Visit us - www.researchjournal.co.in ■ DOI : 10.15740/HAS/IRJAES/8.2/435-439

International Research Journal of Agricultural Economics and Statistics

Volume 8 | Issue 2 | September, 2017 | 435-439 ■ e ISSN-2231-6434

Research Paper

On optimum stratification using mathematical programming approach

■ FAIZAN DANISH AND S.E.H. RIZVI

See end of the paper for authors' affiliations

Correspondence to :
FAIZAN DANISH
Division of Statistics and Computer Science, Faculty of Basic Sciences,
SKUAST-J, Main Campus Chatha, JAMMU (J&K) INDIA
Email: danishstat@gmail.com

Paper History :
Received : *25.05.2017;*
Revised : *22.08.2017;*
Accepted : *29.08.2017*

ABSTRACT : Optimum stratification is a technique which results in minimum possible variance of the estimator for the population characteristic under study. The main objective of stratification is to give a better cross-section of the population so as to gain a higher degree of relative precision. The problem of determining optimum strata boundaries (OSB) was pioneered by Dalenius (1950). The problem of obtaining OSB was recently studied by Khan *et al.* (2009) who formulated the problem as a mathematical programming problem (MPP) by minimizing variance of the estimated population parameter subject to the condition that the sum of the widths of all the strata would be equal to the range of the given distribution under given allocation procedure. In the present investigation the problem of finding OSB has been taken into consideration as the problem of optimum strata width (OSW), using MPP by dynamic programming technique, when the study variable is uniformly distributed. Empirical study has also been taken where it is revealed that with the increase in the number of strata to a fixed number the precision of the method goes on increasing. Also the proposed method proves better than other stratification method (Singh,1967).

KEY WORDS : Mathematical programming problem, Optimum stratification, Optimum strata boundaries, Optimum strata width

HOW TO CITE THIS PAPER : Danish, Faizan and Rizvi, S.E.H. (2017). On optimum stratification using mathematical programming approach. *Internat. Res. J. Agric. Eco. & Stat.,* **8** (2) : 435-439, **DOI : 10.15740/HAS/IRJAES/8.2/435-439.**

INTRODUCTION :

In the sample surveys, the population being investigated may be homogenous or heterogeneous one with respect to characteristic under study. In the latter case stratified random sampling is generally used for selecting the samples. In this technique the whole population is divided into various homogenous sub-populations, known as strata through stratification. The stratification technique which results in minimum possible variance is called optimum stratification. Thus, the optimum stratification of a population consists in dividing the joint domain of stratification variables in such a way that the precision of the estimates is maximum. In achieving this goal, it is usually required that this division be done by cutting the domain of each stratification variable into different intervals. Such stratification has been referred to as interval optimum stratification by Isii and Taga (1969).

The main objective of stratification is to give a better cross-section of the population so as to gain a higher degree of relative precision (Cochran, 1977). The use of stratified sample survey basically involves five different operations:

– The choice of the stratification variable(s).

– The choice of number of strata.

– The determination of the way in which the population is to be stratified.

– Allocation of sample size to each stratum.

– Choice of sampling design in each stratum.

The problem of determining the optimum strata boundaries (OSB),when the study variable itself is used as stratification variable, was first discussed by Dalenius (1950) who obtained solution by using minimal equations, but the exact solution of these equations are not possible because of their implicit nature. Dalenius and Gurney (1951) showed that in some cases the increase in the number of strata leads to a loss in precision, if stratification is not well chosen. Several researchers have attempted to find out approximate solutions such as Mahalanobis (1952); Aoyama (1954); Dalenius and Hodge (1959); Ekman (1959); Serfling (1968) and Singh (1971).

Many authors like Unnithan (1978); Lavallee and Hidiroglou (1988); Hidiriglou and Srinath (1993); Sweet and Sigman (1995); Rivest (2002) and Gupta *et al.* (2005) suggested some iterative procedures to determine OSB. The algorithms require an initial approximation solution to strata as also there is no guarantee that the algorithm which are used will provide the global minimum in the absence of a suitable approximate initial solution. Rizvi *et al.* (2002) tackled the problem of optimum stratification for two study variables using one auxiliary variable as stratification variable.Gunning *et al.*(2004) proposed an alternative approach to approximate stratification. Kozak and Verma (2006) concluded superiority of optimization approach over approximate stratification. Mathematical programming problem (MPP) is a technique for obtaining optimum solution of a problem subject to given constraints. This approach was adopted by Khan *et al.* (2009) in order to determine OSB using auxiliary information. They formulated the problem as an MPP when the number of strata is fixed in advance. By suitable transformation they converted the problem into a multistage decision problem in which at each stage the value of a single decision variable is worked out using dynamic programming problem.

In the present study, study variable itself has been used as stratification variable which is presumed to be distributed uniformly. Further allocation for fixed sample size is considered. The problem been formulated as MPP.

Formulation of the problem :

Let us consider that the population under consideration is divided into L homogeneous non-overlapping strata in order to estimate the population mean. Let y_0 and y_L be the lower and upper bounds of the study variable Y of the given population. Then optimum stratification can be described in order to find the intermediate stratum boundaries $y_1 \leq y_2 \leq .... \leq y_{L-1}$ and

Variance of the sample mean $\bar{y}_{st} Z \overset{L}{\underset{hZ1}{\dot{g}}} W_h \bar{y}_h$
is given by

$$V(\bar{y}_{st}) = \frac{N(\sum_{h=1}^{L} W_h \sigma_h)^2 - n(\sum_{h=1}^{L} W_h \sigma_h^2\ )}{nN} \qquad ..(1)$$

is minimum, presuming that Neyman allocation is used.

Here $\bar{y}_h$ is the sample mean, $W_h = \frac{N_h}{N}$ is the weight and σ_h^2 is the variance of the mean, hth stratum. If the finite population correction (fpc) is ignored, then the minimization of variance given by (1) reduces to minimization of :

$$\sum_{h=1}^{L} W_h \sigma_h \qquad .....(2)$$

The problem of determining OSB is then equivalent to finding L^{-1} intermediate points $y_1 \leq y_2 \leq \dots \leq y_{L-1}$ in the interval $[y_0, y_L]$ such that variance of the sample mean is minimum. Now let:

$$y_L J y_0 Z t \qquad ....(3)$$

Let g (y) be the frequence distributio of the study variable and (y_{h-1}, y_h) be the boundaries of the hth stratum, then:

$$W_h = \int_{y_{h-1}}^{y_h} g(y)dy \qquad ....(4)$$

$$\sigma_h^2 = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} y^2 g(y)dy - \mu_h^2 \qquad ....(5)$$

where,

$$\mu_h = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} yg(y)y \qquad ....(6)$$

For known frequency distribution g (y) of the study variable, (2) can be expressed as a function of boundaries of the hth stratum.

Let $g_h (y_h, y_{h-1}) = W_h \pm_h$ ....(7)

Now using (7) in (2), we get as:

$$\sum_{h=1}^{L} g_h (y_h, y_{h-1}) \qquad ....(8)$$

Define Z_h, the width of the hth stratum, as:

$$Z_h = y_h - y_{h-1},\ h=1,2...,L \qquad ....(9)$$

where, $Z_h \geq 0$

Using (9) we can express (3) as:

$$\sum_{h=1}^{L} Z_h = \sum_{h=1}^{L} (y_h - y_{h-1})$$

$= y_l - y_0 = t$

Using kth stratification point

$y_k = y_0 + z_1 + ... + z_k$, $k=1,2,...L-1$ ....(10)

Now the problem of determining OSB can be expressed as the following MPP

Minimize subject to constraint $\sum_{h=1}^{L} g_h(y_h, y_{h-1})$ ....(11)

$$\sum_{h=1}^{L} Z_h = t$$

and $Z_h \geq 0$, $h=1,2,...,L$

Since y_0 is the initial value of the study variable, the first term, g_1 (z_1, y_0) is the objective function of MPP (11) which is a function of Z_1 only. Similarly the second term g_2 (z_2, y_1) = g_2 (z_2,y_0+ z_1) is a function of Z_2 alone once Z_1 is known. Thus, stating the objective function as a function of Z_h alone we may replace MPP (11) as:

Minimize subject to constraint $\sum_{h=1}^{L} g_h(Z_h)$ ....(12)

$$\sum_{h=1}^{L} Z_h = t$$

and $Z_h \geq 0$ where, $h=1,2,...,L$

Determination of OSB of uniform study variable :

Let the stratification variable Y follows uniform distribution with probability density function (pdf) as:

$$g(y) Z \begin{cases} \frac{1}{b J a}, & a \underline{Y} y \leq b \\ 0, & \end{cases} \quad ......(13)$$

atherwise

Note that here y_0= a and y_L =b

Now (4), (5) and (6) can be written as:

$$W_h Z i_{y_h J1}^{y_h} \frac{1}{b J a} dy$$

$$Z \frac{Z_h}{b J a} \quad ......(14)$$

$$\mu_h Z \frac{1}{W_h} i_{y_h J1}^{y_h} y \frac{1}{b J a} dy$$

$$= \frac{Z_h}{2} \quad ......(15)$$

Thus, the variance expressed can be given as :

$$\sigma_h^2 Z \frac{1}{W_h} i_{x_h J1}^{x_h} y^2 g(y)\, dy \; J \; \mu_h^2$$

$$Z \frac{(Z_h)^2}{12} \quad ....(16)$$

Substituting the values of W_h and σ_h obtained in equations (14) and (15) for uniform distribution, the problem of determining OSB given by (12) can be expressed as:

Minimize subject to constraint $\underset{hZ1}{\overset{L}{g}} \frac{Z_h^2}{2\sqrt{3}(b J a)}$ ....(17)

$$\sum_{h=1}^{L} Z_h = t$$

and $Z_h \geq 0$, $h=1,2,...,L$

where, ‘t’ is obtained by (3) with y_0 = a and y_L =b

Procedure for obtaining optimum solution :

Let us consider the fallowing sub-problem of (11) for first i strata

Minimize subject to constraint $\sum_{h=1}^{i} g_h(Z_h)$ ....(18)

$$\sum_{h=1}^{i} Z_h = t_i$$

and $Z_h \geq 0$, $h=1,2,...,L$

where, $t_i < t$ is the total width available of the division into i strata.

we have

$t_i = t$ for $i = L$

Also

$t_i \quad = z_1 + z_2 + ... + z_i$

$t_{i-1} \quad = z_1 + z_2 + ... + z_{i-1}$

$= t_i - z_i$

$t_{i-2} \quad = z_1 + z_2 + ... + z_{i-2}$

$= t_{i-1} - z_{i-1}$

K

K

K

$t_1 \quad = z_1$

$= t_2 - z_2$

If g (i, t_i) denote the minimum value of the objective function (18), then recurrence relation of the dynamic programming take the form as:

$$g(i,t_i) = \underset{0 \le z_i \le t_i}{\text{Min}} \left[g_i(z_i) + g(i-1, t_i - z_i) \right] \qquad \text{....(19)}$$

Obviously for i=1, using (18)

$$g_1(1, t_1) = g_1(t_1) \qquad \text{....(20)}$$

$$= z_1 = t_1$$

Similarly, for, $i \geq 3$, we have

$$g(i-1, t_{i-1}) = \underset{0 \le z_{i-1} \le t_{i-1}}{\text{Min}} \left\{ g_{i-1}(z_{i-1}) + g(i-2, t_{i-1} - z_{i-1}) \right\}$$

Thus, Z_L is obtained from $g(L-1, t_L - z_L)$ as the optimum width of (L-1)th stratum, Z_{L-1} is obtained $g(L-2, t_{L-1} - z_{L-1})$ as the optimum width of (L-2)th stratum and so on until Z_1 is obtained.

Now, using (19) and (20) in (21), we get

$$g(1,t_1) = \frac{t_1^2}{2\sqrt{3(b-a)}}, \text{for } i = 1 \qquad \text{....(21)}$$

at $Z_1 = t_1$

as $y_i = y_0 = 0, \text{if } i=1$

For ith stage, where i $\geq$ 2

$$g(i,t_i) = \underset{0 \le z_i \le t_i}{\text{Min}} \left\{ \frac{Z_i^2}{2\sqrt{3(b-a)}} + g(i-1, t_i - z_i) \right\} \qquad \text{....(22)}$$

Because $y_{t-1} = y_0 + z_1 + ... + z_{i-1}$

$$= t_i - z_i$$

Similarly for $i \geq 3$

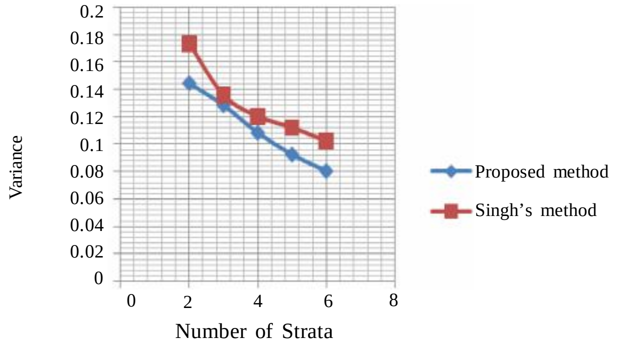

Fig. 1 : Graphical representation of the Table 1

Table 1 : OSW, OSB and optimum value of the objective function for uniform distribution

No. of strata L	Strata width $Z\bar{h}$ [1]	Strata boundary points $y\bar{h} = y\bar{h} - 1 + Z\bar{h}$ [2]	Optimum value of objective function $\sum_{h=1}^{L} W_h \sigma_h$ [3]	Variance obtained by Singh (1967) [4]	% R.E. of [3] over [4]
2	$Z\bar{1} = 0.5000$ $Z\bar{2} = 0.5000$	$y_{\bar{1}} = y_0 + Z\bar{1} = 0.5000$	0.1443	0.1732	120.02
3	$Z\bar{1} = 0.3333$ $Z\bar{2} = 0.3333$ $Z\bar{1} = 0.3333$	$y_{\bar{1}} = y_0 + Z\bar{1} = 0.3333$ $y_{\bar{1}} = y\bar{1} + Z\bar{1} = 0.6666$	0.1283	0.1356	140.95
4	$Z\bar{1} = 0.2500$ $Z\bar{2} = 0.2500$ $Z\bar{2} = 0.2500$ $Z\bar{4} = 0.2500$	$y_{\bar{1}} = y_0 + Z\bar{1} = 0.2500$ $y_{\bar{2}} = y_1 + Z\bar{2} = 0.5000$ $y_{\bar{2}} = y_{\bar{2}} + Z\bar{1} = 0.7500$	0.1082	0.1200	166.43
5	$Z\bar{1} = 0.2000$ $Z\bar{2} = 0.2000$ $Z\bar{2} = 0.2000$ $Z\bar{1} = 0.2000$ $Z\bar{3} = 0.2000$	$y_{\bar{1}} = y_0 + Z\bar{1} = 0.2000$ $y\bar{2} = y\bar{1} + Z\bar{1} = 0.4000$ $y\bar{2} = y\bar{2} + Z\bar{2} = 0.6000$ $y\bar{4} = y\bar{2} + Z\bar{4} = 0.8000$	0.0923	0.1118	193.76
6	$Z\bar{1} = 0.1666$ $Z\bar{2} = 0.1666$ $Z\bar{2} = 0.1666$ $Z\bar{4} = 0.1666$ $Z\bar{5} = 0.1666$ $Z\bar{6} = 0.1666$	$y\bar{1} = y_0 + Z\bar{1} = 0.1666$ $y\bar{2} = y\bar{1} + Z\bar{2} = 0.3333$ $y\bar{2} = y\bar{2} + Z\bar{1} = 0.4999$ $y\bar{4} = y\bar{2} + Z\bar{4} = 0.6665$ $y\bar{5} = y\bar{4} + Z\bar{5} = 0.8332$	0.0802	0.1072	222.86

$$\mathbf{g\,(i-1,t_{i-1}) = \underset{0 \le z_{i-1} \le t_{i-1}}{Min} \left[\frac{Z_{t-1}^2}{2\sqrt{3(b-\alpha)}} + g(i-2,t_{i-1}-z_{i-1}) \right]} \quad ...(23)$$

Numerical illustrations :

The theoretical procedures discussed under section 4 have been illustrated numerically for obtaining the strata width, strata boundary points and optimal value of the objective function. Further, the efficiency comparison has also been made for comparing the proposed method with that of Singh (1967). The results are presented in Table 1.

Conclusion :

In this paper it can be concluded that the proposed method variance has an decreasing trend with respect to the number of strata as shown above in graph told be anyone however, for convenience uniform distribution has been taken into consideration.Also the relative efficiency suggests that the proposed method leads to have more gain in precision than the Singh's method.

Authors' affiliations:
S.E.H. RIZVI, Division of Statistics and Computer Science, Faculty of Basic Sciences, SKUAST-J, CHATHA (J&K) INDIA

LITERATURE CITED :

Aoyama, H. (1954). A study of stratified random sampling. *Annl. Instit. Statist. Mathemat.*, **6** : 1-36.

Cochran, W.G. (1977). *Sampling techniques.* 3rd Ed. John Wiley and Sons, Inc., NEW YORK, U.S.A.

Dalenius, T. (1950). The problem of optimum stratification. *Skandinavisk Aktuarietidskrift*, **33** : 203-213.

Dalenius, T. and Gurney, M. (1951). The problem of optimum stratification II. *Skandinavisk Aktuarietidskrift*, **34** : 133-148.

Dalenius, T. and Hodges, J.L. Jr. (1959). Minimum variance stratification. *J. American Statist. Assoc.*, **54** : 88-101.

Ekman, G. (1959). Approximation expression for the conditional mean and variance over small intervals of a continuous distribution. *Annl. Mathemat. Statist.*, **30** : 1131-1134.

Gupta, R. K., Singh, R. and Mahajan, P. K. (2005).Approximate optimum strata boundaries for ratio and regression estimators. *Aligarh J. Statist.*, **25** : 49-55.

Hidiroglou, M.A. and Srinath, K.P. (1993). Problems associated to sub annual business surveys. *J. Business & Econ. Statist.*, **11** : 397-405.

Isii, K. and Taga, Y. (1969). On optimal stratification for multivariate distributions. *Skand. Akt.*, **52** : 24-38.

Khan, E. A. Khan, M. G. M. and Ahsan, M. J. (2002).Optimum stratification:A mathematical programming approach. *Calcutta Statist. Assoc. Bull.*, **52** : 323-333.

Khan, M.G.M., Ahmad, N. and Khan, S. (2009). Determination the optimum stratum boundaries using mathematical programming. *J. Mathemat. Modelling & Algorithms*, **8** : 409-423.

Kozak, M. and Verma, M.R. (2006). Geometric versus optimization stratification: A comparison of efficiency. *Survey Methodology,* **32**(2) : 157-163.

Lavallee, P. and Hidiriglou, M. (1988). On the stratification of skewed populations. *Survey Methodology*, **14** : 33-43.

Mahalanobis, P.C. (1952). Some aspect of design of sample surveys. *Sankhya*, **12** : 1-17.

Rivest, R.J. (2002). A generalization of Lavallee and Hidiroglou algorithm for stratification in survey. *Survey Methodology*, **28** : 191-198.

Rizvi, S.E.H., Gupta, J.P and Bargava, M. (2002). Optimum stratification based on auxiliary variable for compromise allocation. *Metron*, **28** (1) : 201-215.

Serfling, R.J. (1968). Approximately optimum stratification. *J. American Statist. Assoc.*, **63** : 1298-1309.

Singh, R. (1967). Some contributions to the theory of construction of Strata. Ph.D. Thesis, Indian Agriculture Research Institute, NEW DELHI, INDIA.

Singh, R. (1971). Approximately optimum stratification on auxiliary variable. *J. American Statist. Assoc.*, **66** *:* 829-833.

Sweet, E.M. and Sigman, R.S. (1995).Evaluation of model-assisted procedures for stratifying skewed populations using auxiliary data. *Proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria*, 491–496.

Unnithan, V.K.G. (1978). The minimum variance boundary points of stratification. *Sankhya*, **40** (C) : 60-72.