M_p(t) = E_p(t - Y)^2 = E_p \left[\sum_{i=1}^{N} b_{si} I_{si} Y_i - \sum_{i=1}^{N} Y_i \right]^2$$

$$= E_p \left[\sum_{i=1}^{N} (b_{si} I_{si} - 1) Y_i \right]^2$$

$$= E_p \left[\sum_{i=1}^{N} \sum_{j=1}^{N} (b_{si} I_{si} - 1)(b_{sj} I_{sj} - 1) Y_i Y_j \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} Y_i Y_j \, E_p \left[(b_{si} \cdot I_{si} - 1)(b_{sj} I_{sj} - 1) \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} Y_i Y_j \, d_{ij} \quad \left(\text{where, } d_{ij} = E_p \left[(b_{si} I_{si} - 1)(b_{sj} \cdot I_{sj} - 1) \right] \right.$$

Result

Let $\exists$ $w_i \neq 0$ such that $Z_i = \frac{Y_i}{N_i}$, $i = 1(1)N$

If $Z_i = c$, $\forall$ $i = 1(1)N$, then $M_p(t) = 0$

Illustration 1 (By Ratio Estimation)

Ratio estimator of population total

$$t = \frac{\bar{y}}{\bar{x}} \cdot X = X \cdot \frac{\sum_{i \in s} Y_i}{\sum_{i \in s} X_i}$$

Here, $X_i \neq 0$ $\forall i = 1(1)N$

Define, $Z_i = \frac{Y_i}{X_i}$

$Z_i = c$; $\forall i = 1(1)N$ $\Rightarrow$ $Y_i = cX_i$, $\forall i = 1(1)N$

$\Rightarrow$ $Y = \sum_{i=1}^{N} cX_i = cX$

$$\therefore t = X \frac{\sum_{i \in s} cX_i}{\sum_{i \in s} X_i} = cX$$

$\therefore t - Y = 0$ $\Rightarrow$ $E_p(t-Y)^2 = 0 \longrightarrow MSE(t)$

Illustration 2 (By Hansen-Hurwitz's Estimator)

$$t_{HH} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} \; ; \; 0 < p_i < 1 \;\&\; p_i \neq 0, \forall i$$

Define, $Z_i = \frac{Y_i}{p_i}$ ~~$Z_i = c$; $\forall i$~~

$Z_i = c$ $\forall i$ $\Rightarrow$ $Y_i = cP_i$ $\forall i$ $\Rightarrow$ $\sum_{i=1}^{N} Y_i = c \sum_{i=1}^{N} P_i$ $\Rightarrow$ $Y = c$

$\left[\because \sum_{i=1}^{N} P_i = 1\right]$

$$\& \; t_{HH} = \frac{1}{n} \sum_{i=1}^{n} \frac{cp_i}{p_i} = c$$

$$\therefore E_p(t_{HH} - Y)^2 = 0$$

Illustration

3. Horvitz - Thompson's Estimator :

$$t_{HT} = \sum_{i \in s} \frac{Y_i}{\pi_i} \quad , \pi_i \neq 0 \ \forall i$$

$$Z_i = \frac{Y_i}{\pi_i}$$

$$Z_i = c \Rightarrow Y_i = c\,\pi_i$$

$$\therefore \ Y = c\sum_{i=1}^{N} \pi_i = c\, E_p(\nu(s))$$

$$t_{HT} = c \sum_{i \in s} 1 = c\,\nu(s)$$

$$(t_{HT} - Y)^2 = c^2 \left[\nu(s) - E_p(\nu(s))\right]^2$$

$$E_p(t_{HT} - Y)^2 = c^2\, E_p\left[\nu(s) - E_p(\nu(s))\right]^2$$

$= 0$ only when $\nu(s) = E_p(\nu(s))$ w.p. 1 ie. it is a FES design.

Illustration 4 : Murthy's Estimator

$$t_M = \frac{\sum_{i \in s} Y_i\, p(s|i)}{p(s)}$$

Here we take $W_i = p_i$ & define $Z_i = \frac{Y_i}{p_i}$

$$\Rightarrow Y_i = Z_i\, p_i$$

$$Z_i = c \ \forall i \Rightarrow Y_i = c\,p_i \quad \& \quad \sum_{i=1}^{N} Y_i = c\sum_{i=1}^{N} p_i$$

$$\text{or, } Y = c \quad \left[\because \sum_{i=1}^{N} p_i = 1\right]$$

$$t_M = \frac{c\sum_{i \in s} p_i\, p(s|i)}{p(s)} = c \quad \left[\because p(s) = \sum_{i \in s} p_i\, p(s|i)\right]$$

Hence, $E_p(t_M - Y)^2 = 0$