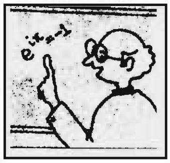
Classroom



In this section of Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.*

R Vasudeva, Department of Statistics, University of Mysore, Mysore 570 006, India.

Stratified Sampling – When the Optimum Allocation Demands More than 100 Percent Sampling!

What can go wrong in an optimization problem when the mathematical model used does not incorporate a constraint that the solution should satisfy? As an example of this it is shown that under a stratified sampling scheme, the optimal allocation procedure can give a sample size allocated to a stratum that exceeds the stratum size itself. The resolution of such a problem is discussed.

In stratified random sampling, the whole population of N units is divided into a finite number of strata, of sizes $N_1, N_2,$ $N_k, k > 1$. Samples are taken from all the strata and the population parameters are estimated based on the sampled observations. One of the criteria of determination of the sample size allocation to each stratum is the minimization of the variance of the estimator. This is done (i) when the total sample size is given or (ii) when the budget allotted for the survey and an appropriate cost model are given. The cost model usually assumed is $C = C_0 + \sum_{h=1}^{k} n_h c_h$, where C denotes the total capital available for the survey, C_0 the overhead cost, c_h the cost of surveying a unit in the hth stratum and n_h the number of units surveyed in the hth stratum.

When the scheme of selection is simple random sampling without replacement (SRSWOR), W G Cochran, in Section 5.8

of his book, *Sampling Techniques*, (3rd edition, Wiley, 1977) explains through an example that under the setup (i) the optimal allocation (Neyman allocation) may produce sample sizes allocated to some strata exceeding the respective stratum sizes. If the number n_h-optimal (n_{oh}), exceeds $N_h, h = h_1, h_2, \cdots, h_r, r < k$, then the solution given is to completely enumerate all the units in such strata and to determine the optimal allocation in the remaining strata by taking $n - \sum_{j=1}^{r} N_{h_j}$ as the sample size. It is shown below that a similar situation may also arise under the setup (ii).

Let $(\sigma_h')^2 = N_h\sigma_h^2/(N_h - 1)$, where σ_h^2 is the variance of the hth stratum, $h = 1, 2, \cdots k$. Then under SRSWOR scheme of selection, an estimate of the population mean is given by

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{y}_h, \tag{1}$$

where $\bar{y}_h$ is the sample mean in the hth stratum. Also, the variance of the estimate is

$$V(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \left[\frac{1}{n_h} - \frac{1}{N_h}\right] (\sigma_h')^2. \tag{2}$$

Under the cost model $C = C_0 + \sum_{h=1}^{k} n_h c_h$, the allocation which minimizes $V(\hat{\bar{Y}}_{st})$ is given by

$$n_h = (C - C_0) \frac{N_h \sigma_h' / c_h^{1/2}}{\sum_{i=1}^{k} N_i \sigma_i' c_i^{1/2}}, \qquad h = 1, 2, \cdots, k. \tag{3}$$

Now consider the example where a population is divided into 3 strata as shown below.

Stratum No. (h)	Stratum size (N_h)	σ_h'	c_h
1	10	2	1.96
2	50	1	1.0
3	50	1	1.0

For $C = 110$ and $C_0 = 10$, the allocation scheme which minimizes $V(\hat{\bar{Y}}_{st})$ yields $n_1 = 11$ (see (3) above); clearly this is more than $N_1 (= 10)$.

Comments

The basic purpose of stratification is to achieve homogeneous strata. However, in practice, a given stratified set-up could be far from ideal. If some of the strata are indeed heterogeneous and further the cost of surveying is rather low in those strata then the sample size allocations to those strata obtained by the mathematical model would be naturally high. At times the sample size allocated to a particular stratum may even exceed the stratum size. A simple explanation for such an anomaly is that the mathematical model used to derive optimal allocation does not incorporate the condition $0 < n_h \leq N_h$, $1 \leq h \leq k$.

This can happen in any optimization problem. A mathematical model may not incorporate a constraint that the optimal solution should satisfy. If the optimal solution so obtained does not satisfy the desired constraint then it is necessary to take certain correcting steps. For our problem, or the one discussed in Cochran's book (1977), whenever the allocated sample size exceeds the corresponding stratum size, the entire stratum is enumerated. For the remaining strata a modified optimization problem is formulated. Another more recent book *Model Assisted Survey Sampling* by Sarndal, Svenson and Wretman (Springer-Verlag. New York, 1992) also discusses this very same problem in its Remark 12.7.1.

T S K V Iyer, Birla Institute of Technology and Science, Pilani, Rajasthan, India

Increase of Entropy and the Arithmetic–Geometric Mean

It is instructive to work out the change in entropy when two identical blocks of the same material but at different temperatures are brought together into thermal contact. The system as a whole is thermally isolated.

Let T_1 and T_2 be the absolute temperatures of the two bodies. Assume $T_1 > T_2$. When they attain thermal equilibrium both will reach a common temperature T given by:

$$T_1 - T = T - T_2$$