Long

ST-1

1) explain Lahiri's method of drawing a sample Discuss about the gain due to pps sampling with replacement compared to simple random sampling

Ans:- If the number of units is very large then cummulativ of sizes may be very clumsy. This is avoided in these method.

Let $M = \max x_i$, $i = 1, 2 \ldots N$

Step ①:- select a number 'i' at random between 1 and N.

Step ②:- Select a random number 'R' between 1 and M. If $R \le x_i$, then unit 'i' is selected. otherwise, reject th unit 'i' and repeat the trial (i.e steps 1 and 2) till a unit is selected.

⟹ The whole process is repeated 'n' times The followir calculations, show that the probability of selecting of unit 'i' remains 'P_i' at each draw.

Let 'q_i' be the probability that unit 'i' gets selected at a trial

$$\therefore q_i = \frac{x_i}{NM}$$

probability that no unit is selected at a trial

$$= 1 - \sum_{i=1}^{N} q_i = 1 - \sum_{i=1}^{N} \frac{x_i}{NM} = 1 - \frac{\bar{X}}{M} = Q.$$

where $\bar{X} = \frac{\sum x_i}{N}$

probability that unit 'i' is selected at a draw,

$$= q_i + q_i Q + q_i Q^2 + \cdots$$

$$= \frac{q_i}{1-Q} \quad \text{where } q < 1$$

Hence substituting we get

$$= \frac{\frac{x_i}{NM}}{1-\left(1-\frac{\bar{X}}{M}\right)} = \frac{\frac{x_i}{NM}}{\frac{\bar{X}}{M}} = \frac{x_i}{N \cdot \bar{X}} = \frac{x_i}{N \cdot \frac{X}{N}} = \frac{x_i}{X} = P_i$$

2) Define Harwitz-Thompson estimator for the population and show that it is unbiased. Also obtain the expression for it's variance.

⇒ Harwitz-Thompson estimator

Suppose, y_i, $i=1,2 \ldots n$ is the value of i^{th} unit in the sample with π_i as the probability of including the i^{th} unit in the sample then Harwitz-Thompson estimator for the population total 'y' is given by $\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$

Proof: Consider the constants $C_1, C_2 \ldots C_N$ to N units of the population, then a general linear function of the sample values $y_1, y_2 \ldots y_n$ is written as

$$L(s) = \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{N} c_i a_i Y_i \longrightarrow (1)$$

where 'a_i' is a arbitrary value which takes 1 and 0

we have $a_i = 1$; if i^{th} unit is included in the sample

$a_i = 0$; otherwise

$$\therefore E(a_i) = 1 \cdot P(a_i=1) + 0 \cdot P(a_i=0)$$

$$E(a_i) = 1 \cdot \pi_i + 0 \cdot (1-\pi_i)$$

$$E(a_i) = \pi_i \longrightarrow (a)$$

$$\Rightarrow E(a_i^2) = 1^2\, P(a_i=0) + 0^2\, P(a_i=0)$$

$$= 1^2 \pi_i + 0^2(1-\pi_i)$$

$$= \pi_i$$

$$\therefore V(a_i) = E(a_i^2) - [E(a_i)]^2$$

$$= \pi_i - \pi_i^2 = \pi_i(1-\pi_i)$$

$$\therefore cov(a_i, a_j) = E(a_i a_j) - E(a_i) \cdot E(a_j)$$

$$= \pi_{ij} - \pi_i \cdot \pi_j$$

Scanned by CamScanner

from (1) $\Rightarrow E[l(s)] = E\left[\sum_{i=1}^{n} c_i y_i\right] = E\left[\sum_{i=1}^{N} c_i a_i Y_i)\right]$

$$= \sum_{i=1}^{N} c_i Y_i E(a_i)$$

$$= \sum_{i=1}^{N} c_i Y_i \pi_i$$

$$= Y$$

Is an unbiased ~~estimator of population total Y~~

This happens only when the constants $C_i = \frac{1}{\pi_i}$; $i = 1, 2 \ldots$

we get the unbiased estimator of population total Y.

$$\therefore \hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{N} \frac{a_i Y_i}{\pi_i}$$

It is called as Horwitz Thompson estimator of population total Y.

⑥ Define Horwitz Thompson estimator and derive $V_{HT}(\hat{Y}_{HT})$

To define $V(\hat{Y}_{HT}) = \sum_{i=1}^{N}(1-\pi_i)\frac{Y_i^2}{\pi_i} + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N}(\pi_{ij} - \pi_i\pi_j)\left(\frac{Y_i}{\pi_i} \cdot \frac{Y_j}{\pi_j}\right)$

Here Horwitz Thompson estimator for population total is given by $\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{N} \frac{a_i Y_i}{\pi_i}$

Now applying variance

$$V(\hat{Y}_{HT}) = V\left(\sum_{i=1}^{N} \frac{a_i Y_i}{\pi_i}\right)$$

$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2} V(a_i) + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j} \operatorname{cov}(a_i, a_j)$$

$$(\because V(\sum a_i x_i) = \sum a_i^2 V(x_i) + \sum\sum a_i a_j \operatorname{cov}(x_i, x_j))$$

$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^{\cancel{2}}} \cancel{\pi_i}(1-\pi_i) + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}(\pi_{ij} - \pi_i \cdot \pi_j)$$

$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}(1-\pi_i) + \sum\sum \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}(\pi_{ij} - \pi_i\pi_j)$$

$$\boxed{\begin{array}{l} \because V(a_i) = \pi_i(1-\pi_i) \\ \operatorname{cov}(a_i, a_j) = \pi_{ij} - \pi_i\pi_j \end{array}}$$

Scanned by CamScanner

7) In PPSWOR S/T (i) $\sum_{i=1}^{N} \pi_i = n$ (ii) $\sum_{j \neq i=1}^{N} \pi_{ij} = (n-1)\pi_i$
(iii) $\sum_{i=1}^{N} \sum_{j>i=1}^{N} \pi_{ij} = \frac{1}{2} n(n-1)$.

(i) consider $\sum_{i=1}^{N} a_i = n$

$$E\left[\sum_{i=1}^{N} a_i\right] = E(n)$$

$$\sum_{i=1}^{N} E(a_i) = n$$

$$\sum_{i=1}^{N} \pi_i = n \qquad (\because E(a_i) = \pi_i)$$

(ii) consider $\sum_{i=1}^{N} a_i = n$, $\sum_{j \neq i=1}^{N} a_j = n - a_i \rightarrow (1)$

$\therefore$ multiply 'a_i' on B.s on eq (1)

$$\sum_{j \neq i=1}^{N} a_i a_j = a_i(n - a_i) = na_i - a_i^2$$

$$\sum_{j \neq i=1}^{N} E(a_i a_j) = n\,E(a_i) - E(a_i^2)$$

$$\sum_{j \neq i=1}^{N} \pi_{ij} = n \cdot \pi_i - \pi_i = (n-1)\pi_i$$

$E(a_i a_j) = \pi_{ij}$
$E(a_i) = \pi_i$
$E(a_i^2) = \pi_i$

(iii) we know that $\sum_{j \neq i=1}^{N} \pi_{ij} = (n-1)\pi_i \rightarrow (1)$

$$\therefore \sum_{j \neq i=1}^{N} \pi_{ij} = 2 \sum_{j>i=1}^{N} \pi_{ij} \rightarrow (2)$$

Taking $\sum_{i=1}^{N}$ on b.s of eq (1) we get

$$\sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \pi_{ij} = \sum_{i=1}^{N} (n-1)\pi_i$$

$$\sum_{i=1}^{N} 2 \sum_{j>i=1}^{N} \pi_{ij} = (n-1) \sum_{i=1}^{N} \pi_i \qquad \text{from (2)}$$

$$\sum_{i=1}^{N} \sum_{j>i=1}^{N} \pi_{ij} = \frac{n(n-1)}{2} \qquad \therefore \sum^{N} \pi_i = n$$

Scanned by CamScanner

Derive Yates and Grundy's ~~estimator~~ form of variance for estimator of population ~~total~~

Yates and Grundy's estimator of $V(\hat{Y}_{HT})$

we have sampling variance of Horwitz Thompson estimator of population total 'Y' is given by

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N} (1-\pi_i)\frac{y_i^2}{\pi_i} + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N}(\pi_{ij}-\pi_i\pi_j)\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}$$

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N} (1-\pi_i)\frac{y_i^2}{\pi_i} + 2\sum_{i=1}^{N}\sum_{j>i=1}^{N}(\pi_{ij}-\pi_i\pi_j)\frac{y_i}{\pi_i}\frac{y_j}{\pi_j} \longrightarrow ①$$

$$\left(\because \sum_{j\neq i=1}^{N}\pi_{ij} = 2\sum_{j>i=1}^{N}\pi_{ij}\right)$$

now we have

$$\sum_{j\neq i=1}^{N}(\pi_i\pi_j-\pi_{ij}) = \pi_i(1-\pi_i)$$

$$(1-\pi_i) = \sum_{i\neq i=1}^{N}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_i}\right) \longrightarrow ⓐ$$

now consider eq① RHS 1st term from (a)

$$\sum_{i=1}^{N}(1-\pi_i)\frac{y_i^2}{\pi_i} = \sum_{i=1}^{N}\sum_{i\neq j=1}^{N}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_i}\right)\frac{y_i^2}{\pi_i}$$

now add $\left(\frac{y_j}{\pi_j}\right)^2$ we have

$$= \sum_{i=1}^{N}\sum_{i\neq j=1}^{N}(\pi_i\pi_j-\pi_{ij})\left(\left(\frac{y_i}{\pi_i}\right)^2+\left(\frac{y_j}{\pi_j}\right)^2\right) \longrightarrow ⓑ$$

sub eqⓑ in eq① we get

$$V(\hat{Y}_{HT}) = \sum\sum(\pi_i\pi_j-\pi_{ij})\left[\left(\frac{y_i}{\pi_i}\right)^2+\left(\frac{y_j}{\pi_j}\right)^2\right] + 2\sum\sum(\pi_{ij}-\pi_i\pi_j)\left(\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\right)$$

$$= \sum\sum(\pi_i\pi_j-\pi_{ij})\left[\left(\frac{y_i}{\pi_i}\right)^2+\left(\frac{y_j}{\pi_j}\right)^2-2\left(\frac{y_i}{\pi_j}\frac{y_j}{\pi_j}\right)\right]$$

$$= \sum\sum(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \longrightarrow ②$$

Scanned by CamScanner

we define $L(s) = \sum\sum \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$

$$E[L(s)] = E\left[\sum\sum \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2\right]$$

$$= \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \longrightarrow (*)$$

Now Replacing RHS of eq (2) with linear function $L(s)$ we get

$$\hat{V}_{YG}(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

(3) Discuss the need for unequal probability sampling suggest an unbiased estimator of the population total under PPSWR scheme and derive its variance

Ans In PPSWR $\hat{Y}_{HH}$ is an unbiased estimator of the population total 'Y' where (i) $\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{P_i}$

(ii) $V(\hat{Y}_{HH}) = \frac{1}{n}\sum P_i\left(\frac{Y_i}{P_i} - Y\right)^2$

(i) proof:- Given $\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{P_i}$

$$\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{N} t_i \frac{Y_i}{P_i}$$

$$E(\hat{Y}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\frac{Y_i}{P_i}E(t_i)$$

$$= \frac{1}{n}\sum_{i=1}^{N}\frac{Y_i}{P_i} \cdot nP_i$$

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

$$= Y$$

$\boxed{E(t_i) = nP_i}$

$\therefore \hat{Y}_{HH}$ is an unbiased estimator of population total Y

$$\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{p_i}$$

$$V(\hat{Y}_{HH}) = V\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}\right)$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i^2} V(t_i) + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\frac{Y_i}{P_i}\frac{Y_j}{P_j}\operatorname{cov}(t_i,t_j)\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i^2} nP_i\cancel{(1-P_i)} + \sum\sum\frac{Y_i}{P_i}\frac{Y_j}{P_j}(-nP_iP_j)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i} - \sum Y_i^2 - \sum\sum Y_iY_j\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i} - \left(\sum Y_i\right)^2\right] \qquad \left[\because \left(\sum Y_i\right)^2 = \sum Y_i^2 + \sum Y_iY_j\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} P_i\left(\frac{Y_i}{P_i} - Y\right)^2$$

(8) ~~As~~ Answer same as above

(9) explain SRSWOR as a particular case of PPSWOR.

Sol) WKT $V(\hat{Y}_{HT}) = \sum_{i=1}^{N}\pi_i(1-\pi_i)\frac{Y_i^2}{\pi_i} + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N}(\pi_{ij}-\pi_i\pi_j)\frac{Y_i}{\pi_i}\cdot\frac{Y_j}{\pi_j}$

Now substituting $\pi_i = \frac{n}{N}$, $\pi_j = \frac{n}{N}$ and $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N}\frac{n}{N}\left(1-\frac{n}{N}\right)\frac{Y_i^2N^2}{n^2} + \sum\sum\left[\frac{n(n-1)}{N(N-1)} - \frac{n}{N}\frac{n}{N}\right]\frac{Y_iN}{n}\cdot\frac{Y_jN}{n}$$

$$= \sum\frac{n}{N}\left(\frac{N-n}{N}\right)\frac{N^2}{n^2}Y_i^2 + \sum\sum\left[\frac{N(n^2-n) - n^2(N-1)}{N^2(N-1)}\right]\frac{N^2}{n^2}Y_iY_j$$

$$= \sum\left(\frac{N-n}{n}\right)Y_i^2 + \sum\sum\left[\frac{\cancel{N}n^2 - Nn - \cancel{N}n^2 + n^2}{(N-1)n^2}\right]\cdot Y_iY_j$$

$$= \sum\left(\frac{N-n}{n}\right)Y_i^2 + \sum\sum\left[\frac{n(N-n)}{(N-1)n^2}\right]Y_iY_j$$

$$= \sum\left(\frac{N-n}{n}\right)Y_i^2 + \sum\sum\left[\frac{N-n}{n(N-1)}\right]Y_iY_j$$

$$= \frac{N-n}{n}\left[\sum Y_i^2 + \frac{1}{N-1}\sum\sum Y_iY_j\right]$$

$$= \cancel{\left(\frac{N-n}{n}\right)\frac{1}{N-1}\,n\sum Y_i^2}$$

$$V(\hat{Y}_{HT}) = \left(\frac{N-n}{n}\right)\frac{1}{N-1}\left[(N-1)\sum Y_i^2 - \sum\sum Y_i Y_j\right]$$

$$= \left(\frac{N-n}{n}\right)\frac{1}{N-1}\left[N\sum Y_i^2 - \sum Y_i^2 - \left(\sum Y_i\right)^2 + \sum Y_i^2\right]$$

$$\left[\because \left(\sum x_i\right)^2 = \sum x_i^2 + \sum\sum x_i x_j\right]$$

$$= \left(\frac{N-n}{n}\right)\frac{1}{N-1}\left[N\sum Y_i^2 - \left(\sum Y_i\right)^2\right]$$

$$= \left(\frac{N-n}{n}\right)\frac{N}{N-1}\left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{N}\right]$$

$$= \left(\frac{N-n}{n}\right)\frac{N}{N-1}\left[\sum Y_i^2 - \frac{N^2\bar{Y}^2}{N}\right]$$

$$= \left(\frac{N-n}{n}\right)\frac{N}{N-1}\left[\sum Y_i^2 - N\bar{Y}^2\right]$$

$$= N\left(\frac{N-n}{n}\right) S^2$$

$$= N^2\left(\frac{N-n}{Nn}\right) S^2$$

$$= N^2\, V(\hat{\bar{y}}_{SRSWOR})$$

$$= V(\hat{Y}_{SRSWOR})$$

$$\therefore V(\hat{Y}_{HT}) = V(\hat{Y}_{SRSWOR})$$

when $\pi_i = \frac{n}{N}$, $\pi_j = \frac{n}{N}$ and $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$, $\forall\, j \neq i = 1, 2 \ldots N$

Hence $\hat{Y}_{SRSWOR}$ becomes a particular case of PPSWOR.

Scanned by CamScanner

Outline two procedures of selecting samples with unequal probabilies without replacement.

Suppose 'n' units are selected from 'N' units of the population with probability proportional of size at each draw without replacing the units selected in the previous draws. This method of selecting the sample is know as "~~pps~~ probability proportional sampling without replacement (PPSWOR)

Let 'N' units in the population denoted by $U_1, U_2 - - U_N$. Let 'y_i' be the variable under study and let the values of 'y_i' be $y_1, y_2 - - y_N$. Let 'x_i' denote the size.

Therefore the sizes of the units be $X_1, X_2 - - X_N$ such that $\sum_{i=1}^{N} X_i = X$. Let the prob. of selecting the units be $P_1, P_2 - - P_N \ni \sum_{i=1}^{N} P_i = 1$ where $P_i = \frac{X_i}{X}$; $i = 1, 2 - - N$. which is the prob. of selecting the ith unit U_i

$\Rightarrow P_i \propto X_i$; $i = 1, 2 - - N$

In PPSWOR the prob. of selecting the ith unit U_i at the 1st draw is given by $P_i = \frac{X_i}{X}$, $i = 1, 2 - - N$ and $\sum X_i = X$.

The prob. of selecting the jth unit 'U_j' at the 2nd draw is given by $\frac{P_j}{1-P_i}$, $\forall\ j \neq i = 1, 2 - - N$ and so on.

Let PPSWOR of size 'n' be selected then the prob. of including the ith unit U_i with PPSWOR of size n is given by $\pi_i = P_i + \sum_{i=1}^{N} \sum_{j\neq i=1}^{N} \frac{P_i P_j}{1-P_j}$ (or)

$$\pi_i = P_i\left[1 + \sum\sum \frac{P_j}{1-P_j}\right]$$

Hence the prob. of including ith unit U_i and jth unit U_j in PPSWOR of size n is given by

~~π_{ij}~~

$$\pi_{ij} = \frac{P_i P_j}{1-P_i} + \frac{P_i P_j}{1-P_j} \; ; \; j \neq i = 1, 2 \ldots N$$

$$= P_i P_j \left[\frac{1}{1-P_i} + \frac{1}{1-P_j}\right]$$

For $i=1$, $j=2$, i.e $i \neq j$ then

$$\pi_{12} = P_1 P_2 \left[\frac{1}{1-P_1} + \frac{1}{1-P_2}\right]$$

The following Relations holds.

① R_1 : $\sum_{i=1}^{N} \pi_i = n$

② R_2 : $\sum_{j \neq i = 1}^{N} \pi_{ij} = (n-1)\pi_i$

③ R_3 : $\sum_{i=1}^{N} \sum_{j \neq i = 1}^{N} \pi_{ij} = \frac{1}{2} n(n-1)$

⑤ Describe Cumulative total method

Procedure

Let the size of the i^{th} unit U_i be X_i for $i = 1, 2 \ldots N$

Such that $X = \sum_{i=1}^{N} X_i$

Step①:- We associate the numbers $1 \rightarrow X_1$ with 1^{st} unit 'U_1' $(X_1+1) \rightarrow X_2$ to the 2^{nd} unit 'U_2' and so on.

Step②:- A number 'K' is choosen at random from $1 \rightarrow X$ wher $X = \sum X_i$ from Random Number Tables and the unit wit which this number 'K' is associated is selected into the sample.

Step③:- For selecting a sample of 'n' units with ppswor the above procedure is repeated 'n' times.

A table of cummulative total of sizes of the units is made. Let $T_i = x_1 + x_2 + \ldots + x_i$ a random num i.e 'K' which is drawn b/w 1 and T_N where $T_N = X$ i.e the sum of the yield of sizes. The U_i is selected if $T_{i-1} < K < T_i$ where These process is repeated for 'n' times.

ort

1) Define PPSWR and PPSWOR sampling.

Ans:- ~~PPSWR:-~~ ~~each sampling unit has unequal probability of selection, the probability being proportion to the size of the~~

PPSWR:- Consider a population consisting of N units and let Y_i be the value of the characteristic under study for the unit U_i of the population for $i = 1, 2 \ldots N$ denoted by $U_1, U_2 \ldots U_N$. Let the values of the variable be $Y_1, Y_2 \ldots Y_N$. Let X_i denote the size and let the sizes of the units $X_1, X_2 \ldots X_N$.

Let the prob. of selecting the unit be $P_1, P_2 \ldots P_N$

where $P_i = \frac{X_i}{X}$ where $X = \sum X_i$,

$\Rightarrow P_i \propto X_i \; ; \; i = 1, 2 \ldots N.$

The i^{th} unit U_i is selected with prob. proportional to its size 'X_i' to be included in the sample.

PPSWOR:- Suppose 'n' units are selected from N units of the population with probability proportional of size at each draw without replacing the units selected in the previous draws. This method of selecting the sample is know as PPSWOR.

2) Explain unbiased estimator of Horwitz Thompson for population mean.

Proof:- WKT $L(S) = \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{N} c_i a_i Y_i$

$$E[L(S)] = E\left[\sum_{i=1}^{N} c_i a_i Y_i\right]$$

$$= \sum_{i=1}^{N} c_i Y_i E(a_i)$$

$$= \sum_{i=1}^{N} c_i Y_i \pi_i$$

$$= \overline{Y}$$

This happens only when $c_i = \frac{1}{N\pi_i}$

$$\therefore E[L(S)] = \overline{Y}$$

Scanned by CamScanner

⑤ S/T SRSWR becomes a special case of PPSWR sam[ple] using Hansen Horwitz estimator

proof:- when $P_i = \frac{1}{N}$, $\forall\ i = 1, 2 - - N$ the we have

$$\hat{Y}_{HH} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{P_i}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\frac{1}{N}}$$

$$= \frac{N}{n} \sum_{i=1}^{n} y_i$$

$$= N \cdot \bar{y}$$

$$\hat{Y}_{HH} = \hat{Y}_{SRSWR}$$

⑧ Define PPS sampling with replacement (PPSWR). Derive an estimator for the variance of Hansen Horwitz estimator of population total.

sol) PPSWR an unbiased estimator of $V(\hat{Y}_{HH})$ for $n > 1$ is given by

$$Est.\ V(\hat{Y}_{HH}) = \hat{V}(\hat{Y}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{P_i} - \hat{Y}_{HH}\right)^2 \quad (or)$$

$$\hat{V}(\hat{Y}_{HH}) = \frac{1}{n(n-1)} \left[\sum_{i=1}^{n} \left(\frac{y_i}{P_i}\right)^2 - n\, \hat{Y}_{HH}^2\right]$$

proof:- consider $\sum_{i=1}^{n} \left(\frac{y_i}{P_i} - \hat{Y}\right)^2 = \sum_{i=1}^{n} \left[\left(\frac{y_i}{P_i} - Y\right) - (\hat{Y} - Y)\right]^2$ $\quad v(\hat{Y}_{HH})$

$$\sum \left(\frac{y_i}{P_i} - \hat{Y}\right)^2 = \sum_{i=1}^{n} \left[\left(\frac{y_i}{P_i} - Y\right)^2\right] - \left[n(\hat{Y} - Y)^2\right]$$

$$\therefore E\left[\sum \left(\frac{y_i}{P_i} - \hat{Y}\right)^2\right] = E\left[\sum_{i=1}^{n} \left(\frac{y_i}{P_i} - Y\right)^2\right] - n\, E(\hat{Y} - Y)^2$$

$$= E\left[\sum_{i=1}^{N} t_i \left(\frac{Y_i}{P_i} - Y\right)^2\right] - n\, V(\hat{Y})$$

$$\cancel{= \sum_{i=1}^{N} \left(\frac{Y_i}{P_i} E(t_i) - Y\right)^2 - n\, V(\hat{Y})} \quad \cancel{(\because E(t_i) = nP_i)}$$

$$\cancel{= \sum_{i=1}^{N} \left(\frac{Y_i}{P_i} nP_i - Y\right)^2 - n\, V(\hat{Y})}$$

$$\cancel{= n^2 \sum_{i=1}^{N} (Y_i - nY)^2 - n\, V(\hat{Y})}$$

$$= \sum_{i=1}^{N} E(t_i) \left(\frac{Y_i}{P_i} - Y\right)^2 - n\, V(\hat{Y}) \quad \therefore E(t_i) = 1$$

$$= \sum_{i=1}^{N} nP_i \left(\frac{Y_i}{P_i} - Y\right)^2 - n\, V(\hat{Y})$$

$$= n^2 \frac{1}{n} \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - Y\right)^2 - n\, V(\hat{Y})$$

$$= n^2\, V(\hat{Y}_{HH}) - n \cdot V(\hat{Y}).$$

$$\therefore E\left[\sum_{i=1}^{n}\left(\frac{y_i}{p_i}-\hat{Y}\right)^2\right] = n^2 V(\hat{Y}_{HH}) - n V(\hat{Y}_{HH})$$

$$E\left[\sum_{i=1}^{n}\left(\frac{y_i}{p_i}-\hat{Y}\right)^2\right] = n(n-1) V(\hat{Y}_{HH})$$

$$\frac{E\left[\sum_{i=1}^{n}\left(\frac{y_i}{p_i}-\hat{Y}\right)^2\right]}{n(n-1)} = V(\hat{Y}_{HH})$$

$$\therefore \hat{V}(\hat{Y}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n}\left(\frac{y_i}{p_i} - \hat{Y}_{HH}\right)^2$$