1. Introduction

Welcome to the series of E-learning modules on Comparison of Systematic Sampling (N equals to n into k) with Simple Random Sampling Without Replacement. In this module, we are going to cover the comparison of Systematic sampling with SRSWOR in terms of population mean square and in terms of intra class correlation coefficient and comparison of the techniques for a population with linear trend.

By the end of this session, you will be able to:

- Explain about Systematic sampling as compared to SRSWOR
- Explain the comparison of Systematic sampling with SRSWOR in terms of Population Mean square
- Explain the comparison of Systematic sampling with SRSWOR in terms of intra class correlation coefficient
- Explain the comparison of Systematic Sampling and SRSWOR for a population with linear trend

Simple random sampling is a basic type of sampling as it can be a component of other more complex sampling methods. The principle of simple random sampling is that every object has the same possibility of being chosen. Simple Random sampling is characterized by the way, in which they are selected.

A sampling technique in which first unit is selected with a help of random numbers and the others are selected automatically according to some pre-designed pattern until the desired sample size is reached is known as systematic random sampling.

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.

Already we proved that Simple Random Sampling Without Replacement is more efficient than that of With Replacement. Now, let us compare Systematic sampling with Simple Random Sampling Without Replacement.

Suppose y bar is the sample mean of Simple Random sample of size n drawn using Without Replacement scheme from a population of size N then its variance is given by,
Variance of y bar is equal to N minus n into S square divided by N into n
Where, S square is a population mean square.

2. Systematic Sampling

Under Systematic sampling,
Variance of yr bar is equal to (N minus 1 by N) into S square minus (n minus 1 by n) into S square systematic
Where, S square systematic is equal to summation r runs from 1 to k, summation i runs from 1 to n (yri minus Yr bar) whole square divided by k into (n minus 1).

Theorem 1:
Systematic sampling is more efficient than SRSWOR if S square systematic is greater than S square
Or prove that,
Variance of yr bar is less than Variance of y bar under SRSWOR if S square systematic is greater than S square.

Proof:
We know that if y bar is the mean of the Simple Random Sample of size n then
Variance of y bar under SRSWOR is equal to N minus n into S square divided by N into n
Consider variance of yr bar less than variance of y bar under SRSWOR
Which implies,
(N minus 1 by N) into S square minus (n minus 1 by n) into S square systematic is less than N minus n into S square divided by N into n
Which implies,
(N minus 1 by N) into S square minus N minus n into S square divided by N into n is less than (n minus 1 by n) into S square systematic.

Which implies S square into [(n minus 1) minus (N minus n) by n]divided by N is less than (n minus 1 by n) into S square systematic.
Which implies, S square by N into [N into n minus n minus N plus n by n] is less than (n minus 1 by n) into S square systematic
Which implies S square into [n minus 1 by n] is less than (n minus 1 by n) into S square systematic
Which implies, S square is less than S square systematic.

Which implies,
S square systematic is greater than S square
Hence, the systematic sampling gives more precise estimate of the population mean as compared with SRSWOR if and only if,
Variance of y bar under SRSWOR minus Variance of yr bar systematic is greater than zero which implies S square systematic is greater than S square.

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 to Cluster Sampling 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, then the successive units in the sample will be repeating more or less the same information.

3. Comparison of SRSWOR with Systematic Sampling in Terms of Intra Class Correlation Coefficient – Part 1

Comparison of SRSWOR with systematic sampling in terms of intra class correlation coefficient
Suppose y bar is the sample mean of Simple Random Sampling of size n drawn using Without replacement scheme from a population of size N then its variance is given by,
Which implies variance of y bar under SRSWOR is equal to (N minus n by N)into (S square by n)
Where,
S square is a population mean square.

Consider
Variance of yr bar is equal to Expected value of (yr bar minus Expected value of yr bar) whole square Which is equal to Expected value of yr bar minus Y bar whole square
Which is equal to Summation , r runs from 1 to k(yr bar minus Y bar) whole square into 1 by k
Which is equal to summation r runs from 1 to k (1 by n) into (summation, i runs from 1 to n yri) minus (Y bar) whole square divided by k.

Variance of yr bar is equal to summation r runs from 1 to k, summation i runs from 1 to n(yri minus Y bar) whole square divided by n square into k
Which is equal to summation r runs from 1 to k, [summation i runs from 1 to n(yri minus Y bar) whole square plus summation (i not equal to j) (yri minus Y bar) into (yrj minus bar)] divided by n square into k.
Which is equal to [summation r runs from 1 to k, summation i runs from 1 to n (yri minus Y bar) whole square plus summation r runs from 1 to k, summation (i not equal to j) (yri minus Y bar) into (yrj minus bar)] divided by n square into k.
Variance of yr bar is equal to[summation r runs from 1 to k, summation i runs from 1 to n(yri minus Y bar) whole square plus summation r runs from 1 to k, summation (i not equal to j)

(yri minus Y bar) into (yrj minus bar)] divided by n square into k. Call this as 1.

Suppose row is the intra-class coefficient then,
Rho is equal to Expected value of (yri minus Y bar) into (yrj minus Y bar) divided by Expected value of (yri minus Y bar) whole square
Which is equal to summation r runs from 1 to k, summation i not equal to j (Yri minus Y bar) into (Yrj minus Y bar) divided by k into n into (n minus 1) whole divided by summation r runs from 1 to k, summation i equal to 1 to n (Yri minus Y bar) whole square divided by k into n.
Rho is equal to summation r runs from 1 to k, summation i not equal to j (Yri minus Y bar) into (Yrj minus Y bar) divided by summation r runs from 1 to k, summation i equal to 1 to n (Yri minus Y bar) whole square into (n minus 1)
Rho is equal to summation r runs from 1 to k, summation i not equal to j (Yri minus Y bar) into (Yrj minus Y bar) divided by n into k into sigma square into (n minus 1)
Because sigma square is equal to summation r runs from 1 to k, summation i not equal to j (Yri minus Y bar) whole square divided by n into k.
Which implies summation r runs from 1 to k, summation i not equal to j (Yri minus Y bar) into ((minus Y bar) is equal to rho into n into k into sigma square into (n minus 1). Call this as 2.

By substituting equation 2 in equation 1 we get,
Variance of yr bar is equal to [(n into k minus 1) into S square plus (n into k minus 1) into (n minus 1) into Rho into S square] by n square into k
Variance of yr bar is equal to [(n into k minus 1) into S square by n square into k into [1 plus (n minus 1) into Rho]

Variance of yr bar is equal to [(N minus 1) into S square by N into n into [1 plus (n minus 1) into Rho]

4. Comparison of SRSWOR with Systematic Sampling in Terms of Intra Class Correlation Coefficient – Part 2

Since variance of yr bar is never less than zero Rho cannot be less than minus 1 by (n minus 1) thus the minimum value of rho is equal to minus 1 by (n minus 1) & in this case variance of yr bar is equal to zero
The relative efficiency of systematic sample mean with respect to SRSWOR is give by the expression E is equal to Variance of yr bar by Variance of y bar SRSWOR
Variance of yr bar by Variance of y bar SRSWOR is equal to (N minus 1 by N into n) into S square into [1 plus (n minus 1) into Rho] by (N minus n into S square by N into n.

Variance of yr bar by Variance of y bar SRSWOR is equal to (N minus 1 by N minus n) into [1 plus (n minus 1) into Rho]
Which is equal to (N minus 1 by N minus n) plus (N minus 1) into (n minus 1) into Rho by (N minus n)
Obviously this depends on the value of Rho.

Variance of yr bar by Variance of y bar SRSWOR is equal to 1
Which implies (N minus 1 by N minus n) into [1 plus (n minus 1) into Rho] is equal to 1
Which implies (N minus 1) into [1 plus (n minus 1) into Rho] which is equal to N minus n
Which implies (N minus 1) plus (N minus 1) into (n minus 1) into Rho minus N plus n is equal to zero
Which implies (n minus 1) plus (N minus 1) into (n minus 1) into Rho is equal to zero.

Divide by (n minus 1)
Which implies 1 plus (N minus 1) into Rho is equal to zero
Rho is equal to minus 1 by (N minus 1)

SRSWOR and Systematic Sampling are equally efficient when Rho is equal to minus 1 by (N minus 1)
That is when Rho is equal to minus 1 by (N minus 1) the two methods give the estimates of equal precision.
For Rho less than minus 1 by (N minus 1) the estimate based on Systematic sampling is more efficient.
For Rho greater than minus 1 by (N minus 1) it reverses.

However, if Rho assumes the minimum possible value that is Rho is equal to minus 1 by (n minus 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 hundred percent. The maximum value of Rho is 1.
If Rho assumes the maximum value then we get the relative efficiency as
E is equal to k minus 1 by n into k minus 1.

5. Population with Linear Trend

Population with Linear Trend:
It was found in some of the experiments that efficiency of the systematic sampling can be increased if there is a trend in the values of population units. For example, if we arrange the villages in order of geographical area, the efficiency of systematic sampling with respect to SRSWOR can be illustrated as follows:

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 is equal to a plus i into b; i is equal to 1,2, up to N where 'a' and 'b' are constants.

Now the population mean,
Y bar is equal to summation Yi by N
Equal to summation (a plus ib) by N
Which is equal to a plus b into (N plus 1) by 2.

The Population variance sigma square is equal to:
Summation over I runs from 1 to N (Yi minus Y bar) whole square by N
Which is equal to summation over I runs

from 1 to N, [(a plus ib) minus (a plus b into (N plus 1) by 2] whole square by N
Which is equal to b square by N into summation over I runs from 1 to N, [i minus (N plus 1) by 2] whole square
Which is equal to b square into (N square minus 1) by twelve.

Variance of y bar under SRSWOR is given by (N minus n) into sigma square by (N minus 1) into n
which is equal to (N minus n) into (N plus 1) into b square by twelve into n
When N is equal to nk
Variance of y bar WOR is equal to (nk minus n) into (nk plus 1) into b square by 12n which is equal to (k minus 1) into (nk plus 1) into b square by twelve.

Values of the units in the Systematic sampling with random start 'r' is given by:
Y rj is equal to a plus (r plus j into k) into b; j is equal to zero,1 up to n minus 1
Sample mean yr bar is equal to summation yrj, j runs from zero to n minus 1 by n
Which is equal to 1 by n summation over j (a plus (r plus j into k) into b)
Which is equal to a plus summation over j (r plus j into k) into b by n
Which is equal to a plus brn by n plus kb by n into n into (n minus 1) by 2
Which is equal to a plus b into (r plus k into (n minus 1) by 2).

Expected value of (yr bar) is equal to Y bar, a population mean
Variance of yr bar systematic is equal to Expected value of (yr bar minus Y bar) whole square
Which is equal to summation over r, (yr bar minus Y bar) whole square by k

Which is equal to summation over r,[(a plus b into (r plus k into (n minus 1) by 2) minus a plus b by 2 into (N plus 1)] whole square by k.

Variance of yr bar systematic is equal to summation over r,[(b into (r plus k into (n minus 1) by 2) minus b by 2 into (N plus 1)] whole square by k
Which is equal to b square by k summation over r, (r plus N by 2 minus k by 2 minus N by2 minus half) whole square
Which is equal to b square by k summation over r, (r minus (k plus 1)by 2) whole square
Which is equal to b square into (k square minus 1) by twelve.

Consider Variance of (y bar) WOR minus Variance of yr bar) sys
which is equal to (k minus 1) into (N plus 1) into b square by twelve minus b square into (k square minus 1) by twelve
Which is equal to b square into (k minus 1) by twelve into [N plus 1 minus (k plus 1)]
Which is equal to b square into (k minus 1) into (N minus k) by twelve is greater than zero
Which implies Variance of(yr bar) systematic is less than or equal to Variance of (y bar) WOR
Hence, when we have a population with linear trend, Systematic is more efficient than SRSWOR.

Here's a summary of our learning in this session:

- Compared Systematic sampling with SRSWOR in terms of population mean square
- Compared Systematic sampling with SRSWOR in terms of intra class correlation coefficient
- Discussed population with linear trend
- Compared Systematic Sampling with SRS for a population with linear trend