1. Introduction

Welcome to the series of E-learning modules on Estimation of Population Mean and Standard Error of the Estimator. In this module, we are going to cover the unbiased estimate of the population mean, Variance of the estimator, Concept of intra class correlation and Standard Error of the estimator.

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

- Explain about unbiased estimator of the population mean under systematic sampling
- Explain the variance of the systematic sample mean in terms of population mean square
- Explain the variance of the systematic sample mean in terms of intra class correlation
- Explain the standard error of the estimator

Systematic sampling 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 are selected automatically according to some pre-designed pattern is known as systematic random sampling. This is a commonly employed technique if the complete and up-to -date list of the sampling units is available.

Suppose that the population is linear in order, in some way such that number can refer to units. Further let N be expressible in the form of N=nk and let the selected random number be r (less than k), k being called as a sampling interval.
In this procedure, a sample comprise of the unit r, r plus k, r plus 2k… r plus(n minus 1) k.

Hence, systematic sampling is a procedure of selecting every kth unit starting from the unit corresponding to a number r selected at random from 1 to k.
k being an integer nearest to N by n, that is k is equal to N by n which implies N is equal to n into k.

NOTE:

1) The number r is known as the random start. If r is a random start then a systematic sample of size n consists of the units r, r plus k, r plus 2k, r plus 3k, up to r plus (n minus 1)k.
2) Under systematic sampling, we can have k possible samples and probability of selecting a sample is 1 by k.
3) Under systematic sampling let yri be the ith unit corresponding to the rth sample r is equal to 1, 2, up to k; i is equal to 1, 2, up to n.

2. Notations

<u>NOTATIONS:</u>
Mean of the rth systematic sample yr bar is equal to summation, I runs from 1 to n yri by n which implies summation yri is equal to n into yr bar
Under systematic sampling
The Population Mean
Y bar is equal to summation, r runs from 1 to k, summation i runs from 1 to n, Yri divided by nk which is equal to summation , r runs from 1 to k, yr bar by k. [Covers whole population].

Sigma square 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 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 divided by N.

S square 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 into k minus 1)
Which 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 minus 1).

<u>Estimation of the Population Mean</u>
<u>Theorem 1:</u> Under Systematic sampling yr bar is an unbiased estimator of the population mean Y bar, that is, Expected value of yr bar is equal to Population Mean Y bar.

<u>Proof:</u> Consider

Expected value of yr bar is equal to summation over r, yr bar into probability of yr bar
Which is equal to summation over r, yr bar into (1 by k)
Which is equal to summation over r (summation i runs from 1 to n yri by n) into (1 by k).
Expected value of yr bar is equal to summation over r, summation over i, y ri by n into k which is equal to Y bar
Therefore, Expected value of yr bar is equal to Y bar
Thus, if N is equal to nk, the systematic sample mean provides an unbiased estimate of the population mean.

3. Variance of the Estimated Mean

Variance of the Estimated Mean
<u>Theorem 2:</u> Under Systematic sampling,
Variance of yr bar is equal to (N minus 1) into S square by N minus (n minus 1) into S square systematic by n
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 by k into (n minus 1) is the mean square among units which lie within the same systematic sample.

The denominator of this variance , k into (n minus 1) is constructed by the usual rules in the analysis of variance each of the k samples contributes (n minus 1) degrees of freedom to the sum of squares in the numerator.

<u>Proof:</u>
S square 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 into k minus 1)
Which is equal to summation, r runs from 1 to k, summation i runs from 1 to n, (yri minus Yr bar plus Yr bar minus Y bar) whole square divided by (n into k minus 1)
Which is equal to summation, r runs from 1 to k, summation i runs from 1 to n, [(yri minus Yr bar) whole square plus (Yr bar minus Y bar) whole square plus 2 into (yri minus Yr bar) into (Yr bar minus Y bar) divided by (n into k minus 1).

S square is equal to summation, r runs from 1 to k, summation i runs from 1 to n, (yri minus Yr bar) whole square plus summation, r runs from 1 to k, summation i runs from 1 to n, (Yr bar minus Y bar) whole square plus 2 into summation r runs from 1 to k (Y r bar minus Y bar) into summation i runs from 1 to n (Yr i minus Y r bar) divided by (n into k minus 1)
Which is equal to summation, r runs from 1 to k, summation i runs from 1 to n, (yri minus Yr bar) whole square plus summation, r runs from 1 to k, summation i runs from 1 to n, (Yr bar minus Y bar) whole square whole divided by (n into k minus 1)
(nk minus 1) into S square is equal to summation, r runs from 1 to k, summation i runs from 1 to n, (yri minus Yr bar) whole square plus summation, r runs from 1 to k, summation i runs from 1 to n, (Yr bar minus Y bar) whole square.
(nk minus 1) into S square is equal to k into (n minus 1) S square systematic plus summation, r runs from 1 to k, n into (Yr bar minus Y bar) whole square. Call this as 1.

We know that,
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)
k into variance of (yr bar) is equal to summation r runs from 1 to k, (yr bar minus Y bar) whole square . Call this as 2.

By substituting equation 2 in equation 1:
(n into k minus 1) into S square is equal to k into (n minus 1) S square systematic plus n into k into Variance of yr bar
Which implies (n into k minus 1) S square minus k into (n minus 1) S square systematic whole divided by n into k is equal to Variance of yr bar
Which implies (n into k minus 1) into S square by n into k minus k into (n minus 1) into S square systematic by nk is equal to Variance of yr bar

Which implies (N minus 1) into S square by N minus (n minus 1) S square systematic by n is equal to Variance of yr bar
Which implies Variance of yr bar is equal to (N minus 1) into S square by N minus (n minus 1) S square systematic by n
Hence proved.

4. Variance of the Estimated Mean - Theorem 4

Theorem 4:
Variance of sample mean under systematic sampling in terms of intra class correlation coefficient or
Prove that :
Variance of yr bar is equal to sigma square into (1 plus (n minus 1) into Rho by n
Where, Rho is an intra-class correlation coefficient between the units of the same systematic sample.
Proof:
The intra-class coefficient Rho
Rho is equal to Expected value of (Yri muns 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, runs from 1 to n(Yri muns Y bar) into (Yrj minus Y bar) into (1 by k into n into (n minus 1) whole divided by (summation, r runs from 1 to k, summation i runs from 1 to n(Yri muns 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, runs from 1 to n(Yri muns Y bar) into (Yrj minus Y bar) divided by summation r runs from 1 to k, summation i runs from 1 to n(Yri muns 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, runs from 1 to n(Yri muns 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 runs from 1 to n(Yri muns Y bar) whole square divided by n into k
Which implies,
Summation r runs from 1 to k, summation i not equal to j, runs from 1 to n(Yri muns Y bar) (Yrj minus Y bar) is equal to Rho into n into k into sigma square into (n minus 1). Call this as 1.

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

Variance of yr bar is equal to summation over r, [into (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 over r, [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 over i, summation over j, i is not equal to j, (yri minus Y bar) into (yrj minus Y 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 over i, and j, i is not equal to j, (yri minus Y bar) into (yrj minus Y bar)] divided by n square into k

Variance of yr bar is equal to [n into k into sigma square plus n into k into (n minus 1) into Rho into sigma square] divided by n square into k. [from equation 1]

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

Thus, we see that a positive intraclass correlation between the units of the same sample inflates the variability of the estimate. Even a small positive correlation may have a larger effect. Due to the multiplier (n minus 1), the increase is quite significant even for small values of Rho.

We have, variance of yr bar is greater than or equal to zero which implies Rho is greater than or equal to minus 1 by n minus 1
Thus, the minimum value of Rho is minus 1 by n minus 1 and in this case variance of yr bar is equal to zero.

Note: Variance of sample mean under systematic sampling in terms of intra class correlation coefficient can also be obtained as:
Variance of yr bar is equal to (n into k minus 1) into S square into (1 plus (n minus 1) into Rho divided by n square into k
where Rho is an intra-class correlation coefficient between the units of the same systematic sample.

In this case Rho can be expressed as:

Rho is equal to summation r runs from 1 to k, summation i not equal to j, runs from 1 to n(Yri

muns Y bar) into (Yrj minus Y bar) divided by (n into k minus 1) into S square into (n minus 1)
Because S square is equal to summation r runs from 1 to k, summation i , runs from 1 to n(Yri muns Y bar) whole square divided by n into k minus 1.

Since variance of a systematic sample mean is expressed in terms of S2 , we can relate it to the variance for a simple random sample.
Systematic sample is precise when units within the same sample are heterogeneous and imprecise when units within the same sample are homogeneous. 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. Standard Error of the Estimator

Standard Error of the estimator

The formulae for the standard errors of the estimated population mean and total are primarily used for three purposes:

1) To compare the precision obtained by Systematic sampling with that given by other methods of sampling.
2) To estimate the size of the sample needed in a survey that is being planned.
3) To estimate the precision actually attained in a survey that has been completed.

Standard error can be defined as the Standard deviation of the distribution.

For example, suppose t is an estimator then Standard Error of t is equal to square root of variance of t.

a) Standard error of estimated population mean

   Standard error of yr bar is equal to square root of variance of yr bar, which is equal to square root of (N minus 1) into S square by N minus n minus 1 into S square systematic by n.

   Where S square systematic is equal to summation r is equal to 1 to k, summation i is equal to 1 to n,(yri minus Yr bar) whole square divided by k into (n minus 1).

b) Standard Error if the estimated population mean can also be expressed in terms of intra-class correlation

   Standard error of yr bar is equal to square root of variance of yr bar which is equal to square root of Sigma square by n into (1 plus (n minus 1) into Rho

   OR in terms of S square

   Standard error of yr bar is equal to square root of variance of yr bar which is equal to square root of (n into k minus 1) into S square by n square into k into (1 plus (n minus 1) into Rho).

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

- Illustrated way of drawing systematic random samples
- Discussed unbiased estimate of the population mean
- Obtained variance of the estimator of the population mean
- Addressed variance of the estimator of the population mean in terms of intra class correlation coefficient
- Discussed the Standard Error of the estimator of the population mean