1. Introduction

Welcome to the series of e-learning modules on Comparison between Stratified Sampling and Simple Random Sampling without Replacement in terms of precision and cost. In this module we are going cover the variance of the estimator of population mean under Stratified random sampling using Proportional and Optimum Allocation, Comparison of Stratified Sampling using two techniques with Simple Random Sampling without Replacement.

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

- Variance of estimator of population mean under stratified sampling with proportional allocation
- Variance of estimator of population mean under stratified sampling with optimum or Neyman's allocation
- Comparison of proportional allocation and optimum allocation under stratified sampling with Simple Random Sampling without Replacement

Basically in a stratified sampling procedure, the population is first partitioned into disjoint classes, called the strata, which together are exhaustive.
Thus each population element should be within one and only one stratum.
Then a simple random sample is taken from each stratum.

The sampling effort may either be a proportional allocation (each simple random sample would contain an amount of variates from a stratum which is proportional to the size of that stratum) or according to optimal allocation, where the target is to have a final sample with the minimum variability possible.

The efficiency of any sampling design is measured in terms of the variance of an estimator.
A sampling design with least variance is said to be more efficient than the others.
Hence we obtain the variances of an estimator of the population mean under proportional and optimum allocations first and then compare the techniques.

If the sample size in the h^{th} stratum is nh is proportional to Nh, then the sample is said to have been selected under Proportional Allocation.

In this method the sample allocation to the h^{th}-stratum is made by

Nh is equal to n into Nh by N

The efficiency of any sampling design is measured in terms of the variance of an estimator.

A sampling design with least variance is said to be more efficient than the others.

Hence we obtain the variances of an estimator of the population mean under proportional and optimum allocations first and then compare the techniques.

2. **Theorem**

The first theorem here states that:
The Variance of the estimated population mean under stratified random sampling when proportional allocation is used is given by:
Variance of Variance of y bar st under proportional allocation is equal to summation, h runs from 1 to k, Wh into Sh square by n minus summation, h runs from 1 to k, Wh into Sh square by N

Proof:
We know that
Variance of y bar st is equal to summation, h runs from 1 to k, Wh square into Sh square by nh minus summation, h runs from 1 to k, Wh square into Sh square by Nh.
Under Proportional Allocation we have,
nh is equal to Nh into n by N and Wh is equal to Nh by N

Then, Variance of y bar st under Proportional Allocation is equal to summation, h runs from 1 to k, (Nh by N) whole square into Sh square by (Nh by N) into n, minus summation h runs from 1 to k, (Nh by N) whole square into Sh square by Nh.
Variance of y bar st under Proportional Allocation is equal to summation , h runs from 1 to k, (Nh by N) into Sh square by n, minus summation h runs from 1 to k, (Nh by N) into Sh square by N
Which is equal to summation, h runs from 1 to k, Wh into Sh square by n minus summation, h runs from 1 to k, Wh into Sh square by N

If the sample size in the h^{th} stratum is directly proportional to the product of the population size in the h^{th} stratum and the population root mean square in the h^{th} stratum i.e., if nh is proportional to Nh into Sh,
Then, the sample is said to have been selected under Optimum allocation or Neyman's allocation.

Therefore, under Optimum Allocation, an expression for the sample size from h^{th} stratum is given by nh is equal to n into Nh into Sh by summation, h runs from 1 to k (nh into Sh)

Theorem 2:

Variance of the estimated population mean under Stratified Random Sampling when Optimum allocation is used is given as:

Variance of y bar st under Optimum allocation is equal to summation, h runs from 1 to k, Wh into Sh whole square by n minus summation, h runs from 1 to k, Wh into Sh square by N

Proof:

We know that:

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

Under Optimum Allocation we have,

nh is equal to n into Nh into Sh by summation, h runs from 1 to k, Nh into Sh and Wh is equal to Nh by N

Then, Variance of y bar st under Optimum allocation is equal to summation, h runs from 1 to k, (Nh by N) whole square into Sh square into summation , h runs from 1 to k, Nh into Sh by (n into Nh into Sh) minus (summation, h runs from 1 to k, (Nh by N) whole square into Sh square) by Nh

Which is equal to summation, h runs from 1 to k, (Nh into Sh) into (summation, h runs from 1 to k, Nh into Sh) by (N into N into n) minus (summation, h runs from 1 to k, Nh into Sh square by (N into N).

By substituting Nh by N as Wh we get

Variance of y bar st under Optimum allocation is equal to summation, h runs from 1 to k, (Wh into Sh) into (summation, h runs from 1 to k, Wh into Sh)by n minus, (summation, h runs from 1 to k, Wh into Sh square) by N

Variance of y bar st under Optimum allocation is equal to summation, h runs from 1 to k, (Wh into Sh) whole square by n minus (summation, h runs from 1 to k, Wh into Sh square) by N.

3. **Comparison of Stratified Sampling – Part 1**

Comparison of Stratified Sampling under Optimum and Proportional allocations and Simple Random Sampling Without Replacement

One can observe that:

Variance of y bar st under optimum allocation is less than or equal to

Variance of y bar st under proportional allocation is less than or equal to

Variance of y bar st under Simple Random Sampling Without Replacement

<u>Proof</u>:

We know that:

Sh square is equal to summation i runs from 1 to Nh (Yhi minus Y bar h) whole square divided by (Nh minus 1)

Which implies (Nh minus 1) into Sh square

is equal to summation i runs from 1 to Nh (Yhi minus Y bar h) whole square .

Call this as equation 1.

Consider S square is equal to summation h runs from 1 to k, summation i runs from 1 to Nh (Yhi minus Y bar) whole square divided by (N minus 1)

By subtracting and adding Y bar h within summation we get,

S square is equal to summation h runs from 1 to k, summation i runs from 1 to Nh (Yhi minus ‘Y bar h’ plus ‘Y bar h’ minus Y bar) whole square divided by (N minus 1).

Which is equal to summation h runs from 1 to k, summation i runs from 1 to Nh (Yhi minus Y bar h) whole square by (N minus 1)

plus summation h runs from 1 to k, summation i runs from 1 to Nh (Y bar h minus Y bar) whole square by (N minus 1) plus two into summation, h runs from 1 to k, summation, i runs from 1 to Nh (Yhi minus Y bar h) into (Y bar h minus Y bar) divided by (N minus 1).

S square is equal to summation h runs from 1 to k, summation i runs from 1 to Nh (Yhi minus

Y bar h) whole square by (N minus 1)
plus summation h runs from 1 to k, summation i runs from 1 to Nh (Y bar h minus Y bar) whole square by (N minus 1)
plus two into summation, h runs from 1 to k, (Y bar h minus Y bar) summation, i runs from 1 to Nh (Yhi minus Y bar h) divided by (N minus 1),

Which is equal to summation h runs from 1 to k, summation i runs from 1 to Nh (Yhi minus Y bar h) whole square by (N minus 1)
plus summation h runs from 1 to k, summation i runs from 1 to Nh (Y bar h minus Y bar) whole square by (N minus 1)
(Since, summation, i runs from 1 to Nh (Yhi minus Y bar h) is equal to zero)

From equation 1 we get,
S square is equal to summation h runs from 1 to k, (Nh minus one) into Sh square by (N minus 1)
plus summation h runs from 1 to k, Nh into (Y bar h minus Y bar) whole square by (N minus 1)

When (N minus 1) and (N_h minus 1) are sufficiently large we can take:
(N minus 1) approximately equal to N and (Nh minus 1) approximately equal to Nh
S square is equal to summation h runs from 1 to k, Nh into Sh square by N
plus summation h runs from 1 to k, Nh into (Y bar h minus Y bar) whole square by N.
S square is equal to summation h runs from 1 to k, Wh into Sh square
plus summation h runs from 1 to k, Wh into (Y bar h minus Y bar) whole square, since Wh is equal to Nh by N

Now by multiplying both LHS and RHS by N minus n by N into n
We get,
(N minus n) into S square by N into n is equal to
N minus n into summation, h runs from 1 to k Wh into Sh square by N into n plus N minus n into summation, h runs from 1 to k Wh into (Y bar h minus Y bar) whole square by N into n
Which is equal to summation, h runs from 1 to k Wh into Sh square by n minus summation, h runs from 1 to k Wh into Sh square by N
plus N minus n into summation, h runs from 1 to k Wh into (Y bar h minus Y bar) whole

square by N into n

But Variance of y bar under Simple Random Sampling without Replacement is equal to (N minus n) into S square by N into n.
Hence, variance of y bar under Simple Random Sampling without Replacement is equal to summation, h runs from 1 to k Wh into Sh square by n minus summation, h runs from 1 to k Wh into Sh square by N plus some positive quantity
Which implies variance of y bar under Simple Random Sampling without Replacement is equal to variance of y bar st under Proportional Allocation plus some positive quantity (from theorem 1)

Hence, Variance of y bar under Simple Random Sampling without Replacement is greater than or equal to variance of y bar st under Proportional Allocation
Or
Variance of y bar st under Proportional Allocation is less than or equal to variance of y bar under Simple Random Sampling without Replacement. Call this as equation two

4. **Comparison of Stratified Sampling – Part 2**

Now, let us prove that Variance of y bar st under Optimum Allocation is less than or equal to Variance of y bar st under Proportional Allocation
We can prove the above by proving:
Variance of y bar st under Proportional Allocation minus Variance of y bar st under Optimum Allocation is greater than or equal to zero.

Variance of y bar st under Proportional Allocation minus Variance of y bar st under Optimum Allocation is equal to summation h runs from 1 to k, Wh into Sh square by n minus summation h runs from 1 to k, Wh into Sh square by N minus (summation h runs from 1 to k, Wh into Sh whole square by n minus summation h runs from 1 to k, Wh into Sh square by N)
Which is equal to summation h runs from 1 to k, Wh into Sh square by n minus (summation h runs from 1 to k, Wh into Sh whole square by n).

So, Variance of y bar st under Proportional Allocation minus Variance of y bar st under Optimum Allocation is equal to
summation h runs from 1 to k, Nh into Sh square by (N into n) minus (summation h runs from 1 to k, Nh into Sh whole square) by N square into n
Which is equal to
(1 by N into n) whole multiplied by summation h runs from 1 to k, Nh into Sh square minus (summation h runs from 1 to k Nh into Sh whole square) by N
Which is equal to
(1 by N into n) whole multiplied by summation h runs from 1 to k, Nh into Sh square minus two into (summation h runs from 1 to k Nh into Sh) whole square by N plus (summation h runs from 1 to k Nh into Sh whole square) by N.

So, Variance of y bar st under Proportional Allocation minus Variance of y bar st under

Optimum Allocation is equal to summation, h runs from 1 to k Nh by N into n whole multiplied by (Sh square) minus (two into Sh) into summation, h runs from 1 to k, Nh into Sh by N plus (summation , h runs from 1 to k, Nh into Sh whole square) by N square.
Which is equal to summation, h runs from 1 to k Nh by N into n whole multiplied by (Sh minus summation , h runs from 1 to k, Nh into Sh by N) whole square
Which is equal to summation, h runs from 1 to k Nh by N into n whole multiplied by (A minus B) whole square
Where A is equal to Sh and B is equal to summation, h runs from 1 to k, Nh into Sh by N which is a positive quantity

Hence, Variance of y bar st under Proportional Allocation minus Variance of y bar st under Optimum Allocation is equal to
summation, h runs from 1 to k Nh by N into n whole multiplied by (A minus B) whole square
Where A is equal to Sh and B is equal to summation, h runs from 1 to k, Nh into Sh by N, which is a positive quantity

That is Variance of y bar st under Proportional Allocation minus Variance of y bar st under Optimum Allocation is greater than or equal to zero
Which implies Variance of y bar st under Proportional Allocation is greater than or equal to Variance of y bar st under Optimum Allocation
Which implies Variance of y bar st under Optimum Allocation is less than or equal to Variance of y bar st under Proportional Allocation .
Call this as equation three.
By comparing equations 2 and 3 we get
Variance of y bar st under optimum allocation is less than or equal to
Variance of y bar st under proportional allocation is less than or equal to Variance of y bar st under Simple Random Sampling Without Replacement

From the above we conclude that Neyman's Optimum Allocation gives better estimates than the Proportional allocation and,
the greater the difference between the stratum standard deviation, more the gain in precision of Neyman's allocation over Proportional allocation

5. **Advantages**

Advantage of Stratified Sampling Over Simple Random Sampling without Replacement:

- Gain in precision
- Flexible in the choice of the sample design for different strata
- Able to get estimates of each stratum in addition to the population estimates

A properly designed sample always leads to good survey results.

- When conducting the survey and the target population shows subpopulations, it is highly recommended to use Stratified Random Sampling to design sample compared to Simple Random Sampling.
- Stratification will give us better estimates for the population parameters compared to those we would get using Simple Random Sampling

Stratification will produce large gains in precision by satisfying the following three conditions
The population is composed of subgroups varying widely in size
The principal variables to be measured are closely related to the sizes of the subgroups
A good measure of size is available for setting up the strata

Compared to unstratified sampling, stratified sampling:

1. Permits the estimation of population parameters and within-strata inferences and comparisons across strata
2. Tends to be more representative of a population
3. Takes advantage of the researchers knowledge about the population
4. Possibly makes for lower data collection costs
5. Permits the researcher to use different sampling procedures within the strata

Unlike unstratified sampling, stratified sampling requires prior information on the stratification variables and more complex analysis procedures.

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

- The derivation of Variance of estimator of the population mean under stratified Random sampling with Proportional Allocation

- Variance of estimator of the population mean under stratified Random sampling with Optimum Allocation
- Addressed the comparison of Proportional, Optimum allocation under Stratified random Sampling with SRSWOR
- Discussed merits of Stratified Random Sampling as compared to Simple Random Sampling