STRATIFIED RANDOM SAMPLING

BY DR RAJIV SAKSENA

DEPARTMENT OF STATISTICS

UNIVERSITY OF LUCKNOW

1. INTRODUCTION

The precision of a simple random sample estimate depends upon (i) the size of the sample and (ii) the variability (or heterogeneity) of the population. The size of the sample cannot be unduly increased; hence the only way to increase the precision of the estimate is to devise procedure which will effectively reduce the variability. One such procedure is known as the procedure of stratified random sampling. It consists in dividing the entire population into several non-overlapping classes or strata and drawing simple random samples from each of these strata and, then, combining all sample units together. It ensures any desired representation of units in all the strata and is intended to give a better cross-section of the population then that of unstratified random sampling. It follows that it will enhance the precision of the estimate since each strata within itself will be more homogeneous than the entire population.

2. REASONS (OR ADVANTAGES) OF STRATIFICATION)

(a) If data known precision are wanted for certain subdivisions of the population, it is advisable to treat each 'subdivision' as a 'population' in its own right.
(b) Administrative convenience may tackle the use of stratification: for illustration, the agency considering the survey may have field offices, each of which can supervise the survey for a part of the population.
(c) Sampling problems may differ markedly in different portions of the population: for example, these may be different types of sampling problems in plains, hilly areas and desserts which may need different approaches.
(d) Stratification is particularly more effective when there are extreme values in the population which can be segregate to from different strata: for example, the adult population may be divided into higher income, lower and unemployed sections.
(e) Stratification will lead to gain in precision since it may be possible to divide a heterogeneous population into sub-populations which are internally homogenous (as compared to whole population).

Let the population consisting of N units be divided into K strata. Let

N_h = number of units in the hth Stratum ($\sum_{1}^{k} N_h = N$)

y_{hi} = value of y in the ith units of the hth Stratum
(i=1,…, N_h ; h = 1, .., k)

$\bar{Y}_h = \sum_{i=1}^{Nh} y_{hi} / N_h$ is the hth stratum mean

$\bar{Y} = \sum_{h=1}^{k} \sum_{i=1}^{Nh} y_{hi} / N$ is the population mean

$S^2_h = \sum_{i=1}^{Nh} (y_{hi} - \bar{Y}_h)^2 / (N_h - 1)$

n_h = number of units in the sample from the hth Stratum ($\sum_{1}^{k} n_h = n$)

$\bar{y}_h = \sum_{i}^{nh} y_{hi} / n_h$ is the sample mean from the hth stratum

$\bar{y} = \sum_{1}^{k} n_h \bar{y}_h / n$ is the overall sample mean

$s_h^2 = \sum_{1}^{nh} (y_{hi} - y_n)^2 / (n_h - 1)$

$W_n = Nn/N$, $f_h = n_h/N_h$, sampling fraction

Definition: The stratified random sampling estimate $\overline{y_{st}}$ of population mean $\bar{Y}$ is defined by

$$\overline{y_{st}} = \frac{\sum_{h=1}^{k} N_h \overline{y_h}}{N}$$

which is, generally, different for $\bar{y}$ unless $\frac{N_h}{N} = \frac{n_h}{n}$ in which case we have "proportional allocation".

Theorem 1: For sampling without replacement (or with replacement) $\overline{y_{st}}$ is an unbiased estimator of $\bar{Y}$

Proof: Since we take simple random samples from each stratum, we have (in both cases)

$$E(\overline{y}_{st}) = E\left(\frac{1}{N}\sum_{h=1}^{k} N_h \overline{y}_h\right)$$

$$= \frac{1}{N}\sum_{h=1}^{k} N_h E(\overline{y}_h)$$

$$= \frac{1}{N}\sum_{h=1}^{k} N_h \overline{Y}_h$$

$$= \overline{Y}$$

Theorem 2: For sampling WOR,

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

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

When finite population correction (f pc) is ignored (n_n / N_h are negligible)

$$V(\overline{y}_{st}) = \sum_{h=1}^{k} W_h^2 \frac{S_h^2}{n_h}$$

Proof : We have, since a s.r.s. is drawn from each stratum,

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

$$= \frac{1}{N^2}\sum_{h=1}^{k} N_h^2 V(\overline{y}_h)$$

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

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

Corollary 1 : $N\overline{y}_{st}$ is the unbiased estimator of population total $N\overline{Y}$ and its variance is

$$V(\overline{y_{st}}) = \sum_{1}^{k} N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

<u>Corollary</u> 2 : An unbiased estimator of $V(\overline{y_{st}})$ is

$$v(\overline{y_{st}}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h (N_h - n_h) \frac{s_h^2}{n_h}$$

<u>Corollary</u> 3 : If sampling is "proportional" we substitute $n_h = \frac{n}{N} Nn$ and then

$$V(\overline{y_{st}}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{N_h} \frac{S_h^2}{n_h}$$

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

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

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

where f = n/N.

In above, if $S^2{}_h = S^2{}_0$ (same for all strata), then

$$V(\overline{Y_{st}}) = (1\text{-}f) \frac{S_o^2}{n}$$

<u>Theorem</u> 3: For sampling WR,

$$V(\overline{y_{st}}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h^2 \frac{\sigma_h^2}{n_h}$$

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

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

<u>Corollary</u> 4: An unbiased estimator of V ($\overline{y_{st}}$) is

$$v(\overline{y_{st}}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h (N_h - 1) \frac{S_h^2}{n_h}$$

<u>Theorem</u> 4: When the characteristic under study is an attribute, the unbiased estimator of population proportion, P, and the variance of estimator are given by the following:

(i) For sampling WOR, sampling proportion p_{st} is unbiased estimator of P with sampling variance given by

$$v(\overline{p_{st}}) = \sum_{h=1}^{k} W_h^2 \left(\frac{N_h - n_h}{N_{h-1}} \right) \frac{P_n Q_n}{n_h}$$

(ii) For sampling WR, sampling proportion p_{st} is an unbiased estimator of P with sampling variance given by

$$v(\overline{p_{st}}) = \sum_{h=1}^{k} W_h^2 \frac{P_n Q_n}{n_h}$$

<u>Proof:</u> Define p_{st} by

$$p_{st} = \sum_{h=1}^{k} W_h^2 \frac{P_n Q_n}{n_h}$$

Where $p_h = \frac{m_h}{n_h}$ = proportion of sample units drawn from the hth Stratum, possessing the

Attribute

$$E = (p_{st}) = P$$

$$V = (p_{st}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h^2 V(p_h)$$

Substituting the value of V (p_h) in the two cases, we get the desired result.

3. SAMPLE ALLOCATION

In stratified sampling, the allocation of the sample to different. Strata, which means the determination of the size of sample to be selected from each stratum, is done by the consideration of there factors viz, (i) Stratum size or the total number of units in the Stratum, (ii) the variability within the Stratum, and (iii) the cost of taking observation per sampling unit in the Stratum.

A good allocation is the where maximum precision is obtained with maximum resources, that is the criterion of allocation is to minimize the total cost for a given variance or minimize the variance for a fixed cost, thus making the effective use of available resources.

Three methods of allocation of sample sizes to different Strata are (a) equal allocation, (b) proportional allocation, and (c) optimum allocation.

(a) <u>Equal allocation</u>: Sometimes, for reasons of administrative or field come convenience, the total sample size n is divided equally among all Strata, i.e.

$$n_h = n/k$$

(b) <u>Proportional allocation</u>: There, the allocation of sample size to different strata is done in proportion to their sizes, i.e.

$$n_h \;\alpha\; N_h$$

$$\text{or} \quad n_h = \frac{nNh}{N}$$

When proportion allocation is used, the variance of the satisfied random sample estimator is

$$V(\bar{y}_{prop}) = \left(1 - \frac{n}{N}\right) \sum_{h=1}^{k} \frac{N_h S_h^2}{N_h}$$

(c) <u>Optimum allocation</u> :

<u>Theorem 5</u> : In stratified random sampling, the variance of $\overline{Yst}$ is minimum, for a fixed total size of the same, if the sample is allocation with

$$n_h \;\alpha\; N_h S_h$$

$$\text{or } n_n = \frac{N_h S_h}{\Sigma N_h S_h} .n$$

when this allocation, called Neyman's allocation, is used the variance of the estimator is

$$V(\bar{y}_{opt}) = \frac{1}{N^2}\left(\sum_{1}^{k} N_h S_h\right)^2 / n - \frac{1}{N^2} \sum_{1}^{k} N_h S_h^2$$

<u>Proof</u>: The problem is to minimize

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

subject to the restriction $\sum^{k} = n$ (fixed) Using the method of Lagrange multiples

1

we have to choose n_h and X so as to minimize

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

$$= \frac{1}{N^2} \sum_{1}^{k} N_h (N_h - n_h) \frac{S_h^2}{n_h} + \lambda \left(\sum_{1}^{k} n_h - n \right)$$

$$\frac{\partial \phi}{\partial n_h} = 0 \text{ gives}$$

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

$$\text{or, } n_h = \frac{N_h S_h}{N\sqrt{\lambda}} \qquad \text{-------------------------- (1)}$$

$$\text{But } \sum_{1}^{k} n_h = n = \sum_{1}^{k} N_h S_h / N\sqrt{\lambda}$$

$$\text{or, } \sqrt{\lambda} = \sum_{1}^{k} N_h S_h / nN$$

Substituting the value of $\sqrt{\lambda}$ in (i) we get

$$n_h = \frac{N_h S_h}{\sum_{1}^{k} N_h S_h} \, . \, n$$

$$\text{or, } n_n \; \alpha \; N_n S_n$$

Substituting the value of n_h from (2) we get

$$V(\bar{y}_{opt}) = V_{\min}(\bar{y}_{st})$$

$$= \frac{1}{N^2} \sum_{1}^{k} \left(\frac{N_h^2 S_h^2}{(N_h S_h) n} \left(\sum_{1}^{k} N_h S_h \right) - N_h S_h^2 \right)$$

$$= \frac{1}{N^2} \left(\sum_{1}^{k} \frac{N_h^2 S_h^2}{n} \right) - \frac{1}{N^2} \sum_{1}^{k} N_h S_h^2$$

Cost function: Suppose the cost per unit in the hth stratum is c_h and we take the cost function as

k

$$C = Co + \sum_{1} c_h n_h$$

Where Co is the overall cost.

Theorem 6: With cost function of the form given in (3), the variance of $\bar{y}_{st}$ is minimize when

$$n_h \; \alpha \; N_h S_h / \sqrt{c_h}$$

$$\text{or,} \qquad n_h = \frac{N_h S_h / \sqrt{c_h}}{\sum_{1}^{k} N_h S_h / \sqrt{c_h}}$$

Proof: Using the method of Lagrange multipliers, we choose n_n and λ so as to minimize

$$\phi = V(\bar{y}_{st}) + \lambda \left(C_o + \sum_{1}^{k} n_h c_h \right)$$

$$= \frac{1}{N^2} \sum_{1}^{k} N_h (N_h - n_h) \frac{S_h^2}{n_h} + \lambda \left(C_o + \sum_{1}^{k} n_h c_h \right)$$

$$\frac{\partial \phi'}{\partial n_h} = 0 \text{ gives}$$

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

$$\text{or,} \quad n_h = \frac{N_h S_h}{N \sqrt{\lambda} \sqrt{C_h}}$$

$$\text{But} \quad \sum_{1}^{k} n_h = n = \sum_{1}^{k} N_h S_h / N \sqrt{\lambda} \sqrt{C_h}$$

$$\text{or,} \quad \sqrt{\lambda} = \frac{1}{nN} \sum_{1}^{k} \frac{N_h S_h}{\sqrt{C_h}}$$

Substituting the value of $\sqrt{\lambda}$ in (4) we get

$$n_h = n \left(\frac{\dfrac{N_h S_h}{\sqrt{C_h}}}{\sum_{1}^{k} \dfrac{N_h S_h}{\sqrt{C_h}}} \right)$$

The theorem states that in a given stratum, take a lager sample if the stratum is larger or the stratum is more variance one if the sampling is cheaper.

Corollary : If Cost per unit remains the same in all strata, $C_h = C$ and we get the Neyman's allocation, $n_h \;\alpha\; N_h S_h$.

Corollary: In Theorem 6 the total sample size n is not foxed. It can be found in two ways:

(a) When total cost C* is fixed

$$C^* = C_o + \sum_1^k C_h n_h$$

$$\text{or, } (C^* - C_o) = \frac{n \sum_1^k N_h S_h \sqrt{C_h}}{\sum_1^k N_h S_h \sqrt{C_h}}$$

$$\text{or, } n = (C^* - C_o) = \frac{n \sum_1^k N_h S_h \sqrt{C_h}}{\sum_1^k N_h S_h \sqrt{C_h}}$$

(b) When Variance is foxed = V*, say

$$V^* = \frac{1}{N^2} \sum_1^k \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2} \sum_1^k N_h S_h^2$$

$$\text{or, } V^* + \frac{1}{N} \sum_1^k W_h S_h^2 = \frac{1}{N^2} \sum_1^k \frac{N_h^2 S_h^2}{\dfrac{n\, N_h S_h \sqrt{C_h}}{\sum_1^k N_h S_h \sqrt{C_h}}}$$

$$= \frac{\left(\sum_1^k W_h S_h \sqrt{C_h}\right)\left(\sum_1^k W_h S_h \sqrt{C_h}\right)}{n}$$

$$\text{or, } n = \frac{\left(\sum_1^k W_h S_h \sqrt{C_h}\right)\left(\sum_1^k W_h S_h \sqrt{C_h}\right)}{V^* + \frac{1}{N} \sum_1^k W_h S_h^2}$$

4. RELATIVE PRECISION OF STRATIFIED AND SIMPLE RANDOM SAMPLING

If intelligently used, stratification will nearly always result in a smaller variance of the estimator than is given by a comparable simple random sample.

Theorem : If the population is large compared to the sample, n/N, n_h/N_h, $1/n_h$ an $1/N$ are negligible and, then

$$V(\bar{y}_{opt}) \leq V(\bar{y}_{prop}) \leq V(\bar{y}_{ran})$$

Where $\bar{y}_{opt}$ is the estimator using Neyman's allocation $n_n \; \alpha \; N_n S_n$ and $\bar{y}_{ran}$ is the SRSWOR estimator of size n .

Proof: We have

$$V(\bar{y}_{ran}) = \frac{N-n}{N} S^2 \cong \frac{S^2}{n}$$

$$V(\bar{y}_{prop}) = \frac{N-n}{N} \sum_{1}^{k} \frac{N_h S_h^2}{nN} \cong \sum_{1}^{k} \frac{N_h S_h^2}{nN}$$

$$V(\bar{y}_{opt}) = \frac{1}{N^2} \left(\sum_{1}^{k} \frac{N_h S_h}{n} \right)^2 - \frac{1}{N^2} \sum_{1}^{k} N_h S_h^2$$

$$\cong \frac{1}{nN^2} \left(\sum_{1}^{k} N_h S_h \right)^2$$

When approximations are used.
We have

$$(N-1) S^2 = \sum_{h} \sum_{i} (y_{ni} - \bar{Y})^2$$

$$= \sum_{h} \sum_{i} (y_{ni} - \bar{Y}_h)^2 + \sum_{h} N_h (\bar{Y}_n - \bar{Y})^2$$

$$= \sum_{h} (N_h - 1) S_h^2 + \sum_{h} N_h (\bar{Y}_n - \bar{Y})^2$$

Since terms in $1/N_k$ $1/N_h$ are neglible, we have

$$NS^2 = \sum_{h=1}^{k} N_h S_h^2 + \sum_{h} N_h (\bar{Y}_n - \bar{Y})^2$$

Therefore

$$V(\bar{y}_{ran}) = \frac{S^2}{n} = \frac{\sum_{1}^{k} N_h S_h^2}{nN} + \frac{\sum_{h=1}^{k} N_h (\bar{Y}_n - \bar{Y})^2}{nN}$$

$$= V(\bar{y}_{prop}) + \sum_{1}^{k} \frac{N_h(\bar{Y}_n - \bar{Y})^2}{nN}$$

which shows that

$$V(\bar{y}_{prop}) \leq V(\bar{y}_{ran})$$

Unless $\bar{Y}_h = \bar{Y}$ for every h.

By definition of optimum allocation, we must have $V(\bar{y}_{opt}) \leq V(\bar{y}_{prop})$.

The difference is

$$V(\bar{y}_{prop}) - V(\bar{y}_{opt}) = \frac{1}{nN}\left(\sum_{1}^{k} N_h S_h^2 - \frac{(\sum N_h S_h)^2}{N} \right)$$

$$= \frac{1}{nN} \sum_{1}^{k} N_h (S_h - \bar{S})^2$$

where, $\bar{S} = \sum_{1}^{k} N_h S_h$

This shows that

$$V(\bar{y}_{opt}) \leq V(\bar{y}_{prop})$$

Unless $S_h = \bar{S}$ for every h, i.e., the strata have equal variability,

Therefore, we get

$$V(\bar{y}_{opt}) \leq V(\bar{y}_{prop}) \leq V(\bar{y}_{ran})$$

Remark : It can be shown that when fpc is not neglected then also above inequality holds.

---

5. RELATIVE PRECISION OF STRATIFIED AND SIMPLE RANDOM SAMPLING

In comparing the precision of stratified and unstratified (simple random) sampling, it was assumed that the population values of stratum means ($\bar{Y}_h$) and variance ($S^2{}_h$) and overall variance S^2 were known. However, in practice, those values are not known and have to be estimate from the sample itself. It is of interest to examine, from a survey, how useful the stratification has been. What is available only a stratified sample and the problem is to estimate the gain in precision due to stratification. An estimate of the variance of the estimate in case of simple random sampling is obtained from the stratified sample and comparison can be made with a situation in which no stratification is done.

We know that an unbiased estimator of V ($\bar{y}_{st}$) is given by

$$v(\bar{y}_{st}) = \sum_{h=1}^{k} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h s_h^2$$

We have to obtain an unbiased estimator of $V(\bar{y}_{ran})$ on the basis of the stratified random sample, when

$$V(\bar{y}_{ran}) = \left(\frac{N-1}{Nn} \right) S^2 \quad \text{------------------------ (ii)}$$

We note that $s^2 = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (y_{ni} - y_h)^2 / (n-1)$

is not an unbiased estimator of S^2.

However,

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

$$= \frac{1}{N-1} \sum_{h=1}^{k} (N_h - 1) S_h^2 + \frac{N}{N-1} \left(\sum_{h=1} W_h (\bar{Y}_n)^2 - (\bar{Y})^2 \right) \quad \text{----------- (iii)}$$

We will first obtain unbiased estimators of $(\bar{Y}_h)^2$ and $(\bar{Y})^2$ occurring in the second term of (iii)

Since $V(\bar{Y}_h) = E(\bar{Y}_h)^2 - (\bar{Y}_h)^2$, an unbiased estimator of $(\bar{Y}_h)^2$ is given by

$$(\bar{Y}_h)^2 = (\bar{y}_h)^2 - V(\bar{y}_h)$$

$$= (\bar{y}_h)^2 - \left(\frac{1}{n_h} - \frac{1}{N_h} \right) s_h^2 \quad \text{-------------------------- (iv)}$$

Similarly, an unbiased estimator of $(\bar{Y})^2$ given by

$$(\bar{Y})^2 = (\bar{y}_{st})^2 - V(\bar{y}_{st})$$

$$= (\bar{y}_{st})^2 - \sum_{h=1}^{k} W_h^2 s_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \quad \text{------------------ (v)}$$

Therefore, using (iv) and (v) we get an unbiased estimator of S^2 as given in (iii) to be

$$S^2 = \frac{1}{N-1}\sum_{h=1}^{k}(N_h-1)s_h^2 + \frac{N}{N-1}\left(\sum_{h=1}^{k} W_h\left\{(\bar{y}_n)^2 - \left(\frac{1}{n_h}-\frac{1}{N_h}\right)s_h^2\right\} - \left\{(\bar{y}_{st})^2 - \sum_{1}^{k} W_h s_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)\right\}\right)$$

$$= \frac{1}{N-1}\sum_{h=1}^{k}(N_h-1)s_h^2 + \frac{N}{N-1}\left(\sum_{h=1}^{k} W_h(\bar{y}_n-\bar{y}_{st})^2 - \sum_{h=1}^{k} W_h(1-W_n)\left(\frac{1}{n_h}-\frac{1}{N_h}\right)s_h^2\right)$$

$$= \frac{1}{N-1}\sum_{h=1}^{k} s_h^2\left\{(N_h-1)+(1-W_n)\right\} + \frac{N}{N-1}\left(\sum_{h=1}^{k} W_h(\bar{y}_n-\bar{y}_{st})^2 - \sum_{h=1}^{k} W_h(1-W_n)\frac{s_h^2}{n_h}\right)$$

$$= \sum_{h=1}^{k} W_h s_h^2 + \frac{N}{N-1}\left(\sum_{h=1}^{k} W_h(\bar{y}_n)^2 - \left(\sum_{1}^{k} W_h\bar{y}_h\right)^2 - \sum_{1}^{k} W_h(1-W_n)\frac{s_h^2}{n_h}\right)$$

Using this unbiased estimator of S^2, an unbiased estimator of $V(\bar{y}_{ran})$ in (ii) is

$$v(\bar{y}_{ran}) = \frac{N-n}{Nn}\left(\sum_{h=1}^{k} W_h s_h^2 + \frac{N}{N-1}\left\{\sum_{h=1}^{k} W_h(\bar{y}_n)^2 - \left(\sum_{1}^{k} W_h\bar{y}_h\right)^2 - \sum_{h=1}^{k} W_h(1-W_n)\frac{s_h^2}{n_h}\right\}\right)$$

$$= \frac{N-n}{n(N-1)}\left(\frac{N-1}{N}\sum_{h=1}^{k} W_h s_h^2 + \left\{\sum_{h=1}^{k} W_h(\bar{y}_n)^2 - \left(\sum_{1}^{k} W_h\bar{y}_h\right)^2 - \sum_{1}^{k} W_h(1-W_n)\frac{s_h^2}{n_h}\right\}\right)$$

If $N>50$, $\frac{N-1}{N} \cong 1$, therefore

$$v(\bar{y}_{ran}) = \frac{N-n}{n(N-1)}\left(\sum_{h=1}^{k} W_h s_h^2 + \sum_{1}^{k} W_h(\bar{y}_n)^2 - \left(\sum_{1}^{k} W_h\bar{y}_h\right)^2 - \sum_{1}^{k} W_h(1-W_n)\frac{s_h^2}{n_h}\right) \quad \text{---(vi)}$$

If $n_h > 50$ for every h=1, ---, k, the last term in above can be dropped and

$$v(\bar{y}_{ran}) = \frac{N-n}{n(N-1)}\left(\sum_{h=1}^{k} W_h s_h^2 + \sum_{1}^{k} W_h(\bar{y}_n)^2 - \left(\sum_{1}^{k} W_h\bar{y}_h\right)^2\right)$$

The estimate of the relative gain in precision due to stratification is thus obtained by

$$\frac{v(\bar{y}_{ran}) - v(\bar{y}_{st})}{v(\bar{y}_{st})}$$

where $v(\bar{y}_{st})$ and $v(\bar{y}_{ran})$ are given in (i) and (vi) respectively.