$$\therefore \quad E\left(\frac{1}{n_i}\right) = \frac{1}{np_i}\left[1 - \frac{E(\varepsilon_i)}{np_i} + \frac{E(\varepsilon_i^2)}{n^2p_i^2}\right] = \frac{1}{np_i}\left[1 + \frac{1-p_i}{np_i}\right] \qquad \text{[From (*) and (**)]}$$

3. *Number of Strata.* It has been pointed out earlier that the precision of stratified sampling under proportional allocation can be increased by increasing the number of strata. However, the number of data should not be increased indefinitely. We have proved in (7·115)

$$V(\bar{y}'_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{k} p_i S_i^2 + \frac{1}{n^2}\sum_{i=1}^{k}(1-p_i)S_i^2$$

If $S_i^2 = S_w^2$ (say), $\forall\, i = 1, 2, \ldots, k$, then

$$V(\bar{y}'_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_w^2 \sum_i p_i + \frac{1}{n^2} \cdot S_w^2 \left(k - \sum_i p_i\right)$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_w^2 + \frac{k-1}{n^2} S_w^2 \qquad \ldots (7\cdot115a)$$

As k increases, S_w^2 decreases but the multiplying factor $(k-1)$ increases the value of $(k-1)S_w^2$ at a relatively faster rate than the rate of decrease of S_w^2. Accordingly for given value of n, we might reach a value of k, beyond which any further increase may not result in any increase in the precision of the estimator.

7·11. SYSTEMATIC SAMPLING

Systematic sampling is a commonly employed technique if the complete and up-to-date list of the sampling units is available. This consists in selecting only the first unit at random, the rest being automatically selected according to some predetermined pattern involving regular spacing of units. Let us suppose that N sampling units are serially numbered from 1 to N in some order and a sample of size n is to be drawn such that

$$N = nk \quad \Rightarrow \quad k = (N/n) \qquad \ldots (7\cdot116)$$

where k, usually called the sampling interval, is an integer.

Systematic sampling consists in drawing a random number, say $i \le k$ and selecting the unit corresponding to this number and every kth unit subsequently. Thus the systematic sample of size n will consist of the units $i, i+k, i+2k, \ldots, i+(n-1)k$.

The random number 'i' is called the *random start* and its value determines, as a matter of fact, the whole sample.

7·11·1. Notations and Terminology. Let y_{ij} denote the observation on the jth unit of the ith sample, $(i = 1, 2, \ldots, k\,;\, j = 1, 2, \ldots, n)$.

$$\bar{y}_{i\cdot} = \text{Mean of the } i\text{th systematic sample} = \frac{1}{n}\sum_{j=1}^{n} y_{ij},\ (i = 1, 2, \ldots, k) \qquad \ldots (7\cdot117)$$

$$\bar{y}_{\cdot\cdot} = \text{The population mean} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_{i\cdot} \qquad \ldots (7\cdot118)$$

S^2 = Population mean square

$$= \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \frac{1}{nk-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 \qquad \ldots (7\cdot119)$$

Scanned with CamScanner

The k possible systematic samples together with their means are as given in *Table 7·9*.

TABLE 7·9 : SYSTEMATIC RANDOM SAMPLE

Random start	*Sample composition (Units in the sample)*	*Probability*	*Mean*
1	$1 \quad 1+k \ldots 1+jk \ldots 1+(n-1)k$	$1/k$	$\bar{y}_{1\cdot}$
2	$2 \quad 2+k \ldots 2+jk \ldots 2+(n-1)k$	$1/k$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	$i \quad i+k \ldots i+jk \ldots i+(n-1)k$	$1/k$	$\bar{y}_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$k \quad 2k \ldots (k+j)k \ldots nk$	$1/k$	$\bar{y}_{k\cdot}$

Thus, k rows of the table give the k-systematic samples. The columns of the above table are also sometimes referred to as k strata.

From the above table it is obvious that each of the N units in the population occurs once in only one of the k samples and thus has an equal chance of being included in the sample. Since the probability of selecting the ith sample ($i = 1, 2, \ldots, k$) as the systematic sample is $1/k$, we get

$$E(\bar{y}_{i\cdot}) = \frac{1}{k}\sum_{i=1}^{k} \bar{y}_{i\cdot} = \bar{y}_{\cdot\cdot} \qquad \ldots (7\cdot120)$$

Thus, if $N = nk$, the sample mean provides an unbiased estimate of the population mean. For notational convenience, we shall write

$$\bar{y}_{sys} = \text{Mean of the systematic sample} = \bar{y}_{i\cdot} \qquad \ldots (7\cdot121)$$

$$\therefore \quad \text{Var}(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \qquad \ldots (7\cdot122)$$

Remark. A systematic sample selected with interval k after a random start yields an equal probability sample, because each element has a probability ($1/k$) of being selected. The sample mean, therefore, is an unbiased estimator of the population mean, if the sample size is fixed. If n is not fixed, ($N \neq nk$), the estimate is not unbiased but is a good estimate usually. [See Circular Systematic Sampling.]

7·11·2. Variance of the Estimated Mean.

Theorem 7·10.

$$\text{Var}(\bar{y}_{sys}) = \frac{N-1}{N} \cdot S^2 - \frac{(n-1)k}{N} \cdot S^2_{wsy} \qquad \ldots (7\cdot123)$$

where

$$S^2_{wsy} = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 \qquad \ldots (7\cdot123a)$$

is the mean square among units which lie within the same systematic sample.

[In *ANOVA* terminology S^2_{wsy} may be interpreted as within (samples) mean sum of squares.]

Proof. We have

$$(N-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i}\sum_{j} (y_{ij} - \bar{y}_{i\cdot} + \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2,$$

Scanned with CamScanner

the co-variance term vanishes, since

$$\sum_i \sum_j (y_{i\cdot} - \bar{y}_{i\cdot})(\bar{y}_{i\cdot} - \bar{y}_{..}) = \sum_{i=1}^{k} \left[(\bar{y}_{i\cdot} - \bar{y}_{..}) \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot}) \right] = 0$$

$$\therefore \quad (n-1) S^2 = \sum_i \sum_j (y_{i\cdot} - \bar{y}_{i\cdot})^2 + n \sum_{i=1}^{k} (\bar{y}_{i\cdot} - \bar{y}_{..})^2$$

$$= k(n-1) S^2_{wsy} + nk \text{ Var}(\bar{y}_{sys}) \text{ [From (7·122) and (7·123a)]}$$

$$\Rightarrow \quad \text{Var}(\bar{y}_{sys}) = \frac{N-1}{N} \cdot S^2 - \frac{k(n-1)}{N} \cdot S^2_{wsy} \qquad (\because N = nk)$$

Cor. Systematic Sampling vs. Simple Random Sampling.

In case of *srswor*, $\text{Var}(\bar{y}_{i\cdot}) = \dfrac{N-n}{Nn} \cdot S^2$, where S^2 is defined in (7·123).

$$\text{Var}(\bar{y}_n) - \text{Var}(\bar{y}_{sys}) = \left(\frac{N-n}{n} - N + 1 \right) \frac{S^2}{N} + \frac{k(n-1)}{N} \cdot S^2_{wsy}$$

$$= \frac{k(n-1)}{N} \cdot S^2_{wsy} - \frac{n-1}{n} \cdot S^2$$

$$= \frac{n-1}{n} \left(S^2_{wsy} - S^2 \right) \qquad (\because N = nk) \qquad \dots (7·124)$$

The systematic sampling gives more precise estimate of the population mean as compared with *srswor* if and only if

$$\text{Var}(\bar{y}_n) - \text{Var}(\bar{y}_{sys}) > 0 \quad \Rightarrow \quad S^2_{wsy} > S^2 \qquad \dots (7·125)$$

This leads to the following important conclusion : "*A systematic sample is more precise than a simple random sample without replacement if the mean square within the systematic sample is larger than the population mean square*". In other words, systematic sampling will yield better results only if the units within the same sample are heterogeneous.

Theorem 7·11.

$$\text{Var}(\bar{y}_{sys}) = \frac{nk-1}{nk} \cdot \frac{S^2}{n} \{1 + (n-1)\rho\} \qquad \dots (7·126)$$

where ρ is the intra-class correlation coefficient between the units of the same systematic sample and is given by :

$$\rho = \frac{\sum_{i=1}^{k} \sum_{j \neq j'=1}^{n} (y_{ij} - \bar{y}_{..})(y_{ij'} - \bar{y}_{..})}{nk(n-1)\sigma^2} \qquad \dots (7·127)$$

$$= \frac{\sum_{i=1}^{k} \sum_{j \neq j'=1}^{n} (y_{ij} - \bar{y}_{..})(y'_{ij} - \bar{y}_{..})}{(n-1)(nk-1) S^2} \qquad \dots (7·127a)$$

$$[\text{Since } N\sigma^2 = (N-1)S^2 \quad \Rightarrow \quad nk\sigma^2 = (nk-1)S^2]$$

Proof. We have

$$\text{Var}(\bar{y}_{sys}) = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_{i\cdot} - \bar{y}_{..})^2 = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{1}{n} \sum_{j=1}^{n} y_{ij} - \bar{y}_{..} \right)^2$$

Scanned with CamScanner

$$\Rightarrow \quad n^2k \operatorname{Var}(\bar{y}_{sys}) = \sum_{i=1}^{k}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})\right]^2 \qquad \ldots(7\cdot127b)$$

$$= \sum_{i=1}^{k}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 + \sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{..})(y'_{ij}-\bar{y}_{..})\right]$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y})(y_{ij'}-\bar{y}_{..})$$

$$= (nk-1)S^2 + (n-1)(nk-1)S^2\rho$$

[Using 7·119 and (7·127*a*)]

$$= (nk-1)S^2\left[1+(n-1)\rho\right]$$

$$\Rightarrow \quad \operatorname{Var}(\bar{y}_{sys}) = \frac{nk-1}{nk}\cdot\frac{S^2}{n}\left[1+(n-1)\rho\right]$$

Thus, we see that a positive intrac-lass correlation between the units of the same sample inflates the variability of the estimate. Due to the multiplier $(n-1)$, the increase is quite significant even for small positive values of ρ.

***Cor.* 1.** We have

$$\operatorname{Var}(\bar{y}_{sys}) \geq 0 \quad \Rightarrow \quad \rho \geq -\frac{1}{n-1}$$

Thus, the minimum value of ρ is $-\dfrac{1}{n-1}$ and in this case $\operatorname{Var}(\bar{y}_{sys}) = 0$.

2. *Systematic Sampling vs srswor.* The relative efficiency of the estimate of the population mean in systematic sampling over *srswor* is given by the expression :

$$E = \frac{\operatorname{Var}(\bar{y}_n)}{\operatorname{Var}(\bar{y}_{sys})} = \frac{N-n}{Nn}\cdot S^2 \div \left[\frac{(nk-1)S^2}{n^2k}\left\{1+(n-1)\rho\right\}\right]$$

$$= \frac{n(k-1)}{(nk-1)\left[1+(n-1)\rho\right]} \qquad \ldots(7\cdot128)$$

Obviously this depends on the value of ρ.

$$E > 1 \quad \Rightarrow \quad \frac{n(k-1)}{(nk-1)\,[1+(n-1)\rho]} > 1$$

$$\Rightarrow \quad (nk-n) > nk-1+(n-1)(nk-1)\rho \quad \Rightarrow \quad -1 > (nk-1)\rho\text{, i.e., } \rho < -\frac{1}{(nk-1)}$$

Thus *systematic sampling would be more efficient as compared with srswor if* $\rho < \dfrac{1}{(nk-1)}$.

On the other hand, *srswor would be superior to systematic sampling if* $\rho > -\dfrac{1}{(nk-1)}$.

However, if ρ assumes the minimum possible value, *i.e.*, $\rho = \{-1/(n-1)\}$ (*c.f.* Cor. 1 above), then $\operatorname{Var}(\bar{y}_{sys}) = 0$ and consequently $E = \infty$. Thus in this case reduction in $\operatorname{Var}(\bar{y}_{sys})$ over *srswor* will be 100%.

If ρ assumes the maximum value, *i.e.*, if $\rho = 1$, then from (7·128), we get

$$E = \frac{k-1}{nk-1} \qquad \ldots(7\cdot128a)$$

Scanned with CamScanner

7·11·3. Systematic Sampling vs Stratified Random Sampling. Let us now regard the population of $N = nk$ units to be divided into n strata corresponding to the n columns of the *Table 7·9* of § 7·11·1 [page 7·77] and suppose that one unit is drawn randomly from each startum, thus giving us a stratified random sample of size n. Then

$$\text{Mean of the } j\text{th stratum} = \bar{y}_{\cdot j} = \frac{1}{k}\sum_{i=1}^{k} y_{ij},\ (j = 1, 2, \ldots, n) \qquad \ldots (7\cdot129)$$

$$\text{Population mean} = \bar{y}_{..} = \frac{1}{nk}\sum_{i}\sum_{j} y_{ij} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{\cdot j} \qquad \ldots (7\cdot129a)$$

$$\text{Stratum mean square} = S_j^2 = \frac{1}{N_j - 1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{\cdot j})^2 = \frac{1}{k-1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{\cdot j})^2 \qquad \ldots (7\cdot129b)$$

$$(\because N_j = k\ ; j = 1, 2, \ldots, n)$$

$$S_{wst}^2 = \text{Pooled mean square between units within strata} = \frac{1}{n(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot j})^2 \qquad \ldots (7\cdot129c)$$

[In terminology of *ANOVA*, S_{wyt}^2 is the within (stratum) mean sum of squares.]

ρ_{wst} is the correlation coefficient between deviations from stratum means of pairs of items that are in the same systematic sample. Thus,

$$\rho_{wst} = \frac{E(y_{ij} - \bar{y}_{\cdot j})(y'_{ij} - \bar{y}'_{j})}{E(y_{ij} - \bar{y}_{\cdot j})^2}$$

$$= \frac{1}{k(n-1)} \cdot \frac{\sum_{i=1}^{k}\sum_{j \neq j' = 1}^{k}(y_{ij} - \bar{y}_{\cdot j})(\bar{y}'_{ij} - \bar{y}'_{\cdot j})}{\frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{k}(y_{ij} - \bar{y}_{\cdot j})^2}$$

$$= \frac{\sum_{i=1}^{k}\sum_{j \neq j' = 1}^{n}(y_{ij} - \bar{y}_{\cdot j})(y_{ij'} - \bar{y}_{\cdot j'})}{(n-1)\,n(k-1)\,S^2_{wst}} \qquad \text{[From (7·129c)]} \qquad \ldots (7\cdot130)$$

Theorem 7.12.

$$Var(\bar{y}_{sys}) = \frac{k-1}{nk}\, S^2_{wst}\,[1 + (n-1)\,P_{wst}] \qquad \ldots (7\cdot131)$$

Proof. We have

$$Var(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i\cdot} - \bar{y}_{..})^2 = \frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{n}\sum_{j=1}^{n} y_{ij} - \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{\cdot j}\right]^2$$

$$= \frac{1}{n^2 k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot j})\right]^2$$

$$= \frac{1}{n^2 k}\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{ij})^2 + \sum_{i=1}^{k}\sum_{j \neq j' = 1}^{n}(y_{ij} - \bar{y}_{ij})(y_{ij'} - \bar{y}_{ij'})\right]$$

$$= \frac{1}{n^2 k}\left[n(k-1)S^2_{wst} + n(n-1)(k-1)\rho_{wst}\,S^2_{wst}\right]$$

$$= \frac{k-1}{nk}\,S^2_{wst}\left[1 + (n-1)\,\rho_{wst}\right]. \qquad \text{[From (7·129c) and (7·130)]}$$

Scanned with CamScanner

Cor. Systematic Sampling vs Stratified Random Sampling. We have c.f. [7·47)]

$$\text{Var}(\bar{y}_{st}) = \sum_{j=1}^{n}\left(\frac{1}{n_j} - \frac{1}{N_j}\right) p_j^2 S_j^2$$

But $N_j = k$ and $n_j = 1, (j = 1, 2, ..., n)$ and $p_j = \dfrac{N_j}{N} = \dfrac{k}{nk} = \dfrac{1}{n}$

$$\therefore \quad \text{Var}(\bar{y}_{st}) = \sum_{j=1}^{n}\left(1 - \frac{1}{k}\right)\frac{1}{n^2} S_j^2 = \frac{k-1}{n^2k}\sum_{j=1}^{n} S_j^2 \quad ...(7·131a)$$

$$= \frac{k-1}{n^2k}\sum_{j=1}^{n}\left[\frac{1}{k-1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2\right] \quad \text{[From (7·129b)]}$$

$$= \frac{1}{n^2k}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{.j})^2 = \frac{k-1}{nk}\cdot S^2_{wst} \quad \text{[From (7·129c)]} \quad ...(7·132)$$

Comparing (7·131) and (7·132), we get

$$E' = \frac{\text{Var}(\bar{y}_{st})}{\text{Var}(\bar{y}_{sys})} = \frac{1}{1 + (n-1)\rho_{wst}} \quad ...(7·133)$$

Thus we see that the relative efficiency of systematic sampling over stratified random sampling depends upon the values of ρ_{wst} and nothing can be concluded in general.

If $\rho_{wst} > 0$, then $E' < 1$ and thus in this case stratified sampling will provide a better estimate of $\bar{y}_{..}$. However, if $\rho_{wst} = 0$, then $E' = 1$ and consequently both systematic sampling and stratified sampling provide estimates of $\bar{y}_{..}$ with equal precision.

Remark. *On comparison of systematic sampling with stratified random sampling or srswor.* As already discussed in the corollaries to *theorems 7·11 and 7·12*, the relative efficiency of the systematic sampling over stratified random sampling or *srswor* depends largely on the properties of the population under study and the conditions under which systematic sampling is superior has also been obtained. Thus without a knowledge of the structure of the population no hard and fast rules can be laid down and no situations can be pinpointed where the use of systematic sampling is to be recommended.

Example 7·21. *The data in Table 7·10 are for small artificial population which exhibits a fairly steady rising trend. Each column represents a systematic sample and the rows are the strata. Compare the precision of systematic sampling, random sampling and stratified sampling.*

Data for 10 systematic samples with $n = 4$, $k = 10$, $N = nk = 40$.

TABLE 7·10

Strata	Systematic Sample Number									
	1	2	3	4	5	6	7	8	9	10
I	0	1	1	2	5	4	7	7	8	6
II	6	8	9	10	13	12	15	16	16	17
III	18	19	20	20	24	23	25	28	29	27
IV	26	30	31	31	33	32	35	37	38	38

Solution. We have

$$\text{Var}(y_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = \frac{1}{k}\left[\sum_{i=1}^{k}\bar{y}_{i.}{}^2 - k\,\bar{y}_{..}{}^2\right] \quad ...(1)$$

$$= \frac{1}{n^2k}\left[\sum_{i=1}^{k}(n\,\bar{y}_{i.})^2 - n^2k\bar{y}_{..}{}^2\right]$$

Scanned with CamScanner

Unit IV : Systematic Sampling.

Systematic sampling is a commonly employed technique if the complete and up-to-date list of the sampling units is available. This consists in selecting only the first unit at random, the rest being automatically selected according to some predetermined pattern involving regular spacing of units.

Let us suppose that N sampling units are serially numbered from 1 to N in some order and a sample of size 'n' is to be drawn such that

$$N = nk \implies k = \frac{N}{n}$$

where k is the sampling interval, an integer.

Systematic sampling consists in drawing a random number, say, $i \le k$ and selecting the unit corresponding to this number and every k^{th} unit subsequently. Thus the systematic sample of size 'n' will consist of the units $i, i+k, i+2k, \ldots, i+(n-1)k$. The random number 'i' is called the random start.

Notations and Terminology

Let y_{ij} denote the observation on the j^{th} unit of the ith sample, $(i=1,2,\ldots,k;\ j=1,2,\ldots,n)$.

Sample mean $= \bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}, \quad (i=1,2,\ldots,k)$ — (1)

Population mean $= \bar{y}_{..} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$ — (2)

Population mean square $= S^2 = \frac{1}{nk-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$ — (3)

Variance $= V(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$ — (4)

Scanned with CamScanner

The 'k' possible systematic samples together with their means are

Random Start	Sample Composition (Units in the sample)	Probability	Mean
1	1 $\quad$ 1+k ... 1+jk ... 1+(n-1)k	$1/k$	$\bar{y}_{1.}$
2	2 $\quad$ 2+k ... 2+jk ... 2+(n-1)k	$1/k$	$\bar{y}_{2}$
⋮		⋮	⋮
i	i $\quad$ i+k ... i+jk ... i+(n-1)k	$1/k$	$\bar{y}_{i.}$
⋮	⋮ $\quad$ ⋮	⋮	⋮
k	k $\quad$ 2k ... (1+j)k ... nk	$1/k$	$\bar{y}_{k.}$

The k rows of the table gives the k-systematic samples. The columns are sometimes referred to as 'n' strata

Theorem: 1

In systematic sampling $V(\bar{y}_{sys}) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsys}$

where $S^2_{wsys} = \frac{1}{k(n-1)} \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$ is the mean square among units which lie within the same systematic sample.

Proof.

With usual identity, the population mean square

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$

$$\Rightarrow (N-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$

Add and subtract $\bar{y}_{i.}$ within the bracket, we get

$$(N-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (\overline{y_{ij} - \bar{y}_{i.}} + \overline{\bar{y}_{i.} - \bar{y}_{..}})^2$$

Scanned with CamScanner

$$(N-1)s^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..})$$

$$(N-1)s^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 + 0$$

since sum of deviations taken from its mean is zero.

Divide both sides by N, we get

$$\frac{N-1}{N}s^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{N} + \cancel{n}\,\frac{\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2}{N=\cancel{n}k}$$

But given $S^2_{wsys} = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$

$$k(n-1)S^2_{wsys} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

and $V(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$.

$$\therefore \frac{N-1}{N}s^2 = \frac{k(n-1)}{N}S^2_{wsys} + \cancel{k}\,\frac{V(\bar{y}_{sys})}{\cancel{k}}$$

$$\Rightarrow V(\bar{y}_{sys}) = \frac{N-1}{N}s^2 - \frac{k(n-1)}{N}S^2_{wsys}$$

and hence the proof

Scanned with CamScanner

Theorem : 2

In systematic sampling

$$V(\bar{y}_{sys}) = \frac{nk-1}{nk} \frac{S^2}{n} \left[1+(n-1)\rho\right]$$

Where ρ is the intra class Correlation Co-eff t between the units of the same systematic sample and is given by

$$\rho = \frac{\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n} (y_{ij}-\bar{y}_{..})(y_{ij'}-\bar{y}_{..})}{nk(n-1)\sigma^2}$$

$$= \frac{\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n} (y_{ij}-\bar{y}_{..})(y_{ij'}-\bar{y}_{..})}{(n-1)(nk-1)S^2}$$

[Since $N\sigma^2 = (N-1)S^2$
$nk\sigma^2 = (nk-1)S^2$]

Proof : We know that

$$V(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 \quad \text{from } (4)$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{n}\sum_{j=1}^{n}y_{ij}-\bar{y}_{..}\right]^2 \quad \because \bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n}y_{ij} \text{ from } (1)$$

$$= \frac{1}{n^2k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n}y_{ij}-n\bar{y}_{..}\right]^2$$

$$= \frac{1}{n^2k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n}y_{ij}-\sum_{j=1}^{n}\bar{y}_{..}\right]^2 = \frac{1}{n^2k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})\right]^2$$

$$= \frac{1}{n^2k}\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{..})(y_{ij'}-\bar{y}_{..})\right] \quad (1)$$

$$\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})\right]^2 = \left[(y_{i1}-\bar{y}_{..})+(y_{i2}-\bar{y}_{..})+\cdots+(y_{in}-\bar{y}_{..})\right]^2$$

$$= \sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 + \sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{..})(y_{ij'}-\bar{y}_{..})$$

But we know that

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$$

$$\Rightarrow (N-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$$

and

$$\rho = \frac{\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{..})(y_{ij'}-\bar{y}_{..})}{(n-1)(nk-1)S^2}$$

Scanned with CamScanner

Substitute these values in ①, we get

$$V(\bar{y}_{sys}) = \frac{1}{n^2k}\left[(N-1)S^2 + \rho(n-1)(nk-1)S^2\right]$$

$$= \frac{1}{n^2k}\left[(N-1)S^2\{1+\rho(n-1)\}\right]$$

$$= \frac{N-1}{nk}\frac{S^2}{n}\left[1+(n-1)\rho\right] \qquad \because nk = N$$

$$= \frac{nk-1}{nk}\frac{S^2}{n}\left[(1+(n-1)\rho\right]$$

and hence the proof.

Systematic Sampling Vs Simple Random Sampling

Theorem : 1

The mean of the systematic sample is more precise than the mean of the SRS ie, $S^2_{wsys} > S^2$

Proof :

We know that the estimated variance

$$V(\bar{y}_{sys}) = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S^2_{wsys}$$

and the variance in SRS is

$$V(\bar{y}) = \frac{N-n}{N}\frac{S^2}{n}$$

The mean of the systematic sample is more precise than mean of the SRS if $V(\bar{y}_{sys}) < V(\bar{y})$

ie., $$\frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S^2_{wsys} < \frac{N-n}{N}\frac{S^2}{n}$$

$$-\frac{k(n-1)}{N}S^2_{wsys} < \frac{N-n}{N}\frac{S^2}{n} - \frac{N-1}{N}S^2$$

Scanned with CamScanner

$$-\frac{k(n-1)}{N} S^2_{wsys} < \frac{S^2}{N}\left[\frac{N-n}{n} - (N-1)\right]$$

$$-\frac{k(n-1)}{N} S^2_{wsys} < \frac{S^2}{N}\left[\frac{N-n-n(N-1)}{n}\right]$$

$$-\frac{k(n-1)}{N} S^2_{wsys} < \frac{S^2}{N}\left[\frac{N-\cancel{n}-nN+\cancel{n}}{n}\right]$$

$$-\frac{k(n-1)}{N} \cdot S^2_{wsys} < \frac{S^2(N-nN)}{Nn}$$

Multiply by -1 on both sides

$$\frac{k(n-1)}{N} \cdot S^2_{wsys} > -\frac{S^2}{n}(1-n)$$

$$S^2_{wsys} > \frac{S^2}{\cancel{n}}\cancel{(n-1)} \times \frac{N=\cancel{n}k}{k\cancel{(n-1)}}$$

$$\therefore S^2_{wsys} > S^2$$

and hence the proof.

Relative Efficiency

Theorem: 2

Systematic sample is more precise as compared with SRSWOR if $\rho < -\frac{1}{N-1}$ (or) $\rho < -\frac{1}{nk-1}$

Proof:

If S~~R~~S is more precise than SRSWOR then relative efficiency is given by

$$E = \frac{V(\bar{y})_R}{V(\bar{y}_{sys})}$$

If $E>1$, then the mean of the variance of systematic sample is more precise than mean of the SRSWOR.

Scanned with CamScanner

We know that $V(\bar{y})_R = \frac{N-n}{N} \frac{S^2}{n}$

and $V(\bar{y}_{sys}) = \frac{N-1}{N} \frac{S^2}{n} \left[1+(n-1)\rho\right]$

(or) $= \frac{nk-1}{nk} \frac{S^2}{n} \left[1+(n-1)\rho\right]$

then the relative efficiency

$$E = \frac{V(\bar{y})_R}{V(\bar{y}_{sys})} = \frac{\frac{N-n}{N} \frac{S^2}{n}}{\frac{nk-1}{nk} \frac{S^2}{n} \left[1+(n-1)\rho\right]}$$

$$= \frac{N-n}{\cancel{N}} \frac{\cancel{S^2}}{\cancel{n}} \times \frac{\cancel{nk}}{nk-1} \frac{\cancel{n}}{\cancel{S^2}} \times \frac{1}{1+(n-1)\rho}$$

$$= \frac{nk-n}{nk-1} \times \frac{1}{1+(n-1)\rho}$$

$$= \frac{n(k-1)}{nk-1} \times \frac{1}{1+(n-1)\rho}$$

if $E>1$, then

$$\frac{n(k-1)}{nk-1} \times \frac{1}{1+(n-1)\rho} > 1$$

$$\frac{1}{1+(n-1)\rho} > \frac{nk-1}{n(k-1)}$$

$$1+(n-1)\rho < \frac{n(k-1)}{nk-1}$$

$$(n-1)\rho < \frac{nk-n}{nk-1} - 1$$

$$\rho < \frac{\cancel{nk}-n-\cancel{nk}+1}{(nk-1)(n-1)}$$

$$\rho < \frac{-n+1}{(nk-1)(n-1)}$$

$$\rho < \frac{-\cancel{(n-1)}}{(nk-1)\cancel{(n-1)}}$$

$$\rho < -\frac{1}{nk-1}$$

$$\rho < -\frac{1}{N-1}$$

Scanned with CamScanner

Remark :

If SRSWOR is more precise than the Systematic sample if $\rho > -\frac{1}{N-1}$, the relative efficiency

$$E = \frac{V(\bar{y})_R}{V(\bar{y}_{sys})} < 1.$$

Systematic Sampling Vs Stratified Sampling

In systematic sample $N = nk$. Let us now regard this population units to be divided into 'n' strata corresponding to the 'n' columns.

Random Start	Sampling Units	Probability	Mean
1	$1, 1+k, \ldots 1+jk, \ldots 1+(n-1)k$	$1/k$	$\bar{y}_{1.}$
2	$2, 2+k, \ldots 2+jk, \ldots 2+(n-1)k$	$1/k$	$\bar{y}_{2.}$
$\vdots$		$\vdots$	$\vdots$
i	$i, i+k \ldots i+jk \ldots i+(n-1)k$	$1/k$	$\bar{y}_{i.}$
$\vdots$			
k	$k, 2k \ldots (1+j)k \ldots nk$	$1/k$	$\bar{y}_{k.}$
Mean	$\bar{y}_{.1}, \bar{y}_{.2} \ldots \bar{y}_{.j} \ldots \bar{y}_{.k}$		

Suppose that one unit is drawn randomly from each stratum, thus giving us a stratified random sample of size n. Then

$$\left.\begin{array}{l}\text{Mean of the}\\ j\text{th stratum}\end{array}\right\} = \bar{y}_{.j} = \frac{1}{k}\sum_{i=1}^{k} y_{ij}, \quad j = 1, 2, \ldots, n$$

$$\text{Population mean} = \bar{y}_{..} = \frac{1}{nk}\sum_{i}\sum_{j} y_{ij} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{.j}$$

$$\left.\begin{array}{l}\text{Stratum}\\ \text{mean square}\end{array}\right\} = s_j^2 = \frac{1}{N_j - 1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2 \quad [\because N_j = k;\ j = 1, 2, \ldots, n]$$

Scanned with CamScanner

S^2_{wst} = Pooled mean square between units within Strata

$$= \frac{1}{n(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2$$

ρ_{wst} is the Correlation Co-eff't between deviations from stratum means of pairs of items that are in the same systematic sample. Thus

$$\rho_{wst} = \frac{E(y_{ij} - \bar{y}_{.j})(y_{ij'} - \bar{y}_{.j'})}{E(y_{ij} - \bar{y}_{.j})^2}$$

$$= \frac{1}{kn(n-1)} \frac{\sum_{i=1}^{k} \sum_{j \neq j'=1}^{n} (y_{ij} - \bar{y}_{.j})(y_{ij'} - y_{.j'})}{\frac{1}{nk} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2}$$

$$= \frac{\sum_{i=1}^{k} \sum_{j \neq j'=1}^{n} (y_{ij} - \bar{y}_{.j})(y_{ij'} - \bar{y}_{.j'})}{(n-1)n(k-1) S^2_{wst}}$$

Theorem: 1

In systematic sample $V(\bar{y}_{sys}) = \frac{k-1}{nk} S^2_{wst} \left[1 + (n-1)\rho_{wst}\right]$

Proof:

$$V(\bar{y}_{sys}) = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \frac{1}{k} \sum_{i=1}^{k} \left[\frac{\sum_{j=1}^{n} y_{ij}}{n} - \frac{\sum_{j=1}^{n} \bar{y}_{.j}}{n} \right]^2$$

$$= \frac{1}{n^2 k} \sum_{i=1}^{k} \left[\sum_{j=1}^{n} y_{ij} - \sum_{j=1}^{n} \bar{y}_{.j} \right]^2 = \frac{1}{n^2 k} \sum_{i=1}^{k} \left[\sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j}) \right]^2$$

$$= \frac{1}{n^2 k} \left[\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2 + \sum_{i=1}^{k} \sum_{j \neq j'=1}^{n} (y_{ij} - \bar{y}_{.j})(y_{ij'} - \bar{y}_{.j'}) \right] \quad (1)$$

But $$S^2_{wst} = \frac{1}{n(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2$$

$$\rho_{wst} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})(y_{ij'} - \bar{y}_{.j'})}{(n-1) S^2_{wst}\, n(k-1)}$$

Scanned with CamScanner

$$V(\bar{y}_{sys}) = \frac{1}{n^2 k}\left[n(k-1) S^2_{wst} + \rho_{wst}(n-1) S^2_{wst}\, n(k-1)\right]$$

$$= \frac{\cancel{n}(k-1) S^2_{wst}}{n^{\cancel{2}} k}\left[1 + (n-1)\rho_{wst}\right]$$

$$= \frac{k-1}{nk} S^2_{wst}\left[1 + (n-1)\rho_{wst}\right] \quad \text{--- (2)}$$

Relative Efficiency

In Stratified random sampling

$$V(\bar{y}_{st}) = \sum_{j=1}^{n}\left(\frac{1}{n_j} - \frac{1}{N_j}\right) P_j^2 S_j^2 \quad \text{--- (1)}$$

But we know that $N_j = k$ and $n_j = 1$, $(j = 1, 2, \dots n)$

and $P_j = \frac{N_j}{N} = \frac{k}{nk} = \frac{1}{n}$

Substitute these values in (1), we get

$$V(\bar{y}_{st}) = \sum_{j=1}^{n}\left(1 - \frac{1}{k}\right)\frac{1}{n^2} S_j^2$$

$$= \sum_{j=1}^{n}\left(\frac{k-1}{n^2 k}\right) S_j^2$$

$$= \frac{k-1}{n^2 k}\sum_{j=1}^{n} S_j^2 \quad \text{--- (2)}$$

But $S_j^2 = \frac{1}{N_j - 1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2$

$$= \frac{1}{k-1}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2$$

Substitute S_j^2 values in (2), we get

$$V(\bar{y}_{st}) = \frac{\cancel{k-1}}{n^2 k}\,\frac{1}{\cancel{k-1}}\sum_{j=1}^{n}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2 \quad \text{--- (3)}$$

Scanned with CamScanner

$$V(\bar{y}_{st}) = \frac{1}{n^2 k} \sum_{j=1}^{n} \sum_{i=1}^{k} (y_{ij} - \bar{y}_{.j})^2$$

But $$S_{wst}^2 = \frac{1}{n(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2$$

$$n(k-1) S_{wst}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{.j})^2$$

Substitute these values in ③, we get

$$V(\bar{y}_{st}) = \frac{1}{n^{\cancel{2}} k} \cancel{n}(k-1) S_{wst}^2$$

$$= \frac{k-1}{nk} S_{wst}^2$$

The relative efficiency is $E' = \dfrac{V(\bar{y}_{st})}{V(\bar{y}_{sys})}$

$$\Rightarrow E' = \frac{\cancel{\frac{k-1}{nk}} \cancel{S_{wst}^2}}{\cancel{\frac{k-1}{nk}} \cancel{S_{wst}^2} \left[1 + (n-1)\rho_{wst}\right]} \rightarrow \text{from ② of Theorem: 1}$$

$$= \frac{1}{1 + (n-1)\rho_{wst}}$$

Thus we see that the relative efficiency of systematic sampling over stratified random sampling depends upon the values of ρ_{wst}.

If $\rho_{wst} > 0$, then $E' < 1$ and thus in this case stratified sampling will provide a better estimate of $\bar{y}_{..}$

If $\rho_{wst} = 0$, then $E' = 1$ and consequently both systematic sampling and stratified sampling provide estimates of $\bar{y}_{..}$ with equal precision.

Scanned with CamScanner

Theorem.

If the population consists of a linear trend, then prove that $V(\bar{y}_{st}) \le V(\bar{y}_{sys}) \le V(\bar{y}_n)_R$

Soln:

Let us suppose that the population has the linear trend given by the model $y_i = i\,;\ (i = 1, 2, \dots, N)$

then $\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

$$\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \frac{1}{\cancel{N}}\left[\frac{\cancel{N}(N+1)}{2}\right]$$

$$= \frac{N+1}{2}$$

Popln. mean square $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y}_N)^2$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2\right]$$

$$= \frac{1}{N-1}\left[\frac{N(N+1)(2N+1)}{6} - N\frac{(N+1)^2}{4}\right]$$

$$= \frac{1}{N-1}\left[\frac{2N(N+1)(2N+1) - 3N(N+1)^2}{12}\right]$$

$$= \frac{N(N+1)}{N-1}\left[\frac{2(2N+1) - 3(N+1)}{12}\right]$$

$$= \frac{N(N+1)}{N-1}\left[\frac{4N+2-3N-3}{12}\right]$$

$$= \frac{N(N+1)}{N-1}\left[\frac{N-1}{12}\right]$$

$$= \frac{N(N+1)}{12}$$

Scanned with CamScanner

$$\therefore V(\bar{y}_n)_R = \left(\frac{1}{n} - \frac{1}{N}\right) s^2$$

$$= \frac{N-n}{Nn} \times \frac{N(N+1)}{12}$$

$$= \frac{(N-n)(N+1)}{12n}$$

$$= \frac{(nk-n)}{n}\,\frac{(nk+1)}{12}$$

$$= (k-1)\left(\frac{nk+1}{12}\right) \quad \text{—— (1)}$$

If the population consists of 'N' units and if the j^{th} stratum consist of 'k' units then

$$V(\bar{y}_{st}) = \frac{k-1}{n^2 k} \sum_{j=1}^{n} S_j^2$$

$$= \frac{k-1}{n^2 k} \sum_{j=1}^{n} \frac{k(k+1)}{12} \qquad \text{since } S^2 = \frac{N(N+1)}{12}$$

$$= \frac{k-1}{n^2 k}\,\frac{nk(k+1)}{12}$$

$$= \frac{(k-1)(k+1)}{12n}$$

$$= \frac{k^2-1}{12n} \quad \text{—— (2)}$$

Consider $V(\bar{y}_{sys})$. In systematic sample,

$\bar{y}_{i\cdot}$ = mean of the i^{th} sample

$$= \frac{1}{n} \sum_{j=1} y_{ij}$$

$$= \frac{1}{n}\left[i + (i+k) + (i+2k) + (i+3k) + \cdots + \{i + (n-1)k\} \right]$$

$$= \frac{1}{n}\left[ni + \{1+2+\cdots+(n-1)\}k \right]$$

$$= \frac{1}{n}\left[ni + \frac{n(n-1)}{2}k \right]$$

$$= i + \frac{(n-1)}{2}k$$

Scanned with CamScanner

$$\bar{y}_{..} = \bar{Y}_N = \frac{1}{N}\Sigma y_i = \frac{1}{N}\Sigma i = \frac{1}{N}\left[\frac{N(N+1)}{2}\right] = \frac{N+1}{2}$$

and population mean $\bar{y}_{..} = \frac{N+1}{2}$

$$\Rightarrow \bar{y}_{i.} - \bar{y}_{..} = i + \frac{n-1}{2}k - \frac{nk+1}{2}$$

$$= i + \cancel{\frac{nk}{2}} - \frac{k}{2} - \cancel{\frac{nk}{2}} - \frac{1}{2}$$

$$= i - \frac{(k+1)}{2}$$

then $V(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2$

$$= \frac{1}{k}\sum_{i=1}^{k}\left[i - \frac{(k+1)}{2}\right]^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left[i^2 + \left(\frac{k+1}{2}\right)^2 - \cancel{2}i\frac{(k+1)}{\cancel{2}}\right]$$

$$= \frac{1}{k}\left[\sum_{i=1}^{k} i^2 + \sum_{i=1}^{k}\frac{(k+1)^2}{4} - (k+1)\sum_{i=1}^{k} i\right]$$

$$= \frac{1}{k}\left[\frac{k(k+1)(2k+1)}{6} + \frac{k(k+1)^2}{4} - (k+1)\frac{k(k+1)}{2}\right]$$

$$= \frac{1}{k}\left[\frac{2k(k+1)(2k+1) + 3k(k+1)^2 - 6(k+1)k(k+1)}{12}\right]$$

$$= \frac{\cancel{k}(k+1)}{\cancel{k}}\left[\frac{2(2k+1) + 3(k+1) - 6(k+1)}{12}\right]$$

$$= (k+1)\left[\frac{4k+2+3k+3-6k-6}{12}\right]$$

$$= (k+1)\left(\frac{k-1}{12}\right)$$

$$= \frac{k^2-1}{12} \quad \text{——— (3)}$$

Comparing (1), (2) & (3) we get

$$\frac{(k-1)(nk+1)}{12} \geq \frac{k^2-1}{12} \geq \frac{k^2-1}{12n}$$

$$\Rightarrow \quad V(\bar{y}_R) \geq V(\bar{y}_{sys}) \geq V(\bar{y}_{st})$$

under the condition that the population is of linear trend.

Scanned with CamScanner

Circular Systematic Sampling

In practical situation, some times N is not expressible in the form $N = nk$. ie, k is not an integer. In this case, the present sampling scheme will give rise to samples of unequal size. 'k' is taken as an Integer nearest to $\frac{N}{n}$. Then a random number is chosen from 1 to k and every k^{th} unit is drawn in the sample. Under this condition, the sample size is not necessarily n and in some cases it may be $(n-1)$.

For example, if $N = 11$, $n = 4$, then $k = \frac{N}{n} = \frac{11}{4} = 2.75 \simeq 3$. The possible samples are $(1, 4, 7, 10)$; $(2, 5, 8, 11)$ and $(3, 6, 9)$, which are not of the same size.

To overcome the difficulty of varying sample size under the situation $N \neq nk$, the procedure is modified slightly by which a sample of constant size is always obtained. The procedure consists in selecting a unit, by a random start, from 1 to N and then thereafter selecting every k^{th} unit, k being an integer nearest to N/n, in a circular manner, until a sample of 'n' units is obtained. This technique is generally known as circular systematic sampling. The sample will then consist of the units corresponding to the serial numbers

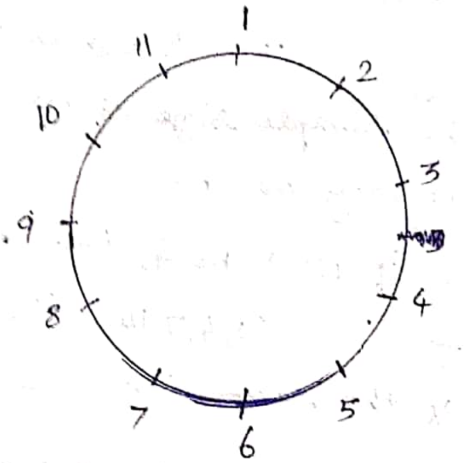
$$\begin{cases} i + jk ; & \text{if } i + jk \leq N \\ i + jk - N ; & \text{if } i + jk > N \end{cases} \quad \text{for } j = 0, 1, 2, \ldots, (n-1).$$

Scanned with CamScanner

For example, let ~~N = 11~~ and n = 4. Then $K = \frac{N}{n} = \frac{11}{4} = 2.75$

The possible samples are

(1, 4, 7, 10); (2, 5, 8, 11); (3, 6, 9, 1); (4, 7, 10, 2); (5, 8, 11, 3);
(6, 9, 1, 4); (7, 10, 2, 5); (8, 11, 3, 6); (9, 1, 4, 7); (10, 2, 5, 8);
(11, 3, 6, 9)



Theory Questions

1. Define Systematic Sampling. Discuss its advantages and Limitations (Apr'2008)
2. State the merits and demerits of Systematic sampling (Apr'
3. What is systematic sampling? Explain how a systemati sample is selected, by giving an example (Nov'2004)
4. State the uses of Systematic sampling (Nov'2004)
5. Compare systematic Sampling and SRS (Nov'2005)
6. " " " and St.R.S (Nov'2005)
7. Explain circular systematic sampling (Nov'2007)

Scanned with CamScanner

Source :

1. S.C. Gupta and V.K. Kapoor: Fundamental of Applied Statistics –Sultan Chand & Sons, Fourth Edition, 2015.