UNIT- III

STRATIFIED RANDOM SAMPLING

The theory of sampling is based on the assumption that population to be sampled is always homogeneous. But in practice, the population is not always homogeneous. In such cases to improve the precision of the estimates, we stratify or group the population into homogeneous sub-groups called as strata in such a way that the units within the group are alike.

The concept or factor which enables to classify the sampling units into various strata is known as stratifying factor.

Ex : Age, income, education, location etc are some of the stratifying factors.

The main objective of stratification is to give a better cross-section of the population so as to gain a higher degree of precision.

Principles of stratification

(i) The strata should be non-overlaping and should together constitute the whole population

(ii) The stratification should be made in such a way that the strata are homogeneous with respect to the characteristic under study

(iii) Administrative convenience may be consider as a basis of stratification.

Definition

The population consisting of 'N' sampling units is divided in to L relatively homogeneous mutually disjoint (non-overlapping) subgroups termed as strata of sizes $N_1, N_2 \ldots N_L$ such that $N = \sum_{h=1}^{L} N_h$. If a simple random sample (WOR) of size n_h, $h = 1, 2 \ldots L$ is drawn from each of the stratum respectively such that $n = \sum_{h=1}^{L} n_h$, the sample is termed as stratified r.s of size n and the technique of drawing such a sample is called Stratified Random Sampling.

Advantages

(i) Stratification may be desired for administrative convenience. It saves time and money because the units selected are geographically localised.

(ii) Stratification brings a gain in precision in the estimate of a characteristic under study.

(iii) Stratification makes it possible to use different sampling designs in different strata

(iv) When there are extreme values in the population they can be grouped in to a seperate stratum

Disadvantages

(i) It requires the knowledge of a proportion of population in each strata. If the population size of the strata is not available, the error increases.

(ii) The stratum should be well defined otherwise the error will be increases.

Notations

Y_{hi} = Value of the i^{th} unit in the h^{th} stratum
$(i = 1, 2 \ldots N_h)$ $(h = 1, 2 \ldots L)$

y_{hi} = Value of the i^{th} sampled unit from h^{th} stratum

N_h = population size of h^{th} stratum

n_h = sample size of h^{th} stratum

$$\sum_{h=1}^{L} n_h = n_1 + n_2 + \ldots + n_L = n$$

f_h = sampling fraction of h^{th} stratum

$W_h = \frac{N_h}{N}$ = stratum weight

$$\sum_{h=1}^{L} W_h = \frac{\sum_{h=1}^{L} N_h}{N} = \frac{N}{N} = 1 \qquad \because \sum_{h=1}^{L} N_h = N_1 + N_2 + \ldots + N_L = N$$

Sample mean of h^{th} stratum $= \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$

Sample mean square of h^{th} stratum $= s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$

Population mean of h^{th} stratum $\bar{Y}_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}$

$N_h \bar{Y}_h = \sum_{i=1}^{N_h} Y_{hi}$

$\frac{\sum_{h=1}^{L} N_h}{N} = \sum_{h=1}^{L} W_h$

population mean square of h^{th} stratum

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$$

Population mean $\bar{Y} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} Y_{hi}}{N}$

$$= \frac{\sum_{h=1}^{L} N_h \bar{Y}_h}{N} = \sum_{h=1}^{L} W_h \bar{Y}_h$$

Population Mean square

$$S^2 = \frac{1}{N-1}\left[\sum_{h=1}^{L}\sum_{i=1}^{N_h}(Y_{hi}-\bar{Y})^2\right]$$

Estimate of the Population Mean

In stratified random Sampling the estimate of the population mean

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$$

The sample mean $\bar{y} = \frac{\sum_{h=1}^{L} n_h \bar{y}_h}{n}$ which is not same as $\bar{y}_{st}$

The difference is that in $\bar{y}_{st}$ the estimates in the individual strata receive correct weights. $\bar{y}$ coincides with $\bar{y}_{st}$ when $f_h = f$ (ie) $\frac{N_h}{N} = \frac{n_h}{n}$.

Theorem : 1

If in every stratum, the sample mean is unbiased, then $\bar{y}_{st}$ is an unbiased estimate of the population mean

ie) $E(\bar{y}_{st}) = \bar{Y}$

Proof

Consider $\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$

Taking Expectations on both sides,

$$E(\bar{y}_{st}) = \frac{\sum_{h=1}^{L} N_h E(\bar{y}_h)}{N} \longrightarrow (1)$$

Since we use SRS for selecting sample for each stratum we have

$$E(\bar{y}_h) = \bar{Y}_h \longrightarrow (2)$$

using (2), (1) can be written as

$$E(\bar{y}_{st}) = \frac{\sum_{h=1}^{L} N_h \bar{Y}_h}{N} = \bar{Y} \quad (\text{By def of } \bar{Y})$$

Thus $\bar{y}_{st}$ is an unbiased estimate of the population mean

Note :

An unbiased estimate of the population total (Y) is given by

$$\hat{Y}_{st} = N \bar{y}_{st}$$

Theorem : 2

In stratified random sampling the variance of the estimate $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

$$= \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

$\frac{1}{N^2} \sum N_h^2 - \sum N_h n_h$

$\frac{1}{N^2} \sum N_h \left(1 - \frac{n_h}{N_h}\right)$

$\sum W_h^2 (1 - f_h)$

Proof

Consider $\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$

$$V(\bar{y}_{st}) = V\left(\frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} \right)$$

$V(\sum a_i x_i) = \sum a_i^2 V(x_i) + \sum \sum a_i a_j \operatorname{cov}(x_i, x_j)$

$$= \frac{1}{N^2} V\left(\sum_{h=1}^{L} N_h \bar{y}_h \right)$$

$$= \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 V(\bar{y}_h) + \sum_{h=1}^{L} \sum_{j \neq h=1}^{L} N_h N_j \operatorname{cov}(\bar{y}_h, \bar{y}_j)$$

$\because$ the sample are drawn independently in different strata, covariance terms vanish (becomes zero)

$$\therefore V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 V(\bar{y}_h) \longrightarrow (1)$$

Since we use SRS for selecting samples from each stratum and also

$$E(\bar{y}_h) = \bar{Y}_h$$

and $$V(\bar{y}_h) = \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} \rightarrow (2)$$

using (2), (1) can be written as

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^{\cancel{2}} \left(\frac{N_h - n_h}{\cancel{N_h}} \right) \frac{S_h^2}{n_h}$$

$$= \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

Note

1. consider $$V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

$$= \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$ $\because W_h = \frac{N_h}{N}$

$f_h = \frac{n_h}{N_h}$

2. $E(s_h^2) = S_h^2$ an estimate of $V(\bar{y}_{st})$ is given by

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{s_h^2}{n_h}$$

Theorem

Obtain the variance of the estimate of the population total.

Proof:

consider $$\hat{Y}_{st} = N \bar{y}_{st}$$

$$V(\hat{Y}_{st}) = N^2 V(\bar{y}_{st}) \rightarrow (1)$$

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h} \rightarrow (2)$$

using (2), (1) can be written as

$$V(\hat{Y}_{st}) = \cancel{N^2} \times \frac{1}{\cancel{N^2}} \sum_{h=1}^{L} N_h (N_h - n_h) S_h^2 / n_h$$

$$= \sum_{h=1}^{L} N_h (N_h - n_h) S_h^2 / n_h$$

Allocation of Sample in different Strata

While Allocating the sample to different strata, the following factors must be consider. The factors are

(i) The size of the Stratum (N_h)

(ii) Variability within the Stratum (S_h)

(iii) cost of inspecting a unit in the stratum

An allocation is said to be best one if and only if it gives maximum precision for the estimates with minimum resources.

There are four different methods of allocation and they are as follows

1. Equal allocation
2. Proportional allocation
3. Neyman allocation (optimum allocation for fixed sample size)
4. Optimal allocation

Equal Allocation

In equal allocation, the size of the Stratum, the Variability and the cost of inspecting a unit are ignored. In this method the sample size 'n' is divided equally among all the Strata.

For stratum h, $n_h = \frac{n}{L}$

$$= \sum \left(\frac{1}{n_h} - \frac{n_h}{N_h n_h} \right) W_h^2 S_h^2$$

$$= \sum \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2$$

We have $V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) W_h^2 S_h^2 \rightarrow (1)$

Put $n_h = \frac{n}{L}$ in (1)

$$V(\bar{y}_{st})_{equal} = \sum_{h=1}^{L} \left(\frac{L}{n} - \frac{1}{N_h}\right) W_h^2 S_h^2$$

Proportional allocation

Proportional allocation is commonly used technique in practise. This allocation assumes that the variability present in the strata as well as the cost of inspecting a unit are equal. Under this allocation, the sample size are ~~equal. Under this allocat~~ selected from each stratum is proportional to the size of the stratum (ie) allocation of n_h's to various strata are proportional if

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_h}{N_h} = \cdots = \frac{n_L}{N_L} = \frac{n}{N}$$

(ie) $\frac{n_h}{N_h} = \frac{n}{N}$

$\sum_{h=1}^{L} \frac{N_h}{N} = 1$

$\Rightarrow n_h = \frac{N_h}{N} \times \frac{n}{1}$

$\Sigma W_h^2 = \Sigma W_h \Sigma W_h$

$\Sigma W_h = \frac{\Sigma N_h}{N}$

Variance

we know $V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) W_h^2 S_h^2$

$$= \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \frac{\sum_{h=1}^{L} W_h \cdot \cancel{N_h} \cdot S_h^2}{\cancel{N_h}\, N}$$

$$= \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \frac{\sum_{h=1}^{L} W_h S_h^2}{N} \longrightarrow (1)$$

Under proportional allocation, $n_h = \frac{N_h}{N} \cdot n = W_h n \rightarrow (2)$

Using (2), (1) can be written as

$$V(\bar{y}_{st})_{Prop} = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{W_h \cdot n} - \sum_{h=1}^{L} \frac{W_h S_h^2}{N}$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} W_h^2 S_h^2$$

Neyman Allocation

In this allocation, the stratum size and variability are consider while allocating the sample to different strata. It is assumed that the sampling cost per unit is the same in all strata and the sample size is fixed. This allocation is also known as minimum variance allocation.

Under this allocation the sample size selected from h^{th} stratum is given by

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n \quad , \quad h = 1, 2 \ldots L$$

Variance : $\quad W_h = \frac{N_h}{N}$

we have $V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) W_h^2 S_h^2$

$$= \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h S_h^2}{N} \longrightarrow (1)$$

For Neyman allocation

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n$$

multiply and divide by 'N' we get

$$n_h = \frac{\left(\frac{N_h}{N}\right) S_h}{\sum_{h=1}^{L}\left(\frac{N_h}{N}\right) S_h} \times n = \frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h} \times n \rightarrow (2)$$

using (2), (1) can be written as

$$V(\bar{y}_{st})_{ney} = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{\frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h} \times n} - \sum_{h=1}^{L} \frac{W_h S_h^2}{N}$$

$$= \sum_{h=1}^{L} \frac{W_h S_h}{n} \cdot \sum_{h=1}^{L} W_h S_h - \sum_{h=1}^{L} \frac{W_h S_h^2}{N}$$

$$= \frac{\left(\sum_{h=1}^{L} W_h S_h\right)^2}{n} - \sum_{h=1}^{L} \frac{W_h S_h^2}{N}$$

Theorem :

In stratified random sampling for the fixed sample of size n, $V(\bar{y}_{st})$ is minimum only when n_h is proportional to $N_h S_h$ (ie) $n_h \propto N_h S_h$

Proof :

we have to minimize $V(\bar{y}_{st})$ subject to the Condition $\sum_{h=1}^{L} n_h = n$

We know that

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2$$

using lagrangian multiplier we can write a function

$$Z = V(\bar{y}_{st}) + \lambda\left(\sum_{h=1}^{L} n_h - n\right)$$

$$z = \frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2 + \lambda\left[\sum_{h=1}^{L} n_h - n\right] \rightarrow (1)$$

minimising $V(\bar{y}_{st})$ is equivalent to minimize z.

For z is minimum, we should have

$$\frac{\partial z}{\partial n_h} = 0 \quad \text{and} \quad \frac{\partial^2 z}{\partial n_h^2} > 0$$

Differentiating eqn (1) w. r. to n_h

$$\frac{\partial z}{\partial n_h} = -\frac{N_h^2 S_h^2}{N^2 n_h^2} + \lambda$$

$$z = \frac{1}{N^2}\frac{N_1^2 S_1^2}{n_1} + \frac{1}{N^2}\frac{N_2^2 S_2^2}{n_2} + \dots + \frac{1}{N^2}\frac{N_h^2 S_h^2}{n_h} + \dots + \frac{1}{N^2}\frac{N_L^2 S_h^2}{n_L} + \dots$$

$$+ \lambda\,(n_1 + n_2 + \dots + n_h + \dots n_L - n)$$

$$\frac{\partial z}{\partial n_h} = -\frac{N_h^2 S_h^2}{N^2 n_h^2} + \lambda$$

$$\frac{\partial z}{\partial n_h} = 0 \Rightarrow \lambda = \frac{N_h^2 S_h^2}{N^2 n_h^2}$$

$$n_h = \frac{N_h S_h}{N\sqrt{\lambda}} \longrightarrow (2)$$

Also
$$\frac{\partial^2 z}{\partial n_h^2} = \frac{2 N_h^2 S_h^2}{N^2 n_h^3} > 0$$

To determine λ, Summing eqn (2) over all the strata,

we have,

$$\sum_{h=1}^{L} n_h = \frac{\sum_{h=1}^{L} N_h S_h}{N\sqrt{\lambda}}$$

$$n = \frac{\sum_{h=1}^{L} N_h S_h}{N\sqrt{\lambda}} \quad \because \sum_{h=1}^{L} n_h = n$$

$$\sqrt{\lambda} = \frac{\sum_{h=1}^{L} N_h S_h}{Nn} \rightarrow (3)$$

using (3), (2) can be written as
$$n_h = \frac{N_h S_h}{N\left(\frac{\sum_{h=1}^{L} N_h S_h}{Nn}\right)}$$

$$= \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n$$

(ie) $n_h \propto N_h S_h$

Optimum Allocation

In this method of allocation, the sample size n_h in the respective strata are determine with a view to minimize the variance for a specified cost or to minimize the cost for a specified variance

A simplest cost function used in stratified random sampling is

$$C = c_0 + \sum_{h=1}^{L} c_h n_h$$

where c_0 = fixed cost

c_h = cost of inspecting a unit in h^{th} stratum

The size of the sample selected from h^{th} stratum is given by

$$n_h = \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$$

Variance:

we know that $$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \frac{1}{N} \sum_{h=1}^{L} W_h S_h^2 \rightarrow (1)$$

under optimum allocation,

$$n_h = \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$$

× & ÷ by N,

$$n_h = \frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h / \sqrt{c_h}} \times n \longrightarrow (2)$$

using (2), (1) can be written as

$$V(\bar{y}_{st}) = \frac{\sum_{h=1}^{L} W_h^2 S_h^2}{\dfrac{W_h S_h/\sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h/\sqrt{c_h}} \times n} - \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2$$

$$= \frac{1}{n}\left(\sum_{h=1}^{L} W_h S_h \sqrt{c_h}\right)\left(\sum_{h=1}^{L} W_h S_h/\sqrt{c_h}\right) - \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2$$

Theorem

In Stratified random sampling for a given cost function $C = c_0 + \sum_{h=1}^{L} c_h n_h$, $V(\bar{y}_{st})$ is minimum only when $n_h \propto W_h S_h/\sqrt{c_h}$

Proof:

Consider $V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2$

The given cost function is

$$C = c_0 + \sum_{h=1}^{L} c_h n_h$$

we have to minimize $V(\bar{y}_{st})$ subject to the condition

$$(C - c_0) = \sum_{h=1}^{L} c_h n_h$$

Using Lagrangian multiplier we can write

$$Z = V(\bar{y}_{st}) + \lambda\left(\sum_{h=1}^{L} c_h n_h - C + c_0\right)$$

$$= \frac{1}{N^2}\sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2 + \lambda\left(\sum_{h=1}^{L} c_h n_h - C + c_0\right) \rightarrow (1)$$

minimizing $V(\bar{y}_{st})$ = minimize Z

Z is minimum only when $\frac{\partial Z}{\partial n_h} = 0$ & $\frac{\partial^2 Z}{\partial n_h^2} > 0$

Differentiating eqn (1) w.r. to n_h

$$\frac{\partial z}{\partial n_h} = -\frac{N_h^2 S_h^2}{N^2 n_h^2} + \lambda C_h = 0$$

$$\lambda C_h = \frac{N_h^2 S_h^2}{N^2 n_h^2}$$

$$n_h = \frac{N_h S_h}{N\sqrt{\lambda}\sqrt{C_h}} \rightarrow (2)$$

also $$\frac{\partial^2 z}{\partial n_h^2} = \frac{2 N_h^2 S_h^2}{N^2 n_h^3} > 0$$

To determine λ, summing equ (2) over all the strata,

we have

$$\sum_{h=1}^{L} n_h = \frac{\sum_{h=1}^{L} N_h S_h}{N.\sqrt{\lambda}\sqrt{C_h}}$$

$$n = \frac{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}}{N\sqrt{\lambda}}$$

$$\sqrt{\lambda} = \frac{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}}{Nn} \rightarrow (3)$$

Using (3), (2) can be written as

$$n_h = \frac{N_h S_h / \sqrt{C_h}}{N \times \left(\frac{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}}{nN} \right)} = \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}} \times n$$

$$\therefore n_h \propto N_h S_h / \sqrt{C_h}.$$

Determination of Sample size

We know that $n_h = \dfrac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$

From the above formula, it is observed that n_h is in terms of n. The value of n_h depends upon whether the sample selected is so as to meet a specified total cost (C) or specified variance (V_0)

Case (i) : The cost is given

For the fixed total cost C, the optimum sample size can be determined as follows,

we have $C = C_0 + \sum_{h=1}^{L} c_h n_h$

$$C - C_0 = \sum_{h=1}^{L} c_h n_h \rightarrow (1)$$

and also $n_h = \dfrac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n \longrightarrow (2)$

Substitute (2) in eqn (1) we get

$$C - C_0 = \sum_{h=1}^{L} c_h \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$$

$$= \frac{\sum_{h=1}^{L} N_h S_h \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$$

$$\frac{1}{n} = \frac{\sum_{h=1}^{L} N_h S_h \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times \frac{1}{C - C_0}$$

$$\Rightarrow n = \frac{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h} \, (c - c_0)}{\sum_{h=1}^{L} N_h S_h \sqrt{c_h}}$$

Case (ii) : Variance is fixed

For the specified variance V_0, the optimum sample can be determined as follows.

Given that $V_0 = V(\bar{y}_{st})$

$$= \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

$$\frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} = V_0 + \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2 \longrightarrow (1)$$

we have $$n_h = \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n$$

Substituting the value of n_h in (1) we get

$$\frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{\left(\frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}} \times n \right)} = V_0 + \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

$$\frac{1}{nN^2} \left(\sum_{h=1}^{L} N_h S_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} N_h S_h / \sqrt{c_h} \right) = V_0 + \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

$$\frac{1}{N^2} \left(\sum_{h=1}^{L} N_h S_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} N_h S_h / \sqrt{c_h} \right) = n \left[V_0 + \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2 \right]$$

$$n = \frac{\frac{1}{N^2} \sum_{h=1}^{L} N_h S_h \sqrt{c_h} \cdot \sum_{h=1}^{L} N_h S_h / \sqrt{c_h}}{V_0 + \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2}$$

Comparison of Simple Random Sampling with Stratified random Sampling under Proportional and optimum allocation for fixed 'n'

Theorem :

If f.P.C is ignore, then S.T

$V_{opt} \leq V_{Pro} \leq V_{random}$ (or) $V_{random} \geq V_{Pro} \geq V_{opt}$

Proof :

The Variance of the estimate of the mean in SRSWOR, Stratified random Sampling under proportional and optimum allocation is given by

$$V_{random} = \left(\frac{1-f}{n}\right) S^2 \longrightarrow (1)$$

$$V_{Pro} = \left(\frac{1-f}{n}\right) \sum_{h=1}^{L} W_h S_h^2 \longrightarrow (2)$$

$$V_{opt} = \frac{\left(\sum_{h=1}^{L} W_h S_h\right)^2}{n} - \frac{\sum_{h=1}^{L} W_h S_h^2}{N} \longrightarrow (3)$$

Consider

$$S^2 = \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y})^2$$

Add and subtract $\bar{Y}_h$ we get

$$(N-1) S^2 = \sum_{h=1}^{L} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h + \bar{Y}_h - \bar{Y})^2$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left[(Y_{hi} - \bar{Y}_h)^2 + (\bar{Y}_h - \bar{Y})^2 + 2(\bar{Y}_{hi} - \bar{Y}_h)(\bar{Y}_h - \bar{Y}) \right]$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2 + \sum_{h=1}^{L} N_h (\bar{Y}_h - \bar{Y})^2 + 0$$

$$(N-1) S^2 = \sum_{h=1}^{L} (N_h - 1) S_h^2 + \sum_{h=1}^{L} N_h (\bar{Y}_h - \bar{Y})^2$$

$$\because S_h^2 = \frac{\sum_{i=1}^{N_h} (\bar{Y}_{hi} - \bar{Y})^2}{N_h - 1}$$

If we assume that N_h and consequently N are sufficiently large so that we can take $N_h - 1 \simeq N_h$ and $N-1 \simeq N$ then we get

$$N S^2 = \sum_{h=1}^{L} N_h S_h^2 + \sum_{h=1}^{L} N_h (\bar{Y}_h - \bar{Y})^2$$

Divide both sides by N, we get

$$S^2 = \sum_{h=1}^{L} W_h S_h^2 + \sum_{h=1}^{L} W_h (\bar{Y}_h - \bar{Y})^2$$

multiply both sides by $\left(\frac{1-f}{n}\right)$ we have

$$\left(\frac{1-f}{n}\right) S^2 = \left(\frac{1-f}{n}\right) \sum_{h=1}^{L} W_h S_h^2 + \left(\frac{1-f}{n}\right) \sum_{h=1}^{L} W_h (\bar{Y}_h - \bar{Y})^2$$

From (1) & (2)

$$V_{random} = V_{prop} + \left(\frac{1-f}{n}\right) \sum_{h=1}^{L} W_h (\bar{Y}_h - \bar{Y})^2$$

$$\therefore V_{random} \geq V_{prop} \longrightarrow (4)$$

When stratum means differs significantly considerable gain in precision is achieved through proportional allocation

Consider $V_{prop} - V_{opt} \geq 0$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} W_h S_h^2 - \left[\frac{1}{n}\left(\sum_{h=1}^{L} W_h S_h\right)^2 - \frac{\sum_{h=1}^{L} W_h S_h^{\cancel{2}}}{N}\right]$$

$$= \frac{1}{n}\left[\sum_{h=1}^{L} W_h S_h^2 - \left(\sum_{h=1}^{L} W_h S_h\right)^2\right]$$

$$= \frac{1}{n} \sum_{h=1}^{L} W_h (S_h - \bar{S})^2 \geq 0 \quad \text{where } \bar{S} = \sum_{h=1}^{L} W_h S_h$$

$$V_{prop} = V_{opt} + \frac{1}{n} \sum_{h=1}^{L} W_h (S_h - \bar{S})^2$$

$$\therefore V_{prop} \geq V_{opt} \longrightarrow (5)$$

$V_{prop} - V_{opt} = \frac{1}{n} \sum W_h (S_h - \bar{S})^2 \geq 0$

$\Rightarrow V_{prop} = V_{opt} + \frac{1}{n} \sum W_h (S_h - \bar{S})^2$

when Stratum standard deviation differs significantly optimum allocation is better than the proportional allocation

From (4) & (5) we have

$V_{rand} \geq V_{prop} \geq V_{opt}$

(or)

$V_{opt} \leq V_{prop} \leq V_{random}$

Hence the proof.

$\frac{1}{n} \sum_{h=1}^{L} W_h S_h^2 - \bar{S}^2$

$\frac{1}{n} \sum W_h S_h^2 + \bar{S}^2 - 2\bar{S}^2$

$\frac{1}{n} \sum W_h S_h^2 + \sum W_h \bar{S}^2 - 2\bar{S} \cdot \sum W_h S_h$

$\because \sum W_h = 1, \ \sum W_h S_h = \bar{S}$

$\frac{1}{n} \sum W_h \left[S_h^2 + \bar{S}^2 - 2\bar{S} S_h \right]$

$\frac{1}{n} \sum W_h (S_h - \bar{S})^2$