Computer Program in MATLAB for Optimum Allocation for Multivariate Stratified Sampling

Manoj Kumar Sharma, Mewar University, Gangrar, Chittorgarh, Rajasthan-312 901 India.

Mohd. Vaseem Ismail PhD, Jamia Hamdard, New Delhi, India.

Vinod Kumar, Delhi Technological University, New Delhi, India.

ABSTRACT

In this paper we develop a MAT LAB computer program for the optimum allocation for multivariate stratified sampling with non-linear cost function –travel cost. The generalized MATLAB program for solving multi variate or multi objective linear programming based on with some major modifications in earlier algorithms. The proposed modified algorithm as well as MATLAB program simplifies the earlier algorithm on Multivariate linear programming problems. The problem of determining the optimum allocations are formulated as Nonlinear Programming Problems , in which each NLPP has a convex objective function and a non-linear cost constraint. The NLLP's are then solved using Lagrange Multiplier technique and the explicit formula for variance is obtained.

Key words: Non linear programming, Multivariate, Stratified sampling, optimum allocation

1. INTRODUCTION

For sample survey, we have extensively used Stratified sampling as it is the most popular among various sampling designs. When a stratified sampling is to be used a we have to deal with three basic problems such as (i) the problem of determining the number of strata, (ii) the problem of cutting the stratum boundaries and (iii) the problem of optimum allocation of sample sizes to various strata. In stratified sampling we have chosen the values of the sample sizes n_h in the respective strata. They may be selected to minimize $V\left(\bar{y}_{st}\right)$ for a specified cost of taking the sample or to minimize the cost for a specified value of $V\left(\bar{y}_{st}\right)$.

The general cost function is of the form

$$\text{Cost } = C = c_o + \sum_{h=1}^{L} c_h n_h^{\alpha}$$

Within any stratum the cost is proportional to the size of sample, but the cost per unit c_h may vary from stratum to stratum. The term c_o represents an overhead cost. If travel costs between units are substantial, empirical and mathematical studies suggest that travel costs are better represented by the expression $\sum_{h=1}^{L} t_h \sqrt{n_h}$ if $\alpha = 1/2$ and c_h is replaced by t_h where t_h is the travel cost per unit (Beardwood et al., 1959).

The method of optimum allocation for multivariate stratified sampling is developed for the non-linear cost function. The problem of determining the optimum allocations are formulated as Nonlinear Programming Problems , in which each NLPP has a convex objective function and a non-linear cost constraint. Several techniques are available for solving these NLPP's , we used Lagrange Multiplier technique to solve the optimum allocation of the value of sample size n_h, at different values of α

2. FORMULATION OF THE PROBLEM

when $C = c_o + \sum_{h=1}^{L} t_h \sqrt{n_h} \quad (\alpha = 1/2)$

In stratified random sampling with a linear cost function, the variance of the estimated mean $\bar{y}_{st}$ is a minimum for a specified cost C, and the cost is a minimum for specified variance $V\left(\bar{y}_{st}\right)$ when $n_h \propto W_h S_h / \sqrt{c_h}$

Suppose that p characteristics are measured on each unit of a population which is partitioned into L strata. Let n_h be the number of units to be drawn with out replacement from the h^{th} stratum $\left(h = 1,2,...,L\right)$. For the j^{th} character an unbiased estimate of the population mean $\bar{Y}_j$ is $\bar{y}_{jst}$ whose variance is given by

$$V\left(\bar{y}_{jst}\right) = \sum_{h=1}^{L} W_h^2 S_{hj}^2 X_h, \qquad j = 1,2,....,p \qquad (2.1)$$

where $W_h = \frac{N_h}{N}, S_{hj}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(y_{hji} - \bar{Y}_j\right)^2$

and $X_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right)$, in usual notations.

Let t_h be the travel cost of enumerating all the p characters on a single unit in the h^{th} stratum. The total cost of survey may be given as

$$C = c_o + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \qquad (2.2)$$

Where c_0 is the overhead cost. If we consider that each sample has constant measurement cost then $\sum_{h=1}^{L} c_h n_h = c \sum_{h=1}^{L} n_h = nc$ which can be merged in c_0 then the equation (2.2) reduces to

$$C = c_o + \sum_{h=1}^{L} t_h \sqrt{n_h} \qquad (2.3)$$

For a fixed budget C_o, the problem of determining an optimum allocation may be expressed as the following NLPP:

Minimize $\quad Z = \sum_{h=1}^{L} W_h^2 S_{hj}^2 X_h$

Subject to $\quad \sum_{h=1}^{L} t_h \sqrt{n_h} \le C_0$

and
$1 \le n_h \le N_h; \qquad h = 1, 2, \ldots\ldots, L$

where $\quad C_o = C - c_o$.

The optimum choice of n_h for an individual characteristic can be determined by minimizing the variance in (2.1) for the given cost in (2.3), or by minimizing the cost for fixed variance. We can use Lagrange multipliers technique to determine the optimum value of n_h.

The Lagrange function ϕ is defined as

$$\phi(n_h, \lambda) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2 + \lambda \left(\sum_{h=1}^{L} t_h \sqrt{n_h} - C_o \right) \qquad (2.5)$$

where λ is a Lagrange multiplier.

The necessary conditions for the solution of the problem are

$$\frac{\partial \phi}{\partial n_h} = -\sum_{h=1}^{L} \frac{1}{n_h^2} W_h^2 S_h^2 + \frac{1}{2} \lambda \sum_{h=1}^{L} \frac{t_h}{\sqrt{n_h}} = 0 \qquad (2.6)$$

$$\frac{\partial \phi}{\partial \lambda} = \sum_{h=1}^{L} t_h \sqrt{n_h} - C_o = 0 \qquad (2.7)$$

Solving (2.6) we get

$$n_h = \left(\frac{2 W_h^2 S_h^2}{\lambda t_h} \right)^{2/3} \qquad (2.8)$$

$$\sum_{h=1}^{L} n_h = n \qquad (2.9)$$

(2.9) gives

$$\lambda = \frac{2}{n^{3/2}} \left\{ \sum_{h=1}^{L} \left(\frac{W_h^2 S_h^2}{t_h} \right)^{2/3} \right\}^{3/2} \qquad (2.10)$$

Putting the value of λ in (2.8) we get

$$n_h = n \left(\frac{W_h^2 S_h^2}{t_h} \right)^{2/3} \sum_{h=1}^{L} \left(\frac{t_h}{W_h^2 S_h^2} \right)^{2/3} \qquad (2.11)$$

This gives

$$V(\bar{y}_{st}) = \frac{1}{n} \left\{ \sum_{h=1}^{L} (W_h S_h t_h)^{2/3} \sum_{h=1}^{L} \left(\frac{W_h^2 S_h^2}{t_h} \right)^{2/3} \right\} \qquad (2.12)$$

(2.4)

Ignoring the term $\frac{1}{N_h}$

3. FORMULATION OF THE PROBLEM

when $\quad C = c_o + \sum_{h=1}^{L} c_h n_h^2 \quad (\alpha = 1)$

Using Lagrange function we get

$$n_h = \left(\frac{W_h^2 S_h^2}{\lambda C_h} \right)^{1/2} \qquad (3.1)$$

$$\lambda = \frac{1}{n^2} \left[\sum_{h=1}^{L} \left(\frac{W_h^2 S_h^2}{C_h} \right)^{1/2} \right]^2 \qquad (3.2)$$

$$V(\bar{y}_{st}) = \frac{1}{n} \sum_{h=1}^{L} (W_h^2 S_h^2 C_h)^{1/2} \sum_{h=1}^{L} \left(\frac{W_h^2 S_h^2}{C_h} \right)^{1/2} \qquad (3.3)$$

4. FORMULATION OF THE PROBLEM

when $C = c_o + \sum_{h=1}^{L} c_h n_h^2 \;\; (\alpha = 2)$

Using Lagrange function we get

$$n_h = n\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/3} \sum_{h=1}^{L}\left(\frac{c_h}{W_h^2 S_h^2}\right)^{1/3} \qquad (4.1)$$

$$\lambda = \frac{1}{2n^3}\left[\sum_{h=1}^{L}\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/3}\right]^{1/3} \qquad (4.2)$$

$$V\left(\bar{y}_{st}\right) = \frac{1}{n}\left\{\sum_{h=1}^{L}\left(W_h^4 S_h^4 c_h\right)^{1/3} \sum_{h=1}^{L}\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/3}\right\} \quad (4.3)$$

5. FORMULATION OF THE PROBLEM

when $C = c_o + \sum_{h=1}^{L} c_h n_h^3 \;\; (\alpha = 3)$

Using Lagrange function we get

$$n_h = n\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/4} \sum_{h=1}^{L}\left(\frac{c_h}{W_h^2 S_h^2}\right)^{1/4} \qquad (5.1)$$

$$\lambda = \frac{1}{3n^4}\left[\sum_{h=1}^{L}\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/4}\right]^{1/4} \qquad (5.2)$$

$$V\left(\bar{y}_{st}\right) = \frac{1}{n}\left\{\sum_{h=1}^{L}\left(W_h^6 S_h^6 c_h\right)^{1/4} \sum_{h=1}^{L}\left(\frac{W_h^2 S_h^2}{c_h}\right)^{1/4}\right\} \quad (5.3)$$

6. NUMERIC EXAMPLES

Consider a population divided in five strata with single characteristic under study for which the values of W_h, S_h, c_h are given in the following table

Table 1.1

Stratum h	W_i	S_h	C_h
1	0.40	4	1
2	0.30	5	2
3	.55	8	4
4	.65	7	5
5	.35	3	6

Let us fix the budget at 100 units

Solving variance for $(\alpha = 1/2,\ \alpha = 1, \alpha = 2, \alpha = 3)$, taking the value of $n = 1000$, we get

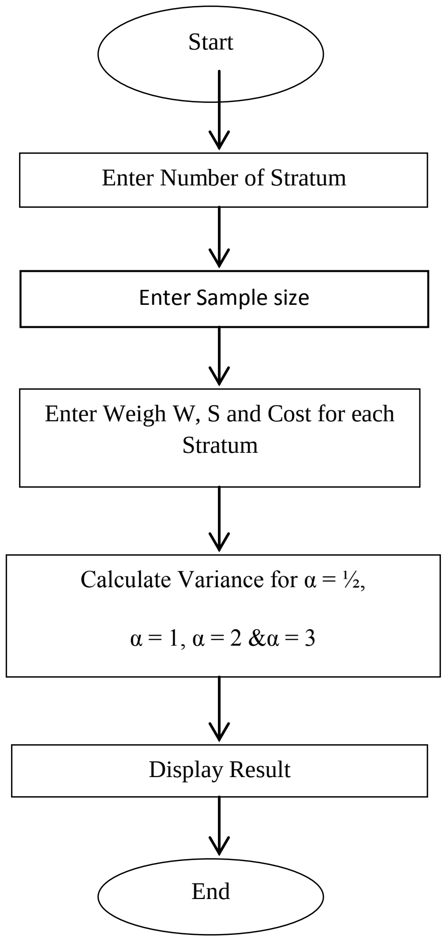
$$V\left(\bar{y}_{st\ \alpha=\frac{1}{2}}\right) = 0.1818$$

$$V\left(\bar{y}_{st\ \alpha=1}\right) = 0.1839$$

$$V\left(\bar{y}_{st\ \alpha=2}\right) = 0.1930$$

$$V\left(\bar{y}_{st\ \alpha=3}\right) = 0.1986$$

7. FLOW CHART FOR DEVELOPED ALGORITHM



8. CODING OF COMPUTER PROGRAM IN MAT LAB FOR DEVELOPED ALGORITHM

```
prompt={'Enter the number of Stratum:'};
name='Input for Goal Programming';
numlines=1;
defaultanswer={'0'};
answer=inputdlg(prompt,name,numlines,defaultanswer);
strings = char(answer);
Stratum_count = str2num(answer{1});
disp(Stratum_count);

prompt={'Enter the sample size:'};
name='Input for Goal Programming';
numlines=1;
defaultanswer={'0'};
answer=inputdlg(prompt,name,numlines,defaultanswer);
strings = char(answer);
Sample_count = str2num(answer{1});
disp(Sample_count);


val1(1,1) = 0;
val1(1,2) = 0;

val1(2,1) = 0;
val1(2,2) = 0;

val1(3,1) = 0;
val1(3,2) = 0;

val1(4,1) = 0;
val1(4,2) = 0;

for i = 1:Stratum_count,
    prompt={ 'Enter the S for the Stratum:',...
             'Enter the S for the Stratum:',...
              'Enter the Cost or the Stratum:'};
    name='Input for Stratum';
    numlines=1;
    defaultanswer={'0','0','0'};

answer=inputdlg(prompt,name,numlines,defaultanswer);
    strings = char(answer);
    Stratum_info(i,1) = str2num(answer{1});
    Stratum_info(i,2) = str2num(answer{2});
    Stratum_info(i,3) = str2num(answer{3});

    Stratum_info(i)
   % for j = 1:Stratum_count,
       val1(1,1) = val1(1,1) +
(Stratum_info(i,1)*Stratum_info(i,2)*Stratum_info(i,3))^(
0.667);
       val1(1,2) = val1(1,2) +
((Stratum_info(i,1)^2*Stratum_info(i,2)^2)/Stratum_info(i,
3))^(0.667);
   % end

   % for j = 1:Stratum_count,
       val1(2,1) = val1(2,1) +
((Stratum_info(i,1)^2)*(Stratum_info(i,2)^2)*Stratum_info
(i,3))^(0.5);
       val1(2,2) = val1(2,2) +
(((Stratum_info(i,1)^2)*(Stratum_info(i,2)^2))/Stratum_inf
o(i,3))^(0.5);
   % end

   % for j = 1:Stratum_count,
       val1(3,1) = val1(3,1) +
(Stratum_info(i,1)^4*Stratum_info(i,2)^4*Stratum_info(i,3
))^(0.334);
       val1(3,2) = val1(3,2) +
((Stratum_info(i,1)^2*Stratum_info(i,2)^2)/Stratum_info(i,
3))^(0.334);
   % end

   % for j = 1:Stratum_count,
       val1(4,1) = val1(4,1) +
(Stratum_info(i,1)^6*Stratum_info(i,2)^6*Stratum_info(i,3
))^(0.25);
       val1(4,2) = val1(4,2) +
((Stratum_info(i,1)^2*Stratum_info(i,2)^2)/Stratum_info(i,
3))^(0.25);
   % end
end

Result(1,1) = val1(1,1)*val1(1,2)/Sample_count;
Result(1,2) = val1(2,1)*val1(2,2)/Sample_count;
Result(1,3) = val1(3,1)*val1(3,2)/Sample_count;
Result(1,4) = val1(4,1)*val1(4,2)/Sample_count;
```

9. CONCLUSION

Therefore we can calculate the minimum variances for each stratum of given cost and optimum allocation of each stratum. First we enter the number of stratum then size of sample. Now we enter the variance of each character, weight of stratum and cost of stratum of this stratum. This can be done for each stratum. After run the program we will find optimum allocation and minimum variance for alpha values.

So for optimum allocation in multivariate stratified sampling we will minimize the cost by minimizing the variance

10. REFERENCES

[1] Dalenius, T. (1953) "The Multivariate Sampling Problem", Skandinavisk Actuarietidskrift, 36, 92-102.

[2] Yates,F.,(1960) “Sampling Methods for census and survey”.London:Charles Griffin & Company.

[3] Aoyama, H.,(1963) “Stratified random sampling with optimum allocation for multivariate populations”, Ann. Inst. Stat. Math., Vol. 14, pp. 251-258.

[4] Chatterjee, S., (1967) “A note on optimum allocation”, Scandinavian Actuarial Journal 50, 40-44.

[5] Kokan, A. R. and Khan, S.,(1967) “Optimum allocation in multivariate surveys: An analytical solution”, Journal of the Royal Statistical Society, Series B 29, 115-125.

[6] Chatterjee, S. (1972) "A Study of Optimum Allocation in Multivariate Stratified Surveys”, Skandinavisk Actuarietidskrift, 55, 73-80.

[7] Khan, M. G. M. and Ahsan, M. J.(2003) “A Note on Optimum Allocation in Multivariate Stratified Sampling”, South Pacific Journal of Natural Science, Vol. 21, pp. 91-95.

[8] Fu, Yan and Diwekar, U.M. (2004)" An efficient sampling approach to multiobjective optimization", Annals of Operation Research.,Vol. 132, pp. 109-134.

[9] Ismail, M.V., and Razzaq, A.,(2004) " Chebyshev solution to a multivariate stratified sampling problem", Pure and Applied Mathematika Sciences. LIX, pp. 1-2.

[10] SAMITA S. and KUMARI W.M.R.(2006) "Multivariate Based Stratification as an Alternative to Multi–stage Stratification in Stratified Random Sampling" Sri Lankan Journal of Applied Statistics, Vol. 7, p. 55-69.

[11] José Antonio, Diaz Garcia & Liliana Ulloa Cortez,(2006) "Optimum Allocation in multivariate stratified sampling: multi-objective programming", Comunicación Técnica No I-06-07/28-03-, (PE/CIMAT).

[12] José A. Diaz Garcia and Liliana Ulloa Cortez,(2008) "Multi-objective optimization for optimum allocation in multivariate stratified sampling", Statistics Canada, Catalogue No. 12-001-X, Vol. 34, No. 2, pp. 215-222, Survey Methodology.

[13] Khan. M. G. M., Khan, E.A. and Ahsan, M. J.(2008) "Optimum allocation in multivariate stratified sampling in presence of Non-response" J. Ind. Soc. Aggril. Statist, 62(1), 42-48.

[14] Charles Day D.(2010) "A Multi-Objective Evolutionary Algorithm for Multivariate Optimal Allocation", Section Survey Research Method – JSM, 3351-3358.