1. Introduction

Welcome to the series of e-learning modules on Comparison of Systematic Sampling, (N equals to n into k) with Stratified Random Sampling . In this module we are going cover the comparison of systematic sampling with stratified sampling in terms of intra-class correlation coefficient, and, comparison of the simple random sampling, stratified sampling and systematic sampling techniques for a population with linear trend.

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

- Differentiate Systematic Sampling from Stratified Sampling
- Explain the two sampling methods in terms of intra-class correlation coefficient
- Compare Simple Random Sampling Without Replacement, Stratified Sampling and Systematic Sampling for a population with linear trend

Systematic Random Sampling is a method of probability sampling in which the defined target population is ordered and the sample is selected according to position using a skip interval.
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.

Systematic Random Sampling is a method of probability sampling in which the defined target population is ordered and the sample is selected according to position using a skip interval
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.

The initial sampling unit is randomly selected or established on the ground. All other sample units are spaced at uniform intervals throughout the area sampled, in which sampling units are easy to locate.
We have already proved that Stratified Random Sampling is more efficient than Simple Random Sampling Without Replacement.
In the slides to follow, we shall compare Systematic Random Sampling with Stratified Random Sampling

Let us now regard the population ‘N is equal to n into k; units to be divided into n strata corresponding to n columns of the table.

Table 1

Random Start	Units in the Sample	Probability	Mean
1	1, 1+k, …, 1+jk, …, 1+(n-1)k	1/k	y1 bar
2	2, 2+k, …, 2+jk, …, 2+(n-1)k	1/k	y2 bar
.			
.			
i	i, i+k, …., i+jk, …, i+(n-1)k	1/k	yi bar
.			
k	k, 2k,…, (1+j)k, …., nk	1/k	yk bar

In the table we have the first column, random start as 1,2,etc...i,k. In the Units in sample column we write the values are 1, (1 plus k), etc (1 plus jk) etc (1plus ‘n minus 1’k) and similarly we fill in the rows for 1, i and k.
In the third column ‘Probability’, we write the probability as ‘1 by k’ for all the columns and in the Mean column we mention the corresponding mean as ‘Y1 bar’, ‘Y2 bar’ and so on to ‘Yi bar’ and ‘Yk bar’

Thus ‘k’ rows of the table gives ‘k’ systematic samples.
The columns of the above table are referred to as ‘n strata’
Suppose that one unit is drawn randomly from each stratum, thus giving us a Stratified random sample of size ‘n’.
Hence here: ‘h is equal to 1,2, etc ,n’ and ‘i is equal to 1,2 etc, k’.
Then Mean of the hth stratum, ‘Yh dot bar’ is equal to ‘summation, i runs from to 1 to k’, Yhi by k.
Population Mean, ‘Y bar’ is equal to ‘summation, h runs from 1 to n’, ‘summation i runs from 1 to k’, ‘Yhi by nk’
which is equal to ‘summation h runs from 1 to n’, ‘Yh dot bar by n’.

2. Stratum Mean Square

Stratum mean square

'Sh square' is equal to 'summation , i runs from 1 to k', 'Yhi minus Yh dot bar whole square' by 'k minus 1'.

'S square wst' is the pooled mean square between units within strata. 'S square wst' is equal to 'summation h runs from 1 to n', 'summation , i runs from 1 to k', 'Yhi minus Yh dot bar whole square' by 'n into 'k minus 1''.

Rho wst is the correlation coefficient between deviations from stratum means of pairs of items that are in the systematic sample. Thus

Rho wst is equal to 'Expected value of ('yhi' minus 'y-bar h-dot') into ('yh-dash i' minus 'y-bar h dash dot') divided by 'Expected value of ('yhi' minus 'y-bar h dot') whole square'.

Rho wst is equal to 'summation i, runs from 1 to k', 'summation h not equal to h dash runs from 1 to n', '(yhi minus y-bar h-dot)into ('yh dash i' minus 'y-bar h-dash dot')' divided by 'k into n into (n minus 1)' ,

divided by 'summation, i runs from 1 to k', 'summation h runs from 1 to n' '(yhi minus y-bar h-dot) whole square' by 'n into k'

Rho wst is equal to 'summation i, runs from 1 to k', 'summation h not equal to h dash runs from 1 to n', (yhi minus y-bar h-dot) into(yh-dash i minus y-bar h-dash dot) divided by 'n into (n minus 1)' into (k minus 1) into S square wst.

Systematic Sampling Vs. Stratified sampling

Variance of (y bar st) is equal to summation, h runs from 1 to n, '1 by nh' minus '1 by Nh' into Wh square into Sh square

But Nh is equal to k; nh is equal to 1, where h is equal to 1,2,etc,n and

Wh is equal to 'Nh by N' which is equal to 'k by nk' which is equal to '1 by n'

Variance of 'y bar st', is equal to summation, h runs from 1 to n, 1 minus '1 by k' into 'Sh square by n square

Variance of 'y bar st' is equal to 'k minus 1' by 'n square into k' into 'summation, h runs from 1 to n, Sh square'

Variance of 'y bar st' is equal to 'k minus 1' by 'n square into k' into summation, 'h runs from 1 to n', 'summation i runs from 1 to k', 'Yhi minus Y-bar h-dot whole square' by 'Nh minus 1'.

Variance of 'y bar st' is equal to 'k minus 1' by 'n square into k' into summation, 'h runs from 1 to n', 'summation i runs from 1 to k', ' Yhi minus Y-bar h-dot whole square' by 'k minus 1'.

Variance of 'y bar st' is equal to 'summation, h runs from 1 to n', 'summation i runs from 1 to k', Yhi minus Y-bar h-dot whole square' by 'n square into k'.

Variance of 'y bar st' is equal to 'k minus 1' by 'n into k' whole multiplied by 'S square wst'. Call this as equation 1.

But, 'variance of yr bar systematic' is equal to to 'k minus 1' by 'n into k' whole multiplied by 'S square wst' into '1' plus 'n minus 1' into 'Rho wst'. Call this as equation 2

Comparing equations (1) and (2) we get,

E is equal to 'Variance of y bar st' divided by variance of 'y bar systematic' which is equal to '1' by '1 plus 'n minus 1' into Rho wst'.

Thus, the relative efficiency of systematic sampling over stratified random sampling

depends on the values of Rho-wst and nothing can be concluded in general.
If Rho-wst is greater than zero, then E is less than 1, and thus, in this case stratified sampling will provide a better estimate of population mean 'Y bar'. However, if Rho-wst is equal to zero, then E is equal to 1, and consequently both systematic sampling and stratified sampling provide estimates of population mean with equal precision.

Population with Linear Trend:

If the population consists solely of a linear trend, it is fairly easy to guess the nature of the results.

It was observed 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.

That is, for example if we arrange the villages in order of geographical area, the efficiency of systematic sampling with respect to stratified random sampling and SRSWOR can be illustrated as follows:

3. Systematic Sampling V/s Stratified Sampling V/s SRSWOR for a Population of Linear Trend

Systematic Sampling V/s Stratified Sampling V/s SRSWOR for a Population of Linear Trend
If the population consists of a linear trend then:
(Variance of ‘y bar st’) is less than or equal to (variance of ‘yr bar systematic’) which is less than or equal to (variance of ‘y bar SRSWOR’).

Consider a hypothetical population where the values of n units has a linear trend.
That is, the values of Y corresponding to the population units are given by:
‘Yi’ is equal to i, ‘i is equal to 1,2, etc,N’.
Now the population mean ‘Y bar’ is equal to ‘summation Yi by N’
‘Equal to summation i by N’
‘Which is equal to (N plus 1) by 2’

The Population Mean square S square is equal to:
‘Summation i, runs from 1 to N’, ‘Yi minus Y bar whole square’ by ‘N minus 1’.
Which is equal to ‘Summation i, runs from 1 to N’, ‘i minus (N plus 1) by 2 whole square’ by ‘N minus 1’
Which is equal to ‘N’ into ‘(N plus 1) into (N minus 1)’ by ‘(N minus 1) into twelve’
Which is equal to ‘N’ into ‘(N plus 1) by twelve’.

For a population with N units, we obtain the population mean square as:
‘N’ into ‘(N plus 1) by twelve’
Hence, for the hth stratum with Nh = k observations in each stratum the population mean square for the hth stratum is:
Sh square is equal to ‘k into (k plus 1)’ by ‘twelve’.
Variance of ‘y bar st’ is equal to ‘k minus 1’ by ‘n square into k’ whole multiplied by ‘summation h runs from 1 to n’, ‘Sh square’.
Variance of ‘y bar st’ is equal to ‘k minus 1’ by ‘n square into k’ whole multiplied by ‘summation h runs from 1 to n’, ‘k into (k plus 1)’ by twelve
which is equal to ‘k minus 1’ into ‘n into k’ into (k plus 1) divided by ‘n square into k into twelve’
which is equal to ‘k square minus 1’ by ‘twelve into n’ . Call this equation (1).
Variance of ‘y bar’ under Simple Random Sampling Without Replacement is given by:
Variance of ‘y bar’ is equal to (N minus n) by (N minus 1) into ‘sigma square by n’
Which is equal to (N minus n) by (N minus 1) into ‘N square minus 1’ by ‘twelve into n’
Which is equal to (N minus n) into (N plus 1) by ‘twelve into n’.
When ‘N’ is equal to ‘n into k’,
Variance of ‘y bar’ under Simple Random Sampling Without Replacement is equal to (‘n into k’ minus n) into (‘n into k’ plus 1) by ‘twelve into n’.
Which is equal to (k minus 1) into (‘n into k’ plus 1) by twelve. Call this as equation (2).

Values of the units in the systematic sampling with random start ‘r’ is given by
‘Yrj’ is equal to ‘r plus j into k’ , where j is equal to ‘zero, 1, etc, n minus 1’
Sample mean ‘y r bar’ is equal to ‘summation j runs from zero to n minus 1’, ‘yrj’ by ‘n’
Which is equal to ‘summation j runs from zero to n minus 1’, ‘r’ plus ‘j into k’ by ‘n’
Which is equal to ‘nr by n’ plus ‘k into n’ into ‘n minus 1’ by ‘n into 2’
Which is equal to ‘r’ plus ‘k into (n minus 1)’ by ‘2’.
Expected value of ‘yr bar’ is equal to ‘Expected value of ‘r’ plus ‘k into (n minus 1)’ by 2’
Which is equal to ‘1 by k’ into ‘summation r runs from 1 to k’, ‘r plus k into (n minus 1) by 2’
Which is equal to ‘summation r runs from 1 to k’, ‘r by k’ plus ‘summation r runs from 1 to k’, (n minus 1) by 2
Which is equal to ‘(k plus 1) by 2’ plus ‘k into (n minus 1) by 2’
Which is equal to ‘(k plus 1)’ plus ‘k into (n minus 1)’ by 2
Which is equal to ‘(N plus 1)’ by 2
Which 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 (‘yr bar’ minus ‘Y bar’ whole square) into ‘1 by k’
Which is equal to ‘1 by k’ into ‘summation r runs from 1 to k’, ‘r’ plus ‘k into (n minus 1) by 2’ minus ‘(1 by 2) into ‘(N plus 1) whole square’.
Which is equal to ‘1 by k’ into ‘summation r runs from 1 to k’, (‘r’ plus ‘N by 2’ minus ‘k by 2’ minus ‘N by 2’ minus ‘1 by 2’ whole square).
Which is equal to ‘1 by k’ into ‘summation r runs from 1 to k’, (‘r’ minus ‘k plus 1’ by 2 whole square)
which is equal to ‘k square minus 1’ by twelve.
Call this as equation (3)

From equations (1), (2) and (3) we get
Variance of ‘y bar st’ is-to ‘variance of y r bar systematic’ is-to ‘variance of y bar SRSWOR’
Which is equal to
‘K square minus 1’ by ‘twelve into n’ is-to ‘K square minus 1 by twelve’ is-to ‘(k minus 1) into (nk plus 1) by twelve’
Which is equal to ‘k plus 1 by n’ is-to (k plus 1) is-to (n into k plus 1)
Which is approximately ‘1 by n’ is-to ‘one’ is-to ‘n’

Which implies ‘Variance of y bar st’ is less than or equal to ‘variance of y r bar systematic’ is less than or equal to ‘variance of y-bar SRSWOR’.
Equality occurs only when ‘n=1’.
Thus, for removing the effect of linear trend, suspected or unsuspected, the systematic sample is much more effective than the simple random sample.
But at the same time it is less effective than stratified random sample.
Thus, if the population is a subject of a linear trend, then stratified random sampling is most effective, with systematic sampling as the next best, in eliminating the effect of the linear trend.

4. Comparison of Systematic Sampling with Stratified Random Sampling

On comparison of systematic sampling with stratified random sampling or Simple Random Sampling Without Replacement, the relative efficiency of the systematic sampling over stratified random sampling or Simple Random Sampling Without Replacement 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 the knowledge of the structure of the population, no hard and fast rules can be laid down and no situations can be pin-pointed where the use of systematic sampling is to be recommended.

Thus, the performance of systematic sampling in relation to that of stratified random sampling or simple random sampling without replacement is greatly dependent on the properties of the population.
There are populations for which systematic sampling is extremely precise and others for which it is less precise than simple random sampling.
For some populations and some values of n, $v(\overline{y})sys$ may even increase when a larger sample is taken - a starting departure from good behavior.
Thus, it is difficult to give general advice about the situation in which systematic sampling is to be recommended.
A knowledge of the structure of the population is necessary for its most effective use.

<u>Population in ‘Random’ Order</u>
Systematic sampling is sometimes used for its convenience, in population in which the numbering of the units is effectively random. This is so in sampling from a file arranged alphabetically by surnames, if the item that is being measured has no relation to the surname of the individual. There will then be no trend or stratification in yi as we proceed along the file and no correlation between neighbouring values

In this situation we would expect Systematic sampling to be essentially equivalent to Simple Random sampling and to have the same variance. For any single finite population with given values of n and k this is not exactly true. Because variance of the sample mean under Systematic sampling which is based on only k degrees of freedom is rather erratic when k is small and may turn out to be either smaller or greater than variance of y bar under Simple Random sampling.

5. Conditions Under Which Systematic Sampling is Recommended

Systematic samples are convenient to draw and to execute. In the light of all the results proved by now, Systematic sampling can safely be recommended in the following situations

1) Where the ordering of the population essentially random or contains atmost a mild stratification. Here Systematic sampling is used for convenience within a little expectation of gain in precision
2) Where stratification with numerous strata is employed and independent systematic sample is drawn from each stratum. The effects of hidden periodicities tend to cancel out in this situation and an estimate of error that is known to be an overestimate can be obtained.
3) For sampling population with variations of a continuous type provided that an estimate of the sampling error is not regularly required

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

- Compared Systematic sampling with Stratified random sampling
- Compared Systematic sampling with SRSWOR and Stratified random sampling for a population with linear trend
- Discussed the situations where it is recommended to use Systematic Sampling .