Frequently Asked Questions

1. What do you mean by Systematic Sampling?

Answer:
Systematic sampling (or interval random sampling) is a probability sampling procedure in which a random selection is made of the first element for the sample, and then subsequent elements are selected using a fixed or systematic interval until the desired sample size is reached.

This is a technique which has a nice feature of selecting a whole sample with just one random start. A sampling technique in which first unit is selected with a help of random numbers and the others get selected automatically according to some pre-designed pattern until the desired sample size is reached is known as systematic random sampling.

2. Write a note on Simple Random sampling.

Answer:
Simple Random Sample is a subset of individuals (a sample) chosen from a larger set that is a population. Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of n individuals has the same probability of being chosen for the sample as any other subset of n individuals. Samples drawn using this technique is known as simple random sampling.

3. Prove that Systematic sampling is more efficient than SRSWOR if $S^2{}_{sym} > S^2$.

Answer:
We know that if $\bar{y}$ is the mean of the Simple Random Sample of size n then $V(\bar{y})_{WOR} = \frac{(N-n)}{N}\frac{S^2}{n}$

Consider $V(\bar{y}_r) < V(\bar{y})_{wor}$

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

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

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

$$\Rightarrow \frac{S^2}{N}\left[\frac{Nn-n-N+n}{n}\right] < \frac{(n-1)S^2{}_{sym}}{n}$$

$$\Rightarrow S^2 \frac{(n-1)}{n} < \frac{(n-1)S^2{}_{sym}}{n}$$

$$\Rightarrow S^2 < S^2{}_{sym}$$

$$\Rightarrow S^2{}_{sym} > S^2$$

Systematic sampling is more efficient than SRSWOR if $S^2{}_{sym} > S^2$

4. When the units within the systematic sample are homogeneous comment on the systematic estimators.

Answer:

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

$$V(\bar{y})_{WOR} - V(\bar{y}_r)_{SYS} > 0 \Rightarrow S^2_{sym} > S^2$$

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.

This important result which applies in general states that Systematic sampling is more precise than Simple Random Sampling if the variance within the Systematic samples is larger than the population variance as a whole. Systematic sampling is precise when the units within the same sample are heterogeneous and is imprecise when they are homogeneous. The result is obvious intuitively. If there is little variation within a systematic sample relative to that in the population, the successive units in the sample are repeating more or less the same information

5. How do you obtain intra class correlation coefficient under Systematic sampling?

Answer:

Suppose ρ is the intra-class coefficient then

$$\rho = \frac{E(Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{E(Y_{ri} - \overline{Y})^2}$$

$$= \frac{\sum_{r=1}^{k} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y}) . \frac{1}{kn(n-1)}}{\sum_{r=1}^{k} \sum_{i=1}^{n} (Y_{ri} - \overline{Y})^2 \frac{1}{kn}}$$

$$= \frac{\sum_{r=1}^{k} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{\sum_{r=1}^{k} \sum_{i=1}^{n} (Y_{ri} - \overline{Y})^2 (n-1)}$$

$$\rho = \frac{\sum_{r=1}^{k} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{nk\sigma^2 (n-1)} \qquad \left[\boxtimes\ \sigma^2 = \frac{\sum_{r=1}^{k} \sum_{i \neq j} (Y_{ri} - \overline{Y})^2}{nk} \right]$$

6. Obtain the minimum value of intra class correlation coefficient under Systematic sampling so that variance of systematic sample mean is zero.

Answer:

Suppose $\bar{y}$ is the sample mean of SRS of size n drawn using WOR scheme from a population of size N then its variance is given by

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

Where S^2 is a population mean square

Consider $V(\bar{y}_r) = E(\bar{y}_r - E(\bar{y}_r))^2$

$$= E(\bar{y}_r - \overline{Y})^2$$

$$= \sum_{r=1}^{k} (\bar{y}_r - \overline{Y})^2 . \frac{1}{k}$$

$$=\frac{1}{k}\sum_{r=1}^{k}(\frac{1}{n}\sum_{i=1}^{n}y_{ri}-\overline{Y})^2$$

$$=\frac{1}{n^2k}\sum_{r=1}^{k}(\sum_{i=1}^{n}(y_{ri}-\overline{Y})^2)$$

$$=\frac{1}{n^2k}\sum_{r=1}^{k}\left[\sum_{i=1}^{n}(y_{ri}-\bar{Y})^2+\sum_{i\neq j}\sum(y_{ri}-\bar{Y})(y_{rj}-\bar{Y})\right]$$

$$=\frac{1}{n^2k}[\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri}-\overline{Y})^2+\sum_{r=1}^{k}\sum_{i\neq j}(y_{ri}-\overline{Y})(y_{rj}-\overline{Y})]$$

$$=\frac{1}{n^2k}[\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri}-\overline{Y})^2+\sum_{r=1}^{k}\sum_{i\neq j}(y_{ri}-\overline{Y})(y_{rj}-\overline{Y})] \longrightarrow 1$$

Suppose ρ is the intra-class coefficient then

$$\rho=\frac{E(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y})}{E(Y_{ri}-\overline{Y})^2}$$

$$=\frac{\sum_{r=1}^{k}\sum_{i\neq j}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y}).\frac{1}{kn(n-1)}}{\sum_{r=1}^{k}\sum_{i=1}^{n}(Y_{ri}-\overline{Y})^2\frac{1}{kn}}$$

$$=\frac{\sum_{r=1}^{k}\sum_{i\neq j}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y})}{\sum_{r=1}^{k}\sum_{i=1}^{n}(Y_{ri}-\overline{Y})^2(n-1)}$$

$$\rho=\frac{\sum_{r=1}^{k}\sum_{i\neq j}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y})}{nk\sigma^2(n-1)} \qquad \left[\boxtimes\ \sigma^2=\frac{\sum_{r=1}^{k}\sum_{i\neq j}(Y_{ri}-\overline{Y})^2}{nk}\right]$$

$$\Rightarrow \sum_{r=1}^{k} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y}) = \rho nk\sigma^2 (n-1) \longrightarrow 2$$

By substituting equation 2 in equation 1 we get

$$V(\bar{y}_r) = \frac{1}{n^2 k}[(nk-1)S^2 + (nk-1)(n-1)\rho S^2]$$

$$V(\bar{y}_r) = \frac{(nk-1)S^2}{n^2 k}[1+(n-1)\rho]$$

$$V(\bar{y}_r) = \frac{(N-1)S^2}{Nn}[1+(n-1)\rho]$$

Since $V(\bar{y}_r)$ is never less than zero ρ cannot be less than $\frac{-1}{n-1}$ thus the minimum value of $\rho = \frac{-1}{(n-1)}$ & in this case $V(\bar{y}_r) = 0$

7. Make a comparative study of Systematic sampling with that of SRSWOR in terms of intra class correlation coefficient.

Answer:

The relative efficiency of systematic sample mean w.r.t. SRSWOR is give by the expression

E = $\frac{V(\bar{y}_r)}{V(\bar{y})_{wor}}$

$$\frac{V(\bar{y}_r)}{V(\bar{y})_{wor}} = \frac{(\frac{N-1}{Nn})S^2[1+(n-1)\rho]}{(\frac{N-n}{N})\frac{S^2}{n}}$$

$$= \frac{N-1}{N-n}[1+(n-1)\rho]$$

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

Obviously this depends on the value of ρ

$$\frac{V(\bar{y}_r)}{V(\bar{y})_{wor}} = 1$$

$$\Rightarrow \frac{N-1}{N-n}[1+(n-1)\rho] = 1$$

$$\Rightarrow (N-1)[1+(n-1)\rho] = N-n$$

$$\Rightarrow N-1+(N-1)(n-1)\rho - N + n = 0$$

$$\Rightarrow (n-1)+(N-1)(n-1)\rho = 0$$

Divide by (n-1)

$$\Rightarrow 1+(N-1)\rho = 0$$

$$\rho = \frac{-1}{(N-1)}$$

SRSWOR and Systematic Sampling are equally efficient when $\rho = \frac{-1}{(N-1)}$

i.e. when $\rho = \frac{-1}{(N-1)}$, the two methods give the estimates of equal precision.

For $\rho < \frac{-1}{(N-1)}$, the estimate based on Systematic sampling is more efficient.

For $\rho > \frac{-1}{(N-1)}$, it reverses.

However if Rho assumes the minimum possible value that is $\rho = \frac{-1}{(n-1)}$, then variance of the systematic sample mean is zero and consequently the relative efficiency is infinity. Thus in this case reduction in variance of Systematic sample mean over Simple Random sampling without replacement will be 100%. The maximum value of ρ is 1.

If ρ assumes the maximum value then we get the relative efficiency as

E = $\frac{k-1}{nk-1}$

8. What do you mean by a population with linear trend?

Answer:

Consider a hypothetical population where the values of n units are in A.P., that is the values of Y corresponding to the population units are given by $Y_i = a + ib; i = 1...N$ where 'a' and 'b' are constants. If the population units are in the above form we call such a population as population with linear trend.

9. Obtain an expression for the mean of the population with linear trend.

Answer:

The population mean

$$\bar{Y} = \frac{\sum Y_i}{N}$$

$$= \frac{\sum (a+ib)}{N}$$

$$= \frac{Na}{N} + \frac{b}{N}\sum_{i=1}^{N} i$$

$$= a + \frac{b(N+1)}{2}$$

10. What is the population variance for the population variance σ^2 ?

Answer:

The Population variance

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

$$= \frac{1}{N} \sum_{i=1}^{N} [(a+ib) - (a + \frac{b(N+1)}{2})]^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} [(ib - \frac{b(N+1)}{2})]^2$$

$$= \frac{b^2}{N} \sum_{i=1}^{N} [i - \frac{(N+1)}{2}]^2$$

$$= \frac{b^2}{N} \sum_{i=1}^{N} [i^2 + \frac{(N+1)^2}{4} - (N+1)i]$$

$$= \frac{b^2}{N} [\sum_{i=1}^{N} i^2 + \sum_{i=1}^{N} \frac{(N+1)^2}{4} - (N+1) \sum_{i=1}^{N} i]$$

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

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

$$= b^2 (N+1) [\frac{4(2N+1) + 6(N+1) - 12(N+1)}{24}]$$

$$= b^2 (N+1) [\frac{4(2N+1) - 6(N+1)}{24}]$$

$$= b^2 (N+1) [\frac{8N + 4 - 6N - 6}{24}]$$

$$= b^2 (N+1)[\frac{2N-2}{24}]$$

$$= \frac{b^2 (N+1)(N-1)}{12}$$

$$= \frac{b^2 (N^2-1)}{12}$$

11. Derive an expression for population mean square for a population with linear trend.

Answer:

Population mean square,

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

$$= \frac{1}{N-1}\sum_{i=1}^{N}[(a+ib)-(a+\frac{b(N+1)}{2})]^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}[(ib-\frac{b(N+1)}{2})]^2$$

$$= \frac{b^2}{N-1}\sum_{i=1}^{N}[i-\frac{(N+1)}{2}]^2$$

$$= \frac{b^2}{N-1}\sum_{i=1}^{N}[i^2+\frac{(N+1)^2}{4}-(N+1)i]$$

$$= \frac{b^2}{N-1}[\sum_{i=1}^{N}i^2+\sum_{i=1}^{N}\frac{(N+1)^2}{4}-(N+1)\sum_{i=1}^{N}i]$$

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

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

$$= \frac{b^2}{N-1}N(N+1)[\frac{8N+4-6N-6}{24}]$$

$$= \frac{b^2 N(N+1)(N-1)}{(N-1)12}$$

$$= \frac{b^2 N(N+1)}{12}$$

12. When N = nk, prove that $V(\bar{y})_{wor} = (k-1)(nk+1)\frac{b^2}{12}$

Answer:

$V(\bar{y})$ Under SRSWOR is given by

$$V(\bar{y}) = (\frac{N-n}{N-1}) \times (\frac{\sigma^2}{n})$$

$$= (\frac{N-n}{N-1}) \times (\frac{b^2(N^2-1)}{12n})$$

$$= (N-n)(N+1)\frac{b^2}{12n}$$

When N=nk

$$V(\bar{y})_{wor} = (nk-n)(nk+1)\frac{b^2}{12n} = (k-1)(nk+1)\frac{b^2}{12}$$

13. Under Systematic sampling for a population with linear trend prove that systematic sample mean is an unbiased estimator of the population mean.

Answer:

Values of the units in the Systematic sampling with random start ‘r’ is given by

$Y_{rj} = a + (r + jk)b$; j=0,1,...,n-1

$$\overline{y_r} = \frac{\sum_{j=0}^{n-1} y_{rj}}{n} = \frac{\sum_{j=0}^{n-1}(a+(r+jk)b)}{n}$$

Sample mean

$$= a + \frac{\sum_{j=0}^{n-1}(r+jk)b}{n}$$

$$= a + \frac{brn}{n} + \frac{kb}{n}\frac{n(n-1)}{2}$$

$$= a + b(r + \frac{k(n-1)}{2})$$

$$E(\overline{y_r}) = E(a + b(r + \frac{k(n-1)}{2}))$$

$$= \frac{1}{k}\sum_{r=1}^{k}(a + b(r + \frac{k(n-1)}{2}))$$

$$\frac{ak}{k} + \frac{b}{k}\sum_{r=1}^{k} r + b\sum_{r=1}^{k}\frac{(n-1)}{2}$$

=

$$a + \frac{b(k+1)}{2} + bk\frac{n-1}{2}$$

=

$$= a + \frac{b[(k+1)+k(n-1)]}{2}$$

$$= a + \frac{b}{2}(N+1) = \overline{Y}$$

Hence for a population with linear trend also systematic sample mean is an unbiased estimator of the population mean.

14. Derive variance of the systematic sample mean $\overline{y_r}$ for a population with linear trend.

Answer:

$$V(\overline{y_r})_{sys} = E(\overline{y_r} - \overline{Y})^2 = \sum_{r=1}^{k}(\overline{y_r} - \overline{Y})^2 \frac{1}{k}$$

$$= \frac{1}{k}\sum_{r=1}^{k}(a + b(r + \frac{k(n-1)}{2}) - a + \frac{b}{2}(N+1))^2$$

$$= \frac{1}{k}\sum_{r=1}^{k}(b(r + \frac{k(n-1)}{2}) - \frac{b}{2}(N+1))^2$$

$$= \frac{b^2}{k}\sum_{r=1}^{k}(r + \frac{N}{2} - \frac{k}{2} - \frac{N}{2} - \frac{1}{2})^2$$

$$= \frac{b^2}{k}\sum_{r=1}^{k}(r - \frac{k+1}{2})^2 = \frac{b^2(k^2-1)}{12}$$

15. Discuss the efficiency of systematic sampling as compared to SRS for population with linear trend.

Answer:

We know that $V(\bar{y})_{wor} = (k-1)(N+1)\frac{b^2}{12}$

$$V(\overline{y_r})_{sys} = \frac{b^2(k^2-1)}{12}$$

Consider $V(\bar{y})_{wor}$ - $V(\overline{y_r})_{sys}$ = $(k-1)(N+1)\frac{b^2}{12}$ - $\frac{b^2(k^2-1)}{12}$

$$= \frac{b^2(k-1)}{12}[N+1-9k+1)]$$

$$= \frac{b^2(k-1)(N-k)}{12} > 0$$

$V(\bar{y})_{wor}$ - $V(\overline{y_r})_{sys}$ >0

$\Rightarrow V(\overline{y_r})_{sys} \leq V(\bar{y})_{wor}$

Hence, systematic sampling is efficient than SRSWOR when the units in the population are in the linear form.