1. Introduction

Welcome to the series of E-learning module on Estimation of Population Mean and Total. In this module, we are going to cover the basic procedure of Stratified Random Sampling, Distribution of observations under Stratified Random Sampling, Notations used for population parameters and estimators under stratification, Unbiased estimates of Population mean and total under Stratified Random Sampling with Simple Random Sampling with and Without Replacement.

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

- Explain the basic procedure of Stratified Random Sampling
- Explain the notations used for parameters and estimators under stratification
- Describe the unbiased estimates of population mean and total under Stratified Sampling with SRSWR
- Describe unbiased estimates of population mean and total under Stratified Sampling with SRSWOR

Stratified Sampling is a process of selecting a sample, which allows identified subgroups in the defined population to be represented in the same proportion that exists in the population.

Let us now discuss some of the principal steps in Stratified Sampling:

1. Identify and define the population.
2. Determine the desired sample size.
3. Identify the variable and subgroups (strata) for which you want to guarantee appropriate, equal representation.
4. Classify all members of the population as members of one identified subgroup.
5. Randomly select (using a table of random numbers) an "appropriate" number of individuals from each of the subgroups. Appropriate means an equal number of individuals.

Stratified random sampling refers to a sampling method that has the following properties:

- The population consists of N elements
- The population is divided into **h** groups, called strata
- Each element of the population can be assigned to one and only one stratum
- The number of observations within each stratum N_h is known, and
    - N is equal to N one **plus** N two **plus** N three **plus etc.** Plus N_h plus etc. plus N_k.
- The researcher obtains a **probability sample** from each stratum.

2. Population Parameters and Sample estimators for Strata

Population Parameters and Sample estimators for Strata
The basic notation for stratified sampling is as follows.
The population consists of N elementary sampling units that are grouped exclusively and exhaustively into k strata such that stratum 1 contains N one elementary sampling units, stratum two contains N two sampling units,..., and stratum h contains N_h elementary sampling units . Hence the population of size N is subdivided into k strata of sizes N one, N two,….,N_h, etc N_k.

Let Yhi_i represent the value of the variable of interest for the i^{th} elementary unit within stratum h and the elementary units in any particular stratum h are labelled from 1 to N.
Or
Y_{hi} is the i^{th} population unit of the population from the h^{th} stratum.
The population units be as shown in the table here:

Figure 1

Stratum No.	Observations							Stratum size	Stratum mean
1	Y_{11}	Y_{12}	...	....	Y_{1i}	...	Y_{1N1}	N_1	
2	Y_{21}	Y_{22}	...	....	Y_{2i}	...	Y_{2N2}	N_2	$\overline{Y_1}$
.	.	.	.	.	.	.	.	.	$\overline{Y_2}$
.	.	.	.	.	.	.	.	.	.
h	Y_{h1}	Y_{h2}	...	....	Y_{hi}	...	Y_{hNh}	N_h	$\overline{Y_h}$
.	.	.	.	.	.	.	.	.	.
k	Y_{k1}	Y_{k2}	...	....	Y_{ki}	...	Y_{kN1}	N_k	$\overline{Y_k}$
	$N = \sum_{h=1}^{k} N_h$								

Suppose we take a sample of size 'n', which is subdivided into various strata having sizes n one, n two,….n_h,…,n_k and Let y_{hi} represent the value of the variable of interest for the i^{th} elementary sample unit within stratum h and the elementary units in any particular stratum h are labelled from 1 to n.

i.e., y_{hi} is the i^{th} sample unit of the population from the h^{th} stratum.

Then, the sample units are as follows:

Figure 2

Stratum No.	Observations							Stratum sample size	Stratum sample mean
1	y_{11}	y_{12}	...	....	y_{1i}	...	y_{1N1}	n_1	
2	y_{21}	y_{22}	...	....	y_{2i}	...	y_{2N2}	n_2	$\overline{y}_1$
.	.	.	.	.	.	.	.	.	$\overline{y}_2$
.	.	.	.	.	.	.	.	.	.
h	y_{h1}	y_{h2}	...	....	y_{hi}	...	y_{hNh}	n_h	$\overline{y}_h$
.	.	.	.	.	.	.	.	.	.
k	y_{k1}	y_{k2}	...	....	y_{ki}	...	y_{kN1}	n_k	$\overline{y}_k$
	$n = \sum_{h=1}^{k} n_h$								

Population mean of the h^{th} stratum is:
Y bar h is equal to summation Y hi, i runs from one to Nh divided by Nh
Which implies, Nh into Y bar h is equal to summation Yhi, i runs from one to Nh. Call this equation as (one).
Population mean square in the h^{th} stratum is:
Sh square is equal to summation (Yhi minus Y bar h) whole square, i runs from one to Nh divided by (Nh minus one).

Similarly, we have population variance in the h^{th} stratum:
Sigma h square is equal to summation (Yhi minus Y bar h) whole square, i runs from one to Nh divided by Nh.
Therefore, we have (Nh minus one) into S h square is equal to Nh into sigma h square.

Similarly, a sample mean of the h^{th} stratum is:
Y bar h is equal to summation yhi , i runs from one to nh divided by nh
Which implies nh into y bar h is equal to summation yhi, i runs from one to nh.

Sample mean square in the h^{th} stratum is:
Sh square is equal to summation (yhi minus y bar h) whole square, i runs from one to nh divided by (nh minus one).

Under Stratified Random Sampling the population mean (Mean of the entire population) is given by, Y bar is equal to summation, h runs from 1 to k summation i runs from 1 to Nh, Yhi divided by N.
Which implies,

Y bar is equal to summation h runs from 1 to k , Nh into Y bar h divided by N is equal to

summation h runs from 1 to k, Wh into Y bar h. From (one)

Where, Wh is equal to Nh divided by N is the stratum weight.

Therefore, Y bar is equal to summation, h runs from 1 to k Wh into Y bar h. Similarly, sample mean y bar is equal to summation h runs from i to k, summation i runs from 1 to nh, yhi divided by n.

Which implies y bar is equal to summation, h runs from 1 to k nh into y bar h divided by n is equal to summation, h runs from 1 to k, wh into y bar h.

Where, wh is equal to nh by n and therefore ybar is equal to summation h runs from 1 to k wh into y bar h.

The total or aggregate amount of a variable within a stratum h is defined as:
Yh is equal to summation Yhi, i runs from 1 to Nh

Entire population total is Y is equal to summation, h runs from 1 to k, summation i runs from 1 to Nh , Yhi

Population variance for the entire population of size N is:

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

3. Properties of the Estimates

Let us now see the Properties of the Estimates
For the population mean per unit the estimate used in Stratified Sampling is:
y bar st is equal to summation, h runs from 1 to k Nh into y bar h divided by N which is equal to summation, h runs from 1 to k, Wh into y bar h
Where, N equals to N1 plus N two plus etc. plus Nk.

The estimate y bar st (st means stratified) is not in general the same as the sample mean. The sample mean is:
y bar is equal to summation, h runs from 1 to k nh into y bar h divided by n which is equal to summation, h runs from 1 to k ,wh into y bar h.

The difference is that in y bar st the estimates from the individual strata receive their correct weights Nh by N. It is evident that y bar coincides with y bar st provided that in every stratum
Nh by n is equal to Nh by N or nh by Nh is equal to n by N
or f h is equal to f
This means that the sampling fraction is the same in all strata. The Principal properties of the estimate y bar st are outlined in the following theorems.

Let us discuss the unbiased estimates of population mean and total under Stratified Sampling with Stratified Random Sampling with Replacement.
Theorem 1: Under Stratified random Sampling with replacement an unbiased estimator of population mean is given by y bar st is equal to summation h runs from 1 to k Wh into y bar h and a stratified sample mean is not unbiased for the population mean.

Proof:

Let us divide the entire population of size N into k strata such that each strata consists of N one, N two,…, Nk number of population observations. Let us draw a sample of size n from each and every stratum of the population using SRSWR.
The number of sample observations selected are n one, n two,…,nk from k strata.

Then we know that,
Y bar is equal to summation, h runs from 1 to k , Wh into Y bar h
y bar is equal to summation, h runs from 1 to k , wh into y bar h and
y bar st is equal to summation, h runs from 1 to k , Wh into y bar h.

We have to prove Expected value of y bar st is equal to Y bar and Expected value of y bar is not equal to Y bar
Consider Expected value of y bar st is equal to Expected value of summation, h runs from 1 to k, Wh into y bar h
Which is equal to summation, h runs from 1 to k, Wh into Expected value of y bar h. Call this as star.

Consider Expected value of y bar h is equal to Expected value of summation , i runs from 1 to nh yhi divided by nh
which is equal to summation , i runs from 1 to nh Expected value of yhi divided by nh. Call this as (2)
Expected value of yhi is equal to summation over i, yhi into pi.

Consider Expected value of y hi, since each y_{hi} takes values Y_{hi} with probability 1 by N_h we have,
Expected value of yhi is equal to summation, i runs from 1 to Nh , Yhi into (1 by Nh) which is equal to Y bar h

By substituting the above in equation (2), we get

Expected value of y bar h is equal to summation, i runs from 1 to nh Y bar h divided by nh which is equal to nh into Y bar h divided by nh which is equal to Y bar h.
Expected value of y bar h is equal to Y bar h Under SRSWR.

By substituting the above equation in (star) we get,
Expected value of y bar st is equal to summation, h runs from 1 to k, Wh into Y bar h, which is equal to Y bar.
Therefore, y bar st is equal to summation, h runs from 1 to k, Wh into y bar h is an unbiased estimator of population mean, Y bar
Therefore, Estimate of Y bar cap is equal to y bar st.

To show that Expected value of y bar is not equal to Y bar
Consider Expected value of y bar is equal to Expected value of summation, h runs from 1 to k , wh into y bar h
which is equal to summation, h runs from 1 to k , wh into Expected value of y bar h
which is equal to summation, h runs from 1 to k , wh into Y bar h under SRSWR which is not equal to Y bar.
Therefore, Expected value of y bar is not equal to Y bar.

That is Under Stratified Random Sampling; sample mean is biased for the population mean.
Hence an unbiased estimator of the population mean is:
y bar st is equal to summation, h runs from 1 to k, Wh into y bar h

and not the sample mean, y bar is equal to summation h runs from 1 to k wh into y bar h.

4. Unbiased Estimates of Population Mean and Total under Stratified Sampling with SRSWR – Theorem 2

Theorem 2:
An unbiased estimator of population total is given by N into y bar st

That is Y cap is equal to N into y bar st.

Proof: The population total under Stratified Random Sampling is given by,
Y is equal to summation, h runs from 1 to k, summation i runs from 1 to Nh Yhi which is equal to N into Y bar
Because Y bar is equal to summation h runs from 1 to k, summation i runs from 1 to Nh , Yhi by N
Unbiased estimator of population total
Y cap is equal to N into Y bar cap which is equal to N into y bar st
Because Expected value of y bar st is equal to Y bar, which implies Y bar cap is equal to y bar st.
Hence the result.

5. Unbiased Estimates of Population Mean and Total under Stratified Sampling with SRSWR – Theorem 3

Let us now discuss the unbiased estimates of population mean and total under Stratified Sampling with Stratified Random Sampling without Replacement.

Theorem 3:

Under Stratified Random Sampling Without Replacement an unbiased estimator of population mean is also given by y bar st is equal to summation h runs from 1 to k Wh into y bar h and a stratified sample mean is not unbiased for the population mean.

Proof:

Let us divide the entire population of size N into k strata such that each strata consists of N one, N two,…,N k number of population observations. Let us draw a sample of size n from each and every stratum of the population using SRSWOR.

The number of sample observations selected are n one, n two,…,nk from k strata.

We have to prove Expected value of y bar st is equal to Y bar and Expected value of y bar is not equal to Y bar

Consider Expected value of y bar st is equal to Expected value of summation , h runs from 1 to k , Wh into y bar h

which is equal to summation , h runs from 1 to k , Wh into Expected value of y bar h. Call this as star.

But under SRSWOR also

Expected value of y bar h is equal to Y bar h.

Already we proved under Simple Random Sampling Without Replacement that the sample mean is an unbiased estimator of the population mean. Since a random sample of size n_h is drawn from the h^{th} stratum using SRSWOR sample mean of the h^{th} stratum must be equal to the population mean of the h^{th} stratum.

By substituting the above equation in (*) we get

Expected value of y bar st is equal to summation, h runs from 1 to k, Wh into Y bar h, which is equal to Y bar.

Therefore, y bar st is equal to summation, h runs from 1 to k, Wh into y bar h is an unbiased estimator of population mean, Y bar

Therefore, Estimate of Y bar is equal to y bar st.

Theorem 4:
An unbiased estimator of population total under Stratified Random Sampling with SRSWOR is given by N into y bar st

That is, Y cap is equal to N into y bar st

Proof:
The population total under Stratified Random Sampling is given by,
Y is equal to N into Y bar
Because Y bar is equal to summation h runs from 1 to k, summation i runs from 1 to Nh, Yhi by N.

Unbiased estimator of population total
Y cap is equal to N into Y bar cap which is equal to N into y bar st
Because Expected value of y bar st is equal to Y bar which implies Y bar cap is equal to y bar st
Hence, an unbiased estimator of population total is given by, N into y bar st

That is Y cap is equal to N into y bar st.

Here's a summary of our learning in this session, where we have:

- Discussed Principal steps in Stratified sampling
- Illustrated distribution of population observations and sample observations in various strata
- Discussed basic notations for population parameters and estimators
- Derived Unbiased estimates of population mean and total under stratified Sampling with SRSWR
- Derived Unbiased estimates of population mean and total under stratified Sampling with SRSWOR