ABSTRACT

In this paper we develop a MATLAB computer program for the optimum allocation for multivariate stratified sampling with non-linear cost function – travel cost. The generalized MATLAB program for solving multivariate 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 $V(\bar{y}_a)$ for a specified cost of taking the sample or to minimize the cost for a specified value of $V(\bar{y}_a)$.

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 \( \alpha = \frac{1}{2} \) 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 $\alpha$.

2. FORMULATION OF THE PROBLEM

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

In stratified random sampling with a linear cost function, the variance of the estimated mean $\bar{y}_a$ is a minimum for a specified cost $C$, and the cost is a minimum for specified variance $V(\bar{y}_a)$ when

\[ n_h \propto \frac{W_h S_h}{\sqrt{c_h}} \]

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 ($h = 1, 2, ..., L$). For the $j^{th}$ character an unbiased estimate of the population mean $\bar{y}_j$ is $\bar{y}_{jst}$ whose variance is given by

\[ V(\bar{y}_j) = \sum_{h=1}^{L} W_h^2 S_{hj}^2 X_h \]

where $W_h = \frac{N_h}{N}, S_{hj}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hji} - \bar{y}_j)^2$
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 \(H \text{th} \) stratum. The total cost of survey may be given as

\[
C = c_o + \sum_{h=1}^{L} c_p n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \tag{2.2}
\]

Where \(c_o\) is the overhead cost. If we consider that each sample has constant measurement cost then

\[
\sum_{h=1}^{L} c_p n_h = c \sum_{h=1}^{L} n_h = nc \quad \text{which can be merged in} \quad c_o
\]

then the equation (2.2) reduces to

\[
C = c_o + \sum_{h=1}^{L} t_h \sqrt{n_h} \tag{2.3}
\]

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

Minimize \(Z = \sum_{h=1}^{L} W_h S_{bh} X_h\)

Subject to \(\sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_o\)

and

\(1 \leq n_h \leq N_h; \quad h = 1, 2, \ldots, L\)

where \(C_o = C - c_o\).

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 \(\phi\) is defined as

\[
\phi(n_h, \lambda) = \sum_{h=1}^{L} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) W_h S_{bh}^2 + \lambda \left( \sum_{h=1}^{L} t_h \sqrt{n_h} - C_o \right) \tag{2.5}
\]

where \(\lambda\) is a Lagrange multiplier.

The necessary conditions for the solution of the problem are

\[
\frac{\partial \phi}{\partial n_h} = -\sum_{h=1}^{L} \frac{1}{n_h^2} W_h S_{bh}^2 + \frac{1}{2} \lambda \sum_{h=1}^{L} t_h \sqrt{n_h} = 0 \tag{2.6}
\]

Solving (2.6) we get

\[
n_h = \left( \frac{2 W_h S_{bh}^2}{\lambda t_h} \right)^{\frac{1}{2}} \tag{2.8}
\]

\[
\sum_{h=1}^{L} n_h = n \tag{2.9}
\]

(2.9) gives

\[
\lambda = \frac{2}{n^{1/2}} \left( \sum_{h=1}^{L} \left( W_h S_{bh}^2 \right)^{1/2} \right)^{1/2} \tag{2.10}
\]

Putting the value of \(\lambda\) in (2.8) we get

\[
n_h = n \left( \frac{W_h S_{bh}^2}{t_h} \right)^{1/2} \sum_{h=1}^{L} \left( \frac{t_h}{W_h S_{bh}^2} \right)^{1/2} \tag{2.11}
\]

This gives

\[
V \left( \bar{y}_n \right) = \frac{1}{n} \left( \sum_{h=1}^{L} \left( W_h S_{bh}^2 \right)^{1/2} \sum_{h=1}^{L} \left( \frac{W_h S_{bh}^2}{C_h} \right)^{1/2} \right)^2 \tag{2.12}
\]

Ignoring the term \(\frac{1}{N_h}\)

### 3. FORMULATION OF THE PROBLEM

When \(C = c_o + \sum_{h=1}^{L} c_p n_h^2 (\alpha = 1)\)

Using Lagrange function we get

\[
n_h = \left( 
\frac{W_h S_{bh}^2}{\lambda C_h}
\right)^{1/2} \tag{3.1}
\]

\[
\lambda = \frac{1}{n T} \left[ \sum_{h=1}^{L} \left( \frac{W_h S_{bh}^2}{C_h} \right)^{1/2} \right]^2 \tag{3.2}
\]

\[
V \left( \bar{y}_n \right) = \frac{1}{n} \sum_{h=1}^{L} \left( W_h S_{bh}^2 \right)^{1/2} \sum_{h=1}^{L} \left( \frac{W_h S_{bh}^2}{C_h} \right)^{1/2} \tag{3.3}
\]
4. FORMULATION OF THE PROBLEM

when \( C = c_o + \sum \alpha c_r n_h^2 \quad (\alpha = 2) \)

Using Lagrange function we get

\[
n_h = n \left( \frac{W_h^2 S_h^2}{c_h} \right)^{\frac{1}{\alpha}} \sum_{h=1}^{L} \left( \frac{c_h}{W_h^2 S_h^2} \right)^{\frac{1}{\alpha}} \tag{4.1} \]

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

\[
V(\bar{y}_x) = \frac{1}{n} \left[ \sum_{h=1}^{L} \left( \frac{W_h^2 S_h^2}{c_h} \right)^{1/\alpha} \sum_{h=1}^{L} \left( \frac{W_h^2 S_h^2}{c_h} \right)^{1/\alpha} \right]^{1/2} \tag{4.3} \]

5. FORMULATION OF THE PROBLEM

when \( C = c_o + \sum \alpha c_r n_h^3 \quad (\alpha = 3) \)

Using Lagrange function we get

\[
n_h = n \left( \frac{W_h^2 S_h}{c_h} \right)^{\frac{1}{3}} \sum_{h=1}^{L} \left( \frac{c_h}{W_h^2 S_h} \right)^{\frac{1}{3}} \tag{5.1} \]

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

\[
V(\bar{y}_x) = \frac{1}{n} \left[ \sum_{h=1}^{L} \left( \frac{W_h^2 S_h}{c_h} \right)^{1/3} \sum_{h=1}^{L} \left( \frac{W_h^2 S_h}{c_h} \right)^{1/3} \right]^{1/4} \tag{5.3} \]

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>
<thead>
<tr>
<th>Stratum h</th>
<th>W_h</th>
<th>S_h</th>
<th>C_h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.30</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Let us fix the budget at 100 units

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

\[
V(\bar{y}_{st} \alpha = \frac{1}{2}) = 0.1818 \]

\[
V(\bar{y}_{st} \alpha = 1) = 0.1839 \]

\[
V(\bar{y}_{st} \alpha = 2) = 0.1930 \]

\[
V(\bar{y}_{st} \alpha = 3) = 0.1986 \]

7. FLOW CHART FOR DEVELOPED ALGORITHM
8. CODING OF COMPUTER PROGRAM IN MATLAB FOR DEVELOPED ALGORITHM

prompt={'Enter the number of Stratum:'};
name='Input for Goal Programming';
umlines=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';
umlines=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';
umlines=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});
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

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


