Methods of sample allocation to different strata :

1. Equal allocation
2. Proportional allocation
3. Optimum allocation

Proportional allocation

Sample size for h-th stratum in case of proportional allocation :

$$n_h = \frac{n}{N} N_h \tag{5.9}$$

Total sample size for fixed total cost :

$$n = \frac{C - c_0}{\sum_{h=1}^{L} W_h c_h} \tag{5.10}$$

Example 5.4

The third student of the group of four, was independently assigned the estimation problem of example 5.1, and was asked to use proportional allocation method . Using the budget and cost information of example 5.2, determine the total number of students that he could afford to select. Also, allocate the sample units to different strata.

Solution

In this case, we have $N_1 = 1300$, $N_2 = 450$, $N_3 = 250$, $N = 2000$, $L = 3$, $C = \$150$, $c_0 = \$24$, and $c_1 = c_2 = c_3 = \$3$. The total sample size that could be possible with the given information, is obtained by using (5.10). As $W_h = N_h/N$, the expression (5.10) can be written as

$$n = \frac{N\,(C - c_0)}{\sum_{h=1}^{L} N_h c_h}$$

$$= \frac{(2000)\,(150 - 24)}{(1300)\,(3) + (450)\,(3) + (250)\,(3)}$$

$$= 42$$

The total number of 42 students to be included in the sample are allocated to each of the 3 strata through (5.9). Thus,

$$n_1 = \left(\frac{n}{N}\right) N_1 = \left(\frac{42}{2000}\right)(1300) = 27.3 \approx 27$$

$$n_2 = \left(\frac{n}{N}\right) N_2 = \left(\frac{42}{2000}\right)(450) = 9.5 \approx 10$$

$$n_3 = \left(\frac{n}{N}\right) N_3 = \left(\frac{42}{2000}\right)(250) = 5.3 \approx 5$$

Therefore, 27, 10, and 5 students would be selected from strata I, II, and III respectively. ■

Example 5.5

In order to estimate the parameters of example 5.1, the total sample size that could be possible with the given budget has been obtained, in example 5.4, as 42 along with its proportional allocation to different strata. Accordingly, the student investigator selected 27 students from stratum I, 10 from stratum II, and 5 from stratum III. The information collected from the students in the sample is given in table 5.5.

Table 5.5 Time (in hours) devoted to study in library during a week

Stratum I					Stratum II		Stratum III
4	5	11	3	2	5	12	18
3	6	0	8	1	9	8	20
10	4	1	2	7	7	10	17
6	9	10	4		16	17	23
8	3	5	6		11	7	10
1	12	4	5				

Estimate the parameters of example 5.1.

Solution

For computations, we prepare table 5.6.

Table 5.6 Calculated values of various statistics for data given in table 5.5

Stratum I			Stratum II			Stratum III		
n_1	=	27	n_2	=	10	n_3	=	5
N_1	=	1300	N_2	=	450	N_3	=	250
W_1	=	..650	W_2	=	.225	W_3	=	.125
$\bar{y}_1$	=	5.185	$\bar{y}_2$	=	10.200	$\bar{y}_3$	=	17.600
s_1^2	=	10.851	s_2^2	=	15.289	s_3^2	=	23.300

Using (5.1), and various values from table 5.6, we find

$$\bar{y}_{st} = \frac{1}{N}(N_1\bar{y}_1 + N_2\bar{y}_2 + N_3\bar{y}_3)$$

$$= \frac{1}{2000}[1300\,(5.185) + 450\,(10.200) + 250\,(17.600)]$$

$$= 7.865$$

The estimate of variance $V(\bar{y}_{st})$ is computed by using (5.3) as

$$v(\bar{y}_{st}) = \frac{W_1^2(N_1 - n_1)\,s_1^2}{N_1 n_1} + \frac{W_2^2\,(N_2 - n_2)\,s_2^2}{N_2 n_2} + \frac{W_3^2(N_3 - n_3)\,s_3^2}{N_3 n_3}$$

$$= \frac{(.650)^2\,(1300 - 27)\,(10.851)}{(1300)\,(27)} + \frac{(.225)^2\,(450 - 10)\,(15.289)}{(450)\,(10)}$$

$$+ \frac{(.125)^2\,(250 - 5)\,(23.300)}{(250)\,(5)}$$

$$= .1663 + .0757 + .0714$$

$$= .3134$$

Utilizing (2.8), we work out the confidence interval from

$$\bar{y}_{st} \pm 2\sqrt{v(\bar{y}_{st})}$$

$$= 7.865 \pm 2\sqrt{.3134}$$

$$= 6.745,\ 8.985$$

Thus, on the average, a PAU student devotes 7.865 hours per week to study in the library. So far as the actual population average is concerned, we are reasonably confident that it will fall in the closed interval [6.745, 8.985] hours. ■

If the cost per unit is same for all the strata, that is, $c_h = c'$ for each h, optimum allocation is known as *Neyman allocation*, after Neyman (1934). For this case, the cost function (5.6) takes more simpler form

$$C = c_0 + c'n \tag{5.13}$$

Minimum variance – Neyman allocation :

$$\left.\begin{aligned} n_h &= n \frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h} \\ &= n \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \end{aligned}\right] \tag{5.14}$$

Total sample size :

$$n = \frac{C - c_o}{c'} \tag{5.15}$$

Example 5.7

The fourth student, in the group of four, was asked to undertake the estimation of parameters considered in examples 5.1 to 5.5 by using Neyman allocation. Again, the cost for contacting the students and gathering the required information is $ 3 per student. The total budget at disposal was $150 including the overhead cost of $24. Using Neyman allocation, the samples of sizes 25, 10, and 7 students were selected from strata I, II, and III respectively (procedure of determining these sample sizes is explained in the solution). The data collected from these three WOR simple random samples, selected from three strata, are presented in table 5.8. Using this information, estimate the average time per week devoted to study in library by a student. Also, set up confidence interval for the population mean. The information on strata mean squares is to be used from example 5.5.

Solution

We are given that

$C = \$150$, $c_o = \$24$, $c' = \$ 3$, $N_1 = 1300$, $N_2 = 450$, $N_3 = 250$, $S_1^2 = 10.851$, $S_2^2 = 15.289$, and $S_3^2 = 23.300$.

Taking cost into account, the total sample size from (5.15) will be

$$n = \frac{C - c_o}{c'} = \frac{150 - 24}{3} = 42$$

Now,

$$\sum_{h=1}^{3} N_h S_h = N_1 S_1 + N_2 S_2 + N_3 S_3$$

$$= (1300)(\sqrt{10.851}) + (450)(\sqrt{15.289}) + (250)(\sqrt{23.300})$$

$$= 7248.615$$

The sample sizes for different strata are then determined using (5.14), where

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

Thus,

$$n_1 = (42)\,\frac{(1300)\left(\sqrt{10.851}\right)}{7248.615} = 24.81 \approx 25$$

$$n_2 = (42)\,\frac{(450)\left(\sqrt{15.289}\right)}{7248.615} = 10.20 \approx 10$$

$$n_3 = (42)\,\frac{(250)\left(\sqrt{23.300}\right)}{7248.615} = 6.99 \approx 7$$

The observations recorded from these selected students are given below in table 5.8.

Table 5.8 Time (in hours) devoted to study in library by selected students during a week

Stratum I					Stratum II		Stratum III	
9	6	8	6	10	9	16	24	25
1	7	3	9	5	14	6	18	22
3	2	5	2	3	13	8	11	
5	4	4	5	4	8	12	19	
4	6	4	1	7	12	11	16	

For convenience, we compute sample estimates for mean and mean square error for each stratum. These are given along with other required information in table 5.9.

Table 5.9 Necessary computations for strata I, II, and III

Stratum I	Stratum II	Stratum III
$n_1 = 25$	$n_2 = 10$	$n_3 = 7$
$N_1 = 1300$	$N_2 = 450$	$N_3 = 250$
$W_1 = .650$	$W_2 = .225$	$W_3 = .125$
$\bar{y}_1 = 4.920$	$\bar{y}_2 = 10.900$	$\bar{y}_3 = 19.286$
$s_1^2 = 5.993$	$s_2^2 = 9.656$	$s_3^2 = 23.892$

By using figures from table 5.9 in (5.1), we work out the estimate of weekly average study time in the library as

$$\bar{y}_{st} = \frac{1}{N}(N_1\bar{y}_1 + N_2\bar{y}_2 + N_3\bar{y}_3)$$

$$= \frac{1}{2000}[1300(4.920) + 450(10.900) + 250(19.286)]$$

$$= 8.061$$

Next we compute estimate of variance $V(\bar{y}_{st})$ from (5.3) as

$$v(\bar{y}_{st}) = \frac{W_1^2(N_1 - n_1)s_1^2}{N_1 n_1} + \frac{W_2^2(N_2 - n_2)s_2^2}{N_2 n_2} + \frac{W_3^2(N_3 - n_3)s_3^2}{N_3 n_3}$$

$$= \frac{(.650)^2(1300 - 25)(5.993)}{(1300)(25)} + \frac{(.225)^2(450 - 10)(9.656)}{(450)(10)}$$

$$+ \frac{(.125)^2(250 - 7)(23.892)}{(250)(7)}$$

$$= .0993 + .0478 + .0518$$

$$= .1989$$

The required confidence interval for population mean is then obtained from

$$\bar{y}_{st} \pm 2\sqrt{v(\bar{y}_{st})}$$

$$= 8.061 \pm 2\sqrt{.1989}$$

$$= 7.169,\ 8.953$$

One can claim that there is approximately 95% chance that the population value of the average weekly study time in the university library will be covered by the closed interval [7.169, 8.953] hours. ■

5.5 RELATIVE EFFICIENCY OF STRATIFIED ESTIMATOR

For examining the usefulness of stratification, we need the sampling variances of the estimators of population mean/total for stratified and unstratified population. The percent relative efficiency of the estimator $\bar{y}_{st}$, with respect to the usual unstratified estimator $\bar{y}$, is then given by

$$RE = \frac{V(\bar{y})}{V(\bar{y}_{st})} (100)$$

where $V(\bar{y})$ and $V(\bar{y}_{st})$, for WOR sampling, are defined in (3.9) and (5.2) respectively. We illustrate below, the various steps involved in the calculation of the above said actual percent relative efficiency for three commonly used sample allocation methods. A relative efficiency figure of well over 100 indicates that the stratification of the population would be effective in reducing the estimation error. For this purpose, we consider a hypothetical situation where the study variable values are known for all population units.

HW

- 5.6 ,5.8 and 5.9, and 5.11 assuming fixed cost.