SIAP

Statistical Institute for Asia and the Pacific

GLOBAL STRATEGY FOR AGRICULTURE AND RURAL STATISTICS

Regional Training Course on Sampling Methods for Producing Core Data Items for Agricultural and Rural Statistics

Jakarta, Indonesia ,29Sep-10 October 2014.

STRATIFIED SAMPLING

Bal Goel

INTRODUCTION

In case of simple random sampling without replacement, the sampling variance of the sample mean is

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right)S^2 \qquad (1.1)$$

Clearly, the variance decreases as the sample size (n) increases while the population variability S^2 decreases. Now one of the objectives of a good sampling technique is to reduce the sampling variance. So we have to either increase 'n' or decrease S^2. Apart from the sample size, therefore, the only way of increasing the precision of an estimate is to devise sampling procedure which will effectively reduce S^2 i.e. the heterogeneity in the population. In fact, S^2 is a population parameter and is inherent with the population, therefore, it cannot be decreased. Instead, the population may be divided into number of groups (called strata), thereby, controlling variability within each group. This procedure is known as stratification.

Contd.

Stratified sampling is a common procedure in sample surveys. The procedure enables one to draw a sample with any desired degree of representation of the different parts of the population by taking them as strata. In stratified sampling, the population consisting of N units is first divided into K sub-populations of $N_1, N_2, \ldots, N_K$ units respectively. These sub-populations are non-overlapping and together they comprise the whole of the population i.e. $\sum_{i=1}^{K} N_i = N$. These sub-populations are called strata. To obtain full benefit from stratification, the values of N_i's must be known. When the strata have been determined, a sample is drawn from each stratum, the drawings being made independently in different strata. If a simple random sample is taken in each stratum then the procedure is termed as stratified random sampling. As the sampling variance of the estimate of mean or total depends on within strata variation, the stratification of population is done in such a way that strata are homogeneous within themselves with respect to the variable under study. However, in many practical situations it is usually difficult to stratify with respect to the variable under consideration especially because of physical and cost consideration. Generally, the stratification is done according to administrative groupings, geographical regions and on the basis of auxiliary characters correlated with the character under study.

- 1. Since the population is first divided into various strata, and then samples are from each stratum there is little possibility of any essential group of population being completely excluded. Hence, stratif ication ensures that a better cross section of the population is represented in the sample as compared to that under unstratified sampling
- 2. The stratification makes it possible to use different sampling designs in different strata thereby enabling effective utilization of the available auxiliary information. It is particularly true in cases where the extent and nature of the available information vary from stratum to stratum. Separate estimates obtained for different strata can be combined into a precise estimate for the whole population.
- 3. When a survey organization has field offices in several zones it might be desirable to treat the zones as strata from the point of view of administrative convenience, as it will facilitate the supervision and organization of the field work

- 4. When there are extreme values in population, these can be grouped into a separate stratum thereby reducing the variability within other strata.
- 5. The geographical and topographical considerations may also be the reason for resorting to stratification. There may be different types of sampling problems in plains, deserts and hilly areas. These may need different approaches for their resolution. Hence, it would be advantageous to form separate stratum for each of such areas.
- 6. Since the variability within strata is considerably reduced the stratification normally provides more efficient estimates than the usual unstratified sampling
- 7. The cost of conducting the survey is expected to be less for stratified sampling when strata are formed keeping administrative convenience in mind.

Let $N_1, N_2, \ldots, N_K$ denote the size of K strata, such that $\sum_{i=1}^{K} N_i = N$ (total number of units in the population). Let y_{ij} denotes the j-th observation in the i-th stratum. Denote $\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{N_i} y_{ij}$ as population mean.

Again

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{K}\sum_{j=1}^{N_i}\left(y_{ij} - \bar{Y}_N\right)^2$$ = population mean square

$$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}\left(y_{ij} - \bar{Y}_{N_i}\right)^2$$ = population mean square for the i-th stratum

where $\bar{Y}_{N_i} = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$ = population mean for the i-th stratum.

Select a simple random sample of size n_1 from N_1 units in the first stratum, of size n_2 from N_2 units in the second stratum,…, of size n_i from N_i units in the i-th stratum and so on such that

$n_1+n_2 + ...+n_i+…+n_K = n$ (sample size).

The population mean $\bar{Y}_N$ can be written as

$$\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{K} N_i \bar{Y}_{N_i} = \sum_{i=1}^{K} P_i \bar{Y}_{N_i}$$

(1.2)

where $P_i = N_i / N$. Since in each stratum, the samples have been drawn by simple random sampling, $\bar{y}_{n_i} \left(= \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}\right)$, sample mean for i-th stratum, is an unbiased estimator of $\bar{Y}_{N_i}$ and obviously the stratified sample mean

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{K} N_i \bar{y}_{n_i} = \sum_{i=1}^{K} P_i \bar{y}_{n_i}$$

(1.3)

which is the weighted mean of the strata sample means with strata size as the weights, will be an appropriate estimator of the population mean.

The $\bar{y}_{st}$ is an unbiased estimator of $\bar{Y}_N$.

Since the sample in the i-th stratum has been drawn by simple random sampling without replacement, so

$$V\left(\bar{y}_{n_i}\right) = \left(\frac{1}{n_i} - \frac{1}{\mathrm{N}_\mathrm{i}}\right) \mathrm{S}_\mathrm{i}^2$$

(1.4)

The sampling variance of $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st}) = \sum_{i=1}^{K} P_i^2 V(\bar{y}_{n_i}) = \sum_{i=1}^{K} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2$$

(1.5)

Since sample mean square for the i-th stratum, $s_i^2 = \frac{1}{n_i - 1} \sum_{j}^{n_i} (y_{ij} - \bar{y}_{n_i})^2$ unbiasedly estimates S_i^2, it follows that an unbiased estimator of $V(\bar{y}_{st})$ is given by

$$\hat{V}(\bar{y}_{st}) = \sum_{i=1}^{K} P_i^2 \hat{V}(\bar{y}_{n_i}) = \sum_{i=1}^{K} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) s_i^2 .$$

(1.6)

From the above, it is clear that the sampling variance of stratified sample mean depends on S_i's, variabilities within the strata which suggests that the smaller the S_i's, i.e. the more homogeneous the strata, greater will be the precision of the stratified sample mean.

2. EXAMPLE 1

The data given below pertains to the total geographical area in 20 villages of a block. Treating this as population of 20 units, stratify this population in three strata taking stratum sizes to be villages with geographical area, 50 ha. or less, villages with geographical area more than 50 ha. & up to 100 ha. and villages having geographical area more than 100 ha. A sample of 6 villages is to be selected by taking 2 villages in each stratum. Compare the efficiency of stratified sampling with corresponding unstratified simple random sampling.

Contd.

2. EXAMPLE 1

Village Sl.No.	01	02	03	04	05	06	07	08	09	10
Geographical Area (in ha.)	020	080	050	100	150	070	020	250	220	010
Village Sl.No.	11	12	13	14	15	16	17	18	19	20
Geographical Area (in ha.)	050	140	080	020	050	030	070	090	100	220

SOLUTION :

Clearly, N = 20, n = 6

Population Mean $\overline{Y}_N = \frac{1}{20} \sum_{i=1}^{20} y_i = \frac{1820}{20} = 91\, ha$

Population Mean Square $S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \overline{Y}_N\right)^2$

$$S^2 = \frac{1}{N-1}\left(\sum_{i}^{N} y_i^2 - N\overline{Y}_N^2\right) = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i^2 - N.\overline{Y}_N^2\right) = \frac{96180}{19} = 5062\, ha^2$$

Sampling Variance of Simple Random Sample Mean is given by

Contd.

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 = 590\ ha^2$$

Now stratify the population according to given strata sizes into three strata as follows:

Stratum Number	Allocated sampling units							
I $(\leq 50ha)$	020	050	020	010	050	020	050	030
II $(\succ 50ha\ and \leq 100ha)$	080	100	070	080	070	090	100	
III $(\succ 100ha)$	150	250	220	140	220			

Clearly,

$$N_1 = 8, \quad \bar{Y}_{N_1} = 031.3\,ha, \quad S_1^2 = 0270\,ha^2$$

$$N_2 = 7, \quad \bar{Y}_{N_2} = 084.3\,ha, \quad S_2^2 = 0161\,ha^2$$

$$N_3 = 5, \quad \bar{Y}_{N_3} = 196.0\,ha, \quad S_3^2 = 2330\,ha^2$$

Contd.

From each stratum, a sample of 2 villages is to be selected so $n_1 = n_2 = n_3 = 2$.

Now

$$V(\bar{y}_{st}) = \sum_{i=1}^{K} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_i^2 = 67\, ha^2$$

Obviously the stratification has reduced the sampling variance of the sample mean from 590 ha^2 (in case of simple random sampling) to 67 ha^2 (in case of stratified sampling) i.e. a reduction of about 89 per cent.

3. ALLOCATION OF SAMPLE SIZE TO DIFFERENT STRATA

In stratified sampling, having decided the strata and the sample size, the next question which a survey statistician has to face is regarding the method of selection within each stratum and the allocation of sample to different strata. The expression for the variance of stratified sample mean shows that the precision of a stratified sample for given strata depends upon the n_i's which can be fixed at will.

Contd.

The guiding principle in the determination of the n_i's is to choose them in such a manner so as to provide an estimate of the population mean with the desired degree of precision for a minimum cost or to provide an estimate with maximum precision for a given cost, thus making the most effective use of the available resources. The allocation of the sample to different strata made according to this principle is called the principle of optimum allocation.

The cost of a survey is a function of strata sample sizes just as the variance. The manner in which cost will vary with total sample size and with its allocation among the different strata will depend upon the type of survey. In yield estimation surveys, the major item in the survey cost consists of labour charges for harvesting of produce and as such survey cost is found to be approximately proportional to the number of crop cutting experiments (CCE). Cost per CCE may, however, vary in different strata depending upon labour availability. Under such situations, the total cost may be represented by

$$C = \sum_{i=1}^{K} c_i n_i \qquad (3.1)$$

Contd.

where, c_i is the cost per CCE in the i-th stratum. When c_i is same from stratum to stratum, say c, the total cost of a survey is given by

$$C = c \, . \, n \qquad (3.2)$$

The cost function will change in form, if travel cost, field staff salary, statistical analysis etc. are to be paid for.

The optimum values of n_i can be obtained by minimizing the $V(\bar{y}_{st})$ for fixed cost C as

$$\frac{P_i S_i}{\sqrt{n_i}} = \sqrt{\mu \, \mathrm{c_i}} \, n_i \quad (i=1,2,\ldots,K) \qquad \text{or} \qquad n_i = \frac{P_i S_i}{\sqrt{\mu . c_i}} \quad (i = 1,2,...,K)$$

(3.3)

where μ is some constant.

From the above, one can easily infer that:

Contd.

where, c_i is the cost per CCE in the i-th stratum. When c_i is same from stratum to stratum, say c, the total cost of a survey is given by

$$C = c \cdot n \qquad (3.2)$$

The cost function will change in form, if travel cost, field staff salary, statistical analysis etc. are to be paid for.

The optimum values of n_i can be obtained by minimizing the $V(\bar{y}_{st})$ for fixed cost C as

$$\frac{P_i S_i}{\sqrt{n_i}} = \sqrt{\mu \, c_i} \, n_i \quad (i=1,2,\ldots,K) \qquad \text{or} \qquad n_i = \frac{P_i S_i}{\sqrt{\mu . c_i}} \quad (i= 1,2,...,K) \qquad (3.3)$$

where μ is some constant.

From the above, one can easily infer that:

Contd.

where μ is some constant.

From the above, one can easily infer that:

- the larger the stratum size, the larger should be the size of the sample to be selected from that stratum ;
- the larger the stratum variability, the larger should be the size of the sample from that stratum and
- the cheaper the cost per sampling unit in a stratum, the larger should be the sample from that stratum.

The exact value of n_i for maximising precision for a fixed cost C_0 can be obtained by evaluating $\frac{1}{\sqrt{\mu}}$ the constant of proportionality as

The exact value of n_i for maximising precision for a fixed cost C_0 can be obtained by evaluating $\frac{1}{\sqrt{\mu}}$ the constant of proportionality as

$$n_i = \frac{PS_{ii}}{\sqrt{c_i}} \frac{C_0}{\sum_{i=1}^{K} P_i S_i \sqrt{c_i}} \qquad (3.4)$$

where C_0 is the total cost of the survey. The total sample size, n, is given by

Contd.

$$n = \sum_{i=1}^{K} n_i = \frac{C_0 \sum_{i=1}^{K} (P_i S_i / \sqrt{c_i})}{\sum_{i=1}^{K} P_i S_i \sqrt{c_i}} \qquad (3.5)$$

The allocation of sample size 'n' according to equation (3.5) is known as optimum allocation.

When c_i is the same from stratum to stratum i.e. $c_i = c$ (say), the cost function takes the form C = c . n, or in other words, the cost of survey is proportional to the size of the sample, the optimum values of n_i 's are given by

$$n_i = n \frac{P_i S_i}{\sum_{i=1}^{K} P_i S_i} \qquad (3.6)$$

The allocation of the sample according to the above formula is known as Neyman Allocation. On substituting for n_i in $V(\bar{y}_{st})$ expression, we obtain

Contd.

$$V_N(\bar{y}_{st}) = \frac{1}{n}(\sum_{i=1}^{K} P_i S_i)^2 - \frac{1}{N}\sum_{i=1}^{K} P_i S_i^2 \qquad (3.7)$$

where the subscript N stands for the stratification with Neyman Allocation.

As S_i 's being population values, are not known, so another approach of allocation is to allocate larger sample sizes for larger strata i.e. $n_i \alpha N_i$ or $n_i = n.\frac{N_i}{N}$ (i=1,2,...,K). The allocation of sample size 'n' according to above is known as proportional allocation and $V(\bar{y}_{st})$ in this case becomes

$$V_P(\bar{y}_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{K} P_i S_i^2 \qquad (3.8)$$

where the subscript P indicates the stratification with proportional allocation.

4. EXAMPLE 2

As follows are the number of pepper standards for the randomly selected villages in each of the three strata of Trivandrum Zone.

Stratum	Total number of villages in the stratum	Number of villages selected from the stratum	Number of pepper standards in each of the selected villages
1	441	11	41, 116, 19, 15, 144, 159, 212, 57, 28, 76, 119
2	405	12	39, 70, 38, 37, 161, 38, 27, 119, 36, 128, 30, 208
3	103	7	252, 386, 192, 296, 115, 159, 120

Estimate the total number of pepper standards along with its standard error in Trivandrum Zone. Also, estimate the gain in precision due to stratification.

SOLUTION:

For calculating estimate of variance, the following Tables are prepared:

Stratum	N_i	n_i	$\sum_j y_{ij}$	$\bar{y}_{n_i}$	$N_i \bar{y}_{n_i}$	$\sum_j y_{ij}^2$	$\sum_j y_{ij} / n_i$	s_i^2
1	441	11	986	89.63	39529.81	130654	88381.45	4227.25
2	405	12	743	91.92	25481.22	70469	46004.08	2224.08
3	103	7	1520	217.71	22365.71	389886	330057.14	9971.48
Total	N=949	n=30			87376.74			

Stratum	$\left(\frac{1}{n_i} - \frac{1}{N_i}\right)$	$N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)s_i^2$	$N_i\ s_i^2$	$\frac{N_i\ s_i^2}{n_i}$	$N_i\ \bar{y}_{n_i}^2$	$\frac{N_i^2\ s_i^2}{n_i}$
1	0.0886	72839892.71	1864219.48	169474.48	3543275.94	74738245.68
2	0.0809	29512745.97	900753.74	75062.81	1577720.88	30400438.62
3	0.1331	14080301.61	1027062.04	146723.15	4856556.93	15112484.24
Total		116432940.29	3792035.06	391260.44	9977553.75	120251168.54

Contd.

Estimate of total number of pepper standards is given by

$$\hat{Y}_{st} = \sum_{i=1}^{K} N_i \, \bar{y}_{n_i} = 87376.54$$

The estimate of variance of $\hat{Y}_{st}$ is

$$\hat{V}\left(\hat{Y}_{st}\right) = \sum_{i=1}^{K} N_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) s_i^2 = 116432940.29$$

So standard error of $\hat{Y}_{st} = \sqrt{116432940.29} = 10790.41$

Estimate of variance for total number of pepper standards when sampling has been done using SRSWOR

$$\hat{V}\left(\hat{Y}_{SRS}\right) = \left(\frac{1}{n} - \frac{1}{N} \right) \left[N \left\{ \sum_{i=1}^{K} N_i s_i^2 - \sum_{i=1}^{K} \frac{N_i \, s_i^2}{n_i} + \sum_{i=1}^{K} N_i \bar{y}_{n_i}^2 \right\} + \sum_{i=1}^{K} \frac{N_i s_i^2}{n_i} - \left(\sum_{i=1}^{K} N_i \bar{y}_{n_i} \right)^2 \right]$$

$$= 167269559.37$$

Contd.

Thus, the percentage gain in precision due to stratification

$$= \frac{\hat{V}\left(\hat{Y}_{SRS}\right) - \hat{V}\left(\hat{Y}_{st}\right)}{\hat{V}\left(\hat{Y}_{st}\right)} \text{ x } 100$$

$$= \frac{167269559.37 - 116432940.29}{116432940.29} x100 = 43.65\%$$

5. EXAMPLE 3

Data from a complete census of 2,000 farms under wheat crop (in hectares) in a region are given in the following Table. The farms were stratified on the basis of farm size into seven strata. The total number of farms (N_i) in each stratum is given. The population strata averages for area under wheat per farm ($\bar{Y}_{N_i}$) and the corresponding standard deviations (S_i) for each stratum are also given:

Contd.

Stratum	Farm size (ha.)	Total number of farms (N_i)	Average area under wheat (ha.) ( )	Standard Deviation of area under wheat (ha.) (S_i)
1	00-16	394	5.4	8.3
2	16-32	461	16.3	13.3
3	32-48	381	24.3	15.1
4	48-64	334	34.5	19.8
5	64-80	169	42.1	24.5
6	80-96	113	50.1	26.0
7	> 96	148	63.8	35.2

(A) Allocate the total sample of 200 farms

i) in proportion to the size of the strata (proportional allocation),

ii) in proportion to the product N_i S_i (Neyman Allocation).

(B) Calculate the sampling variance of the estimate of average area under wheat for the region based on a random sample of 200 farms,

i) if the farms were selected by simple random sampling without stratification,

ii) if the farms were selected by simple random sampling within each stratum and allocated in proportion to (a) N_i and (b) N_i .S_i in each stratum.

(C) Also, obtain the relative gains in efficiency resulting from (a) and (b) of (ii) as compared to (i) as well as of (b) as compared to (a).

Contd.

Solution: Here N=2000, n=200.

We prepare the following Table:

Stratum number	No. of holdings (N_i)	P_i	Mean Area ($\overline{Y}_{N_i}$)	S_i	$N_i\overline{Y}_{N_i}$	P_iS_i	nP_i	$n\frac{P_iS_i}{\sum_{i=1}^{K}P_iS_i}$	$\left(\overline{Y}_{N_i}-\overline{Y}\right)^2$	$P_i\left(\overline{Y}_{N_i}-\overline{Y}\right)^2$	$P_iS_i^2$
1	394	0.1970	5.4	8.3	2127.6	1.64	40	19	437.65	86.22	13.57
2	461	0.2305	16.3	13.3	7514.3	3.06	46	36	100.04	23.06	40.77
3	381	0.1905	24.3	15.1	9258.3	2.88	38	34	4.08	0.78	13.44
4	334	0.1670	34.5	19.8	11523.0	3.31	33	39	66.91	11.17	65.47
5	169	0.0845	42.1	24.5	7114.9	2.07	17	24	249.01	21.04	50.72
6	113	0.0565	50.1	26.0	5661.3	1.47	11	17	565.49	31.95	38.19
7	148	0.0740	63.8	35.2	9442.4	2.61	15	31	1404.75	103.95	91.69
Total	2000	1.0000				17.05	200	200		278.17	313.85

Contd.

(A) (i) When sample from each stratum is to be selected in proportional to the size of the stratum, i.e. $n_i \alpha N_i$ or $n_i = n.P_i$ from the Table, the number of villages to be selected from the seven strata are 40,46,38,33,17,11 and 15 respectively, total sample size being 200 villages. For example,

$$n_3 = n\frac{N_3}{N} = 200 \times \frac{381}{2000} = 38.1 \Rightarrow n_3 = 38$$

(A) (ii) In case, the sampling units from each stratum are selected in proportion to

N_i S_i , the sample size n_i is given as

$$n_i = n.\frac{P_i S_i}{\sum_{i=1}^{K} P_i S_i}$$

and here the sample of 200 villages is to be allocated to the seven strata as 19, 36, 34, 39, 24, 17 and 31 respectively. For example,

Contd.

$$n_1 = n\frac{N_1 S_1}{\sum_{i=1}^{K} N_i S_i} = 200 \times \frac{3270.2}{34055.9} = 19.20 \Rightarrow n_1 = 19$$

(B) (i) In case of simple random sampling without stratification, variance of the estimate of population mean, when N_i and N are sufficiently large, is given by

$$V_{srs}(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\left[\sum_{i=1}^{K} \frac{N_i - 1}{N - 1} S_i^{\ 2} + \sum_{i=1}^{K} \frac{N_i}{N - 1}\left(\bar{Y}_{N_i} - \bar{Y}\right)^2\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\left[\sum_{i=1}^{K} P_i S_i^2 + \sum_{i=1}^{K} P_i\left(\bar{Y}_{N_i} - \bar{Y}\right)^2\right].$$

On substituting the values from the above Table, we get

Contd.

$$V_{srs}(\bar{y}_n) = (\frac{1}{200} - \frac{1}{2000})(313.85 + 278.17) = 2.6641$$

Therefore, standard error of the estimated average area

$$SE_{srs}(\bar{y}_n) = \sqrt{V_{srs}(\bar{y}_n)} = \sqrt{2.6641} = 1.63$$

(B)(ii) (a) When sample from each stratum is selected in proportion to the size of the stratum i.e. proportional allocation, the variance of the estimated mean is given by

$$V_P(\bar{y}_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} P_i S_i^2$$

and hence, in this case

$$V_P(\bar{y}_{st}) = \left(\frac{1}{200} - \frac{1}{2000}\right) x\, 313.85 = 1.4123$$

Contd.

$$V_{srs}(\bar{y}_n) = (\frac{1}{200} - \frac{1}{2000})(313.85 + 278.17) = 2.6641$$

Therefore, standard error of the estimated average area

$$SE_{srs}(\bar{y}_n) = \sqrt{V_{srs}(\bar{y}_n)} = \sqrt{2.6641} = 1.63$$

(B)(ii) (a) When sample from each stratum is selected in proportion to the size of the stratum i.e. proportional allocation, the variance of the estimated mean is given by

$$V_P(\bar{y}_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} P_i S_i^2$$

and hence, in this case

$$V_P(\bar{y}_{st}) = \left(\frac{1}{200} - \frac{1}{2000}\right) x\, 313.85 = 1.4123$$

Contd.

On substituting the values from the Table,

$$V_N(\bar{y}_{st}) = \frac{1}{200} x (17.05)^2 - \frac{1}{2000} x 313.85 = 1.2966$$

Therefore, standard error of the estimated average area

$$SE_N(\bar{y}_{st}) = \sqrt{V_N(\bar{y}_{st})} = \sqrt{1.2966} = 1.14$$

(C) The relative gain in efficiency of using stratified sampling when sample from each stratum is allocated in proportion to N_i , over simple random sampling

$$RE_{PA/SRS} = \frac{V_{srs}(\bar{y}_n) - V_P(\hat{\bar{Y}}_{st})}{V_P(\bar{y}_{st})} x 100$$

$$= \frac{2.6641 - 1.4123}{1.4123} x 100 = 46.99\%$$

Contd.

The relative gain in efficiency of using stratified sampling when sample from each stratum is allocated in proportion to $N_i S_i$, over simple random sampling

$$RE_{NA/SRS} = \frac{V_{srs}\left(\bar{y}_n\right) - V_N\left(\bar{y}_{st}\right)}{V_N\left(\bar{y}_{st}\right)} x100$$

$$= \frac{2.6641 - 1.2966}{1.2966} x100 = 51.33\%$$

The relative gain in efficiency of using stratified sampling with Neyman allocation as compared to using proportional allocation

$$RE_{NA/PA} = \frac{V_P\left(\bar{y}_{st}\right) - V_N\left(\bar{y}_{st}\right)}{V_N\left(\bar{y}_{st}\right)} x100$$

$$= \frac{1.4123 - 1.2966}{1.2966} x100 = 8.19\%$$

Contd.

THANKS FOR YOUR ATTENTION