CHAPTER III

OPTIMUM ALLOCATION IN STRATIFIED SAMPLING USING
SEPARABLE PROGRAMMING

3.1 INTRODUCTION:

In stratified random sampling with L strata the sampler has to determine the sample sizes \( n_h \); \( h=1,2,\ldots,L \) in advance. The values of \( n_h \) may be determined either to maximize the precision of the estimate for given cost or to minimize the cost of the survey for fixed precision. The allocation of the sample sizes \( n_h \) according to the above criteria is known as optimum allocation.

The optimum allocation, when the cost of the survey is expressed as a linear function of \( n_h \), is well known in literature. In this chapter it is assumed that the cost function is quadratic in \( \sqrt{n_h} \). The lower and upper bounds on the values of \( n_h \) are also taken into account. The problem of optimum allocation is then formulated as a mathematical programming problem and its approximate solution is
obtained by using separable programming technique.

In multivariate stratified random sampling where more than one characters are to be measured on each sampled unit, no single optimality criterion is available to work out the optimum allocations. The section 3.2 of this chapter presents an overview of the optimum allocation in multivariate stratified sampling using various criteria.

In the subsequent sections of this chapter the problem of optimum allocation in stratified sampling for univariate case is formulated as a separable programming problem. A solution procedure is indicated by using Simplex method by approximating the nonlinear functions into linear functions. Further extension of the technique for the multivariate case is also indicated. The objective is to find sample sizes that minimizes the total cost of the survey for a desired precision of the estimated population mean. Apart from the measurement cost, the total cost of the survey also includes the travelling cost within strata.

3.2 AN OVERVIEW:

Consider the problem of optimum allocation in univariate stratified sampling in estimating the overall population mean;
\[
\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} \sum_{j=1}^{N_h} y_{hj} = \sum_{h=1}^{L} \bar{W}_h \bar{y}_h
\]

where \( y_{hj} \) = value of the \( j \)th population unit in the \( h \)th stratum; \( h = 1, 2, \ldots, L \); \( j = 1, 2, \ldots, N_h \)

\( N_h \) = size of the \( h \)th stratum

\( \bar{W}_h = \frac{N_h}{N} \) = stratum weight

\( \bar{y}_h = \frac{1}{N_h} \sum_{j=1}^{N_h} y_{hj} \) = stratum mean

\( N = \sum_{h=1}^{L} N_h \) = population size

If independent, without replacement simple random samples of sizes \( n_h \); \( h = 1, 2, \ldots, L \) are drawn to construct the unbiased estimate

\[
\bar{Y}_{ac} = \sum_{h=1}^{L} \bar{W}_h \bar{y}_h
\]

for the population mean \( \bar{Y} \) where
then the well known optimum allocation which minimizes the variance of $\bar{y}_s$ for fixed cost $C_0$ given by Neyman (1934) is

$$n_h = \frac{C_0 (N_h S_h/\sqrt{C_h})}{\sum_{h=1}^{L} (N_h S_h/\sqrt{C_h})} ; \quad h=1,2,\ldots,L \quad \ldots \quad (3.2.1)$$

where

$$C_0 = \sum_{h=1}^{L} c_h n_h \quad \text{is the total cost of the survey}$$

$$c_h = \text{cost of measuring per unit in the } h\text{th stratum}$$

$$S_h^2 = \frac{1}{N_h-1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2 \quad \text{stratum variance in the } h\text{th stratum}.$$
The problem of optimum allocation in multivariate stratified random sampling may be expressed in the following general way (Zacks (1970)).

Let \( \theta_h = \theta_h(y_{h1}, \ldots, y_{hn_h}) \); \( h=1,2,\ldots,L \) be specified parametric functions

where \( N_h = \text{size of the } h\text{th stratum} \)
and \( y_{nj} = \text{value of the } j\text{th population unit of the } h\text{th stratum} \).

Consider the problem of estimating \( p \) linearly independent functions

\[
L_j(Y) = \sum_{h=1}^{L} a_{jh} \theta_h; \quad j=1,2,\ldots,p
\]

from the stratified sample, where \( Y = (Y_1, \ldots, Y_n) \) is the vector of values in the population.

In univariate case, that is when \( p=1 \), the Neyman allocation given in (3.2.1) gives

\[
n_h \propto |a_{1h}| \phi(S_h^2, c_h); \quad h=1,2,\ldots,L
\]

where \( \phi(S_h^2, c_h) \) is specified function of \( S_h^2 \) and \( c_h \).
It can be seen that even if $S_h$ are known the Neyman allocation will vary for different functions $L_j; j = 1, 2, \ldots, p$. The problem is then to select a single optimality criterion which is suitable for estimating all $L_j$. Many criteria are available in literature to choose a reasonable allocation. Some of them are indicated below.

Dalenius (1953) suggested to minimize the total relative loss in precision by applying an allocation $n = (n_1, n_2, \ldots, n_L)$ other than Neyman allocation such that the cost of the survey $\sum_{h=1}^{L} C_h n_h \leq C_0$. The problem may be mathematically expressed as:

"Find the $L$ component vector $n = (n_1, n_2, \ldots, n_L)$ which Minimize

$$Q(n) = \frac{\sum_{j=1}^{L} \{V(L_j|n) - V_{opt}(L_j)\}}{V_{opt}(L_j)}$$

Subject to,

$$\sum_{h=1}^{L} C_h n_h \leq C_0$$

where $V(L_j|n)$ and $V_{opt}(L_j)$ denote the variances under an allocation $n = (n_1, n_2, \ldots, n_L)$ and the Neyman allocation.
Cochran (1963) suggested the use of the average of the individual optimum allocations for different characters under study as an alternative to the optimum allocation for all characters. He showed that due to the flatness of the variance function the average allocation gives results almost as precise as the individual optimum allocations.

Dalenius (1957) considered a weighted linear function of the loss in precision relative to optimum allocation of the type \( \sum_{j=1}^{p} w_j L_j \) where \( w_j \) are weights assigned to different characteristics according to their importance and discussed its minimization subject to cost constraint.

Ghosh (1958) worked out the multivariate optimum allocation by minimizing the generalized variance of the estimates of the population means for a fixed total sample size.

Yates (1960) suggested two criteria for working out optimum allocation in multivariate cases in the situations where the individual optimum allocations differ so much that there is no obvious compromise. The first approach applies to surveys where the loss due to an error in the estimates can be measured in terms of money or utility. For
estimating the population means, with $p$ variates and a quadratic loss function he expressed the total expected loss as a linear function of the variances of the estimates, that is

$$L = \sum_{j=1}^{p} a_j V(\bar{y}_{jst})$$

This expected loss is minimized under the linear cost constraint

$$C_0 + \sum_{h=1}^{L} c_h n_h \leq C$$

In the second approach, tolerance limits are specified for each variance and the total cost of the survey is minimized. In this case we have the problem of minimizing $C_0 + \sum_{h=1}^{L} c_h n_h$

under the $p$ constraints

$$V(\bar{y}_{jst}) \leq v_j \quad j=1,2,\ldots,p,$$

and the restriction

$$0 \leq n_h \leq N_h \quad h=1,2,\ldots,L,$$

where $v_j$ is the pre-fixed tolerance limit of $V(\bar{y}_{jst})$. 
Several authors e.g. Hartely and Hocking (1963), Kokan (1963), Chatterjee (1966), Zukhovitsky and Ardeyeva (1966), Huddleston, and et al (1970), etc. gave algorithms for solving such problems.

Kokan (1963) suggested the minimization of the total sampling cost

\[ C = C_0 + \sum_{h=1}^{L} C_h R_h \]

subject to the constraints on the individual variances \( V(\bar{L}_j|n) \), \( j=1,2,\ldots,p \). The condition that he imposes on each variance \( V(\bar{L}_j|x) \) are such that all \( \bar{L}_j \) satisfy the proportional closeness (relative precision) requirement, namely:

\[ P \left( \frac{|\bar{L}_j - L_j|}{\lambda |L_j|} \right) \geq 1-\alpha \quad \text{for all } j=1,2,\ldots,L, \]

0< \( \lambda <1 \), and 0< \( \alpha <1 \). In Kokan's approach the functions \( L_j \) do not have to be linearly independent and \( p \) can be greater than \( L \).

Folks and Antle (1965) solved the problem of allocation by determining the set of efficient points \( A(n_1,n_2,\ldots,n_L) \), where
Kokan and Khan (1967) gave an analytical solution using n-dimensional geometry to the problem of optimum allocation in multivariate stratified random sampling and in few other sampling procedures for estimating the population mean. The optimality criteria is similar to that of Yates (1960).

Huddleston, Claypool and Hocking (1970) determined the optimal sample allocation by using convex programming methods and compared the allocation with other existing allocation system.

Ahsan (1975-1976) gave a variation of the problem discussed in Kokan and Khan (1967) and worked out the solution on the same lines.

Ericson (1965) used prior information for obtaining optimum allocation in univariate stratified sampling. The multivariate case was also discussed under the assumption that the strata are sufficiently similar.
Ahsan and Khan (1977), Ahsan (1978), and Ahsan and Khan (1982) also developed algorithms to obtain optimum allocation in multivariate stratified random sampling using prior information.

Khan and Islam (1980) formulated the problem of optimum allocation when multiple characters are measured into a nonlinear programming problem with multiple objective functions. A solution has been given similar to the STEP method for linear programming.

Dayal (1982) gave a procedure of determining sample sizes and its optimum allocation to various strata under stratified random sampling when stratum level estimates are also required. He consider the desired precision in terms of relative variance for the hth stratum which is given by

\[ \{ V(\overline{Y}_h | \overline{Y}_h^2) \} \leq b_h ; \ h = 1, 2, \ldots, L \]

where \( b_h \)'s are given constants.

This gives \( n_h \geq a_h ; \ h = 1, 2, \ldots, L, \) where \( a_h \)'s are constants depending upon \( b_h, N_h \) and \( S_h. \) Thus the problem is to minimize \( V(\overline{y}_1) \) subject to the constraints,
\[
\sum_{h=1}^{h} n_h = n
\]

and \( a_h - n_h \leq 0 \)

Liao and Sendranksk (1983) determine the optimal allocation of a total sample of size \( n \) among the strata by maximizing for fixed \( n \) the probability \( P \), that specified variance constraints are satisfied and by minimizing the total sample size \( n \) for a desired level of \( P \). Both exact and approximate solutions are presented and compared using a sequence of numerical examples.

Rao (1984) gave an alternative method of deriving Neyman's optimum allocation. In order to obtain this allocation, the within stratum standard deviation of the study variable are required which are unknown and are substituted by known quantities. He also discussed the effects of deviations from optimum allocation and other related problems.

Mukherjee and Rao (1985) consider the problem of judging how good an actual compromise allocation is as compared to Neyman allocation by obtaining a bound to the ratio of the corresponding variances.

Dayal (1985) shows how the values of auxiliary characteristic linearly related to the study variable, can
be used in the allocation of the sample. He also found that proportional allocation can be more efficient than an approximation to Neyman's allocation by using estimates of standard deviations of the study variable from a previous survey or approximations to them from some variable related to the study variable.

Omule (1985) expressed and solved the procedure of determining the optimum sample size in each stratum in stratified sampling for several variables as a multistage decision process through dynamic programming.

Bethel (1989) using the Kuhn-Tucker theorem, gave an expression for the optimum allocation in terms of lagrangian multipliers. Using this representation, the partial derivative of the cost function with respect to the jth variance constraint is found to be,

\[
-2a_j^* g(x^*)
\]

\[
\frac{1}{v_j}
\]

where \( g(x^*) \) is the cost of the optimum allocation and \( a_j^* \) and \( v_j \) are respectively the jth normalized lagrangian multiplier and the upper bound on the precision of the jth variable. A simple computational algorithm and its convergence properties are also discussed.
3.3 FORMULATION OF THE PROBLEM:

Let the population of size $N$ be divided into $L$ strata of sizes $N_h; h=1,2,\ldots,L$ and that the sampling within each stratum is independent simple random sampling without replacement. Further, for the $h$th stratum, let,

- $N_h =$ the stratum size
- $n_h =$ the sample size
- $f_h = \frac{n_h}{N_h}$, the sampling fraction.

$y_{hj} =$ the value of the characteristic under study for the $j$th unit, $j=1,2,\ldots,N_h$.

$W_h = \frac{N_h}{N} =$ stratum weight

$\bar{Y}_h = \frac{1}{N_h} \sum_{j=1}^{N_h} y_{hj} =$ the stratum mean

$\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} \sum_{j=1}^{N_h} y_{hj}

= \sum_{h=1}^{L} W_h \bar{Y}_h

= \text{over all population mean}

$\bar{\bar{y}}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj},$ the sample mean based on $n_h$ units,
which is an unbiased estimate of $\bar{Y}_h$.

$$S_h^2 = \frac{1}{N_h-1} \sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^2$$

is the stratum variance

$$S_h^2 = \frac{1}{n_h-1} \sum_{j=1}^{n_h} (y_{hj} - \bar{Y}_h)^2$$

is the sample mean square

based on $n_h$ units, used as an unbiased estimate of $S_h^2$.

An unbiased estimator of $\bar{Y}$ is given by,

$$\bar{Y}_{st} = \sum_{h=1}^{L} w_h \bar{Y}_h$$

The variance of $\bar{Y}_{st}$, ignoring the fpc is,

$$V(\bar{Y}_{st}) = \sum_{h=1}^{L} \frac{w_h^2 S_h^2}{n_h}$$

(3.3.1)

If $W_h$ and $S_h$ are known, $V(\bar{Y}_{st})$ is a function of $n_h$ only. If $S_h$ are not known $s_h$ may be substituted for them.

The cost function is assumed to be of the form

$$C = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h}$$

(3.3.2)

where,

- $C$ = total cost of the survey
- $c_h$ = measurment cost per unit in the $h$th stratum
- $t_h$ = travelling cost within $h$th stratum.

If travel costs within strata are substantial Beardwood et
al. (1959) suggested that a fairly good approximation to the total travel cost is $\sum_{h=1}^{L} t_h \sqrt{n_h}$.

The problem is to find $n_h$ that minimizes (3.3.2) for a fixed tolerance limit on the value of $V(\bar{y}_{st})$ given in (3.3.1), where $n_h$ is bounded as, $2 \leq n_h \leq N_h$, $h=1,2,\ldots,L$. The upper limit on $n_h$ is put to avoid oversampling whereas the lower limit gives an opportunity to estimate the stratum variance if not known.

Consequently our allocation problem may be expressed as the following nonlinear programming problem (NLPP).

Minimize $C(n_h) = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h}$ \ldots (3.3.3)

Subject to,

$$\sum_{h=1}^{L} \frac{\hat{w}_h^2 \hat{S}_h^2}{n_h} \leq \nu \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.3.4)$$

and $2 \leq n_h \leq N_h \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.3.5)$

where $\nu$ is the desired tolerance limit for $V(\bar{y}_{st})$.

With $x_h = \sqrt{n_h}$, the nonlinear programming problem
(3.3.3)-(3.3.5) becomes,

Minimize \( C = \sum_{h=1}^{L} c_h x_h^2 + \sum_{h=1}^{L} t_h x_h \) \ldots \ldots (3.3.6)

Subject to,

\[
\sum_{h=1}^{L} \frac{w_h^2 s_h^2}{x_h^2} \leq V \ldots \ldots \ldots \ldots \ldots \ldots (3.3.7)
\]

\[
\sqrt{2} \leq x_h \leq \sqrt{N_h} \ldots \ldots \ldots \ldots \ldots \ldots (3.3.8)
\]

As \( c_h > 0, h=1,2,...L \), the objective function (3.3.6) is convex. The nonlinear constraint (3.3.7) is also convex. Hence any local minimum for NLPP (3.3.6) - (3.3.8) will be a global minimum. Both the objective function and the constraint function are separable in \( x_h \). The separability of the functions allow us to apply separable programming technique which will yield an approximate global minimum for the NLPP. The separable programming problem may be stated as:

Minimize \( \sum_{h=1}^{L} f_h(x_h) \) \ldots \ldots \ldots \ldots (3.3.9)
Subject to,

\[ \sum_{h=1}^{t} g_h(x_h) \leq v \]  \hspace{1cm} (3.3.10)

\[ \sqrt{2} \leq x_h \leq \sqrt{N_h} \]  \hspace{1cm} (3.3.11)

where \[ f_h(x_h) = c_h x_h^2 + t_h x_h \]

and \[ g_h(x_h) = \frac{w_h^2 s_h^2}{x_h^2} \]

3.4 APPROXIMATE SOLUTION USING SEPARABLE PROGRAMMING:

Any nonlinear function can be approximated by a piecewise linear function without much difficulty if the nonlinear function is well behaved. The basic approximation technique to be used will be that of replacing the functions \( f_h(x_h) \) and \( g_h(x_h) \) by linear approximations, thereby reducing the problem to a form which can be solved by simplex method with a simple modification (Hadley (1970), Kambo (1984)).

Let us define the subset \( H \) of the set \( \{1, 2, \ldots, L\} \) by \( H = \{ h : f_h(x_h) \text{ are linear} \} \). For each \( h \in H \), let the feasible range of the variable \( x_h \) be given by the interval \( [a_h, b_h] \) and let
us choose a set of $n_h$ grid points $a_{hr}$ ($r=1, 2, \ldots, n_h$) such that,

$$a_h = a_{h1} < a_{h2} < \ldots < a_{hn} = b_h$$

It should be noted that the grid points need not be equally spaced and that varying grid lengths can be used for the different variables. Let us consider the grid interval $[a_{hr}, a_{h, r+1}]$. Every point $x_h$ in this interval can be expressed uniquely as,

$$x_h = \lambda_{hr} a_{hr} + \lambda_{h, r+1} a_{h, r+1} \quad \ldots \ldots \quad (3.4.1)$$

where $\lambda_{hr} + \lambda_{h, r+1} = 1$

and $\lambda_{hr} \geq 0, \lambda_{h, r+1} \geq 0$

Each function $f_h(x_h), h \notin H$ can be approximated in the grid interval $[a_{hr}, a_{h, r+1}]$ by the linear approximation,

$$\hat{f}_h(x_h) = \lambda_{hr} f_h(a_{hr}) + \lambda_{h, r+1} f_h(a_{h, r+1}) \quad \ldots \quad (3.4.2)$$

where $x_h$ is given by (3.4.1). It is easily seen that, for each $h \notin H$ and any $x_h \in [a_h, b_h]$, the entire piecewise linear approximation $\hat{f}_h(x_h)$ with breaks at the grid points can be written as,
\[ f_h(x_h) = \sum_{h \in H} \lambda_{hr} f_h(a_{hr}) \]

where

\[ x_h = \sum_{r=1}^{n_h} \lambda_{hr} x_{hr} \]

\[ \sum_{r=1}^{n_h} \lambda_{hr} = 1 \text{ and } \lambda_{hr} \geq 0 \text{ for } r = 1, 2, \ldots, n_h. \]

provided, for each h, at most two adjacent \( \lambda_{hr} \)'s are positive. This condition ensures that the linear approximation occurs only between adjacent grid points.

Replacing the nonlinear function \( f_h(x_h), h \in H \) by their piecewise linear approximation \( \hat{f}_h(x_h) \) given by (3.4.2) in the separable programming problem given by (3.3.9)-(3.3.10), we obtain the following approximated program in the variables \( \lambda_{hr} (h \in H) \) and \( x_h (h \in H) \) as,

Minimize \[ Z = \sum_{h \in H} f_h(x_h) + \sum_{h \in H} \sum_{r=1}^{n_h} \lambda_{hr} f_h(a_{hr}) \ldots. (3.4.3) \]
Subject to,
\[
\sum_{h \in H} g_h(x_h) + \sum_{h \in H} \sum_{r=1}^{n_h} \lambda_{hr} g_h(a_{hr}) \leq v \quad \ldots \ldots \quad (3.4.4)
\]
\[
\sum_{r=1}^{n_h} \lambda_{hr} = 1 \quad h \notin H \quad \ldots \ldots \ldots \ldots \quad (3.4.5)
\]
\[
\lambda_{hr} \geq 0 \quad (\text{for } r = 1, 2, \ldots, n_h; h \notin H) \quad (3.4.6)
\]
and the additional restriction, namely, for each \( h \notin H \), no more than two adjacent \( \lambda_{hr} \) can be positive, that is if \( \lambda_{hr} \) is positive then only \( \lambda_{h,r-1} \) or \( \lambda_{h,r-1} \) can be positive.

The approximated programming problem (3.4.3)-(3.4.6) is a linear programming problem. Therefore the problem (3.4.3)-(3.4.6) can be solved by the simplex method by using the following restricted basis entry rule. A non basic variable \( \lambda_{hr} \) is introduced into the basis only if it improves the value of the objective function and if, for each \( h \notin H \), the new basis has no more than two adjacent \( \lambda_{hr} \)'s that are positive. The optimal values \( x_h^* \) (\( h \notin H \)) and \( \lambda_{hr}^0 \) (\( r = 1, 2, \ldots, n_h; h \notin H \)) obtained by solving (3.4.3)-(3.4.6) yield an approximate optimal solution \( \hat{x} \) to the problem (3.3.9)-(3.3.11), where the components \( \hat{x}_h \) of \( \hat{x} \) are given by,
\[ R = \begin{cases} x_h^0 & \text{if } h \in H \\ \sum_{r=1}^{n} \lambda_h^r a_{hr} & \text{if } h \notin H \end{cases} \]

It may be remarked here that the foregoing procedure of solving approximated separable programming guarantees an approximate local maximum or minimum. It is only when the functions \( f_h(x_h) \) and \( g_h(x_h) \) have the appropriate convexity or concavity properties which assures us that a local optimum is also a global optimum that we can find a global optimum for the approximating problem, and hence an approximated optimal solution. The use of finer grid points improves the approximate solution but at the expense of more computational efforts. For more accurate solutions, it is usually advantageous to initially solve a problem with a rather coarse grid points and then re-solve it by using a finer grids only in the neighborhood of the initial approximate solution.

3.5 A NUMERICAL EXAMPLE:

In a survey the population is stratified into two strata (\( L=2 \)). The following information are available,
The separable programming problem (3.3.9)-(3.3.11) thus become

Minimize \( 4x_1^2 + 9x_2^2 + 5x_1 + 10x_2 \) \( \ldots \ldots \) (3.5.1)

Subject to,

\[
\frac{16}{x_1^2} + \frac{144}{x_2^2} \leq 1 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.5.2)
\]

\[
\begin{cases}
\sqrt{2} \leq x_1 \leq \sqrt{200} \\
\sqrt{2} \leq x_2 \leq \sqrt{300}
\end{cases} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.5.3)
\]

It is clear from the constraint (3.5.2) that \( x_1 > 4 \) and \( x_2 > 12 \).

Let us take \( x_1 \geq 4.1 \) and \( x_2 \geq 12.1 \) which implies that the ranges of the variables are,

\[
4.1 \leq x_1 \leq 14.1, \text{ and } \\
12.1 \leq x_2 \leq 17.3
\]
Let us use the grid points for \( x \), as 4.1, 6.1, 8.1, 10.1, 12.1, 14.1, and that for \( x^2 \) as 12.1, 14.1, 16.1, and 17.3. Also \( a_{11} = 4.1, a_{12} = 6.1, a_{13} = 8.1, a_{14} = 10.1, a_{15} = 12.1, \) and \( a_{16} = 14.1. \) Similarly, \( a_{21} = 12.1, a_{22} = 14.1, a_{23} = 16.1, \) and \( a_{24} = 17.3. \)

The piecewise linear approximation to the functions

\[
\begin{align*}
f_1(x) &= 4x^2 + 5x \\
f_2(x) &= 9x^2 + 10x \\
g_{11}(x) &= \frac{16}{x^2} \\
g_{12}(x) &= \frac{144}{x^2} \quad \text{are} \\
\hat{f}_1(x) &= 87.74\lambda_{11} + 179.34\lambda_{12} + 302.94\lambda_{13} + 45854\lambda_{14} + 646.14\lambda_{15} + 865.74\lambda_{16} \\
f_2(x) &= 1438.69\lambda_{21} + 1930.29\lambda_{22} + 2493.89\lambda_{23} + 2866.61\lambda_{24} \\
\hat{g}_{11}(x) &= 0.95\lambda_{11} + 0.43\lambda_{12} + 0.24\lambda_{13} + 0.16\lambda_{14} + 0.11\lambda_{15} + 0.08\lambda_{16} \\
\hat{g}_{12}(x) &= 0.98\lambda_{21} + 0.72\lambda_{22} + 0.56\lambda_{23} + 0.48\lambda_{24}
\end{align*}
\]

Thus the approximated separable program is,

Minimize \( \hat{f}_1(x) + \hat{f}_2(x) \) \( \ldots \) \( (3.5.4) \)

Subject to,

\[
\begin{align*}
\hat{g}_{11}(x) + \hat{g}_{12}(x) &\leq 1 \quad \ldots \quad (3.5.5) \\
\lambda_{11} + \lambda_{12} + \lambda_{13} + \lambda_{14} + \lambda_{15} + \lambda_{16} &= 1 \quad \ldots \quad (3.5.6)
\end{align*}
\]
\[
\lambda_{21} + \lambda_{22} + \lambda_{23} + \lambda_{24} = 1 \quad \ldots \ldots \ldots \quad (3.5.7)
\]
\[
\begin{align*}
\lambda_{12} \geq 0 \\
\lambda_{22} \geq 0
\end{align*} \quad \ldots \ldots \ldots \ldots \quad (3.5.8)
\]

Introducing the slack variable \( x_3 \geq 0 \) in the constraint (3.5.5) and artificial variables \( x_4 \geq 0, x_5 \geq 0 \) in the constraint (3.5.6) and (3.5.7). The objective function (3.5.4) is minimized by Charne's M-technique following the restricted basis entry rule. The following are the simplex tableaus.

**TABLE 3.5.1**

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>( C_0 )</th>
<th>( X_0 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( \lambda_3 )</th>
<th>( \lambda_4 )</th>
<th>( \lambda_5 )</th>
<th>( \lambda_6 )</th>
<th>( \lambda_7 )</th>
<th>( \lambda_8 )</th>
<th>( \lambda_9 )</th>
<th>( x_4 )</th>
<th>( x_5 )</th>
<th>( x_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>0</td>
<td>1</td>
<td>.9</td>
<td>.43</td>
<td>.24</td>
<td>.16</td>
<td>.11</td>
<td>.08</td>
<td>.96</td>
<td>.72</td>
<td>.56</td>
<td>.48</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( x_5 )</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>M-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>87</td>
<td>179</td>
<td>30</td>
<td>45</td>
<td>64</td>
<td>86</td>
<td>14</td>
<td>19</td>
<td>249</td>
<td>286</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.3</td>
<td>2.</td>
<td>8.</td>
<td>6.</td>
<td>5.</td>
<td>38</td>
<td>30</td>
<td>3.9</td>
<td>6.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 3.5.2

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>$C_b$</th>
<th>$x_a$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_3$</th>
<th>$\lambda_4$</th>
<th>$\lambda_5$</th>
<th>$\lambda_6$</th>
<th>$\lambda_7$</th>
<th>$\lambda_8$</th>
<th>$\lambda_9$</th>
<th>$\lambda_{10}$</th>
<th>$\lambda_{11}$</th>
<th>$\lambda_{12}$</th>
<th>$\lambda_{13}$</th>
<th>$\lambda_{14}$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0</td>
<td>.0</td>
<td>.52</td>
<td>.71</td>
<td>.79</td>
<td>.84</td>
<td>.87</td>
<td>.98</td>
<td>.72</td>
<td>.56</td>
<td>.43</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>87.74</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3$</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 3.5.3

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>$C_b$</th>
<th>$x_a$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_3$</th>
<th>$\lambda_4$</th>
<th>$\lambda_5$</th>
<th>$\lambda_6$</th>
<th>$\lambda_7$</th>
<th>$\lambda_8$</th>
<th>$\lambda_9$</th>
<th>$\lambda_{10}$</th>
<th>$\lambda_{11}$</th>
<th>$\lambda_{12}$</th>
<th>$\lambda_{13}$</th>
<th>$\lambda_{14}$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>143.869</td>
<td>.057</td>
<td>0</td>
<td>-.53</td>
<td>-.72</td>
<td>.81</td>
<td>-.84</td>
<td>-.89</td>
<td>1</td>
<td>.73</td>
<td>.57</td>
<td>.49</td>
<td>1.0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>87.74</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3$</td>
<td>M</td>
<td>.95</td>
<td>0</td>
<td>.53</td>
<td>.72</td>
<td>.81</td>
<td>.86</td>
<td>.89</td>
<td>0</td>
<td>.27</td>
<td>.43</td>
<td>.51</td>
<td>-1.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

61
In Table 3.5.4, according to the usual simplex criteria, \( \lambda_4 \) should enter the next basis since the corresponding \( Z_j - C_j \) has the most positive value (i.e., \( -51M - 5161 \)) and \( x_5 \) should leave the basis since the ratio \( (b_3/a_{35}) = 0.06/0.51 = 0.11 \) is the smallest one. But according to the restricted basis entry rule the two nonzero \( \lambda_r \) must have adjacent subscript \( r \). Hence \( \lambda_4 \) can not enter the basis as \( \lambda_4 \) remains in the basis. The next most positive cost coefficient corresponds to \( \lambda_1 \), but it also can not enter since \( \lambda_1 \) remains in the basis. Finally \( \lambda_2 \) enters the basis and \( x_1 \) leaves the basis. The remaining simplex tableaus are

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>( C_j )</th>
<th>( x_0 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( \lambda_3 )</th>
<th>( \lambda_4 )</th>
<th>( \lambda_5 )</th>
<th>( \lambda_6 )</th>
<th>( \lambda_7 )</th>
<th>( \lambda_8 )</th>
<th>( x_1 )</th>
<th>( x_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
<td>143.86</td>
<td>0.94</td>
<td>0.8</td>
<td>0.36</td>
<td>0.17</td>
<td>0.08</td>
<td>0.03</td>
<td>0</td>
<td>1</td>
<td>0.73</td>
<td>0.57</td>
<td>0.49</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>865.74</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>M</td>
<td>0.06</td>
<td>-0.89</td>
<td>-0.36</td>
<td>-0.17</td>
<td>-0.08</td>
<td>-0.03</td>
<td>0</td>
<td>0</td>
<td>0.27</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>0</td>
<td>0</td>
<td>0.43</td>
<td>-80</td>
<td>5161</td>
</tr>
</tbody>
</table>
Thus \[ \lambda_1 = \sum_{r=1}^{k} \lambda_{1r} a_{1r} = 1 \times 10.1 = 10.1 \]

and \[ \lambda_2 = \sum_{r=1}^{k} \lambda_{2r} a_{2r} = 0.48 \times 12.1 + 0.52 \times 14.1 = 13.4 \]

Therefore the required sample size for the two strata are \( n_1 = 102.01 \approx 102 \)
and \( n_2 = 172.6 \approx 173 \).

The minimum cost needed for the survey is \( C(n_h) = 2147 \).

For better approximated result we may take finer grid points in the neighbourhood of the current optimum. Let the new grid points for \( x_1 \) be 8.1, 9.1, 10.1, 11.1 and 12.1 and that for \( x_2 \) be 12.1, 13.1 and 14.1. Approximating the nonlinear functions again in the above grid points and applying the method mentioned above we obtain the following result,

\[ \hat{x}_1 = 9.1 \text{ and } \hat{x}_2 = 13.35 \]

Thus the required optimum sample size for the two strata are \( n_1 = 82.8 \approx 83 \) and \( n_2 = 178.22 \approx 178 \) with \( C(n_h) = 2114 \).

The number of variables increases rapidly if the number of strata increases. But this problem can be handled easily on a high speed digital computer. Since the restricted entry conditions can be coded for a computer, it is possible to modify
TABLE 3.5.5

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>C_i</th>
<th>X_i</th>
<th>λ_1</th>
<th>λ_2</th>
<th>λ_3</th>
<th>λ_4</th>
<th>λ_5</th>
<th>λ_6</th>
<th>λ_7</th>
<th>λ_8</th>
<th>λ_9</th>
<th>x_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ_1</td>
<td>143.869</td>
<td>0.78</td>
<td>3.3</td>
<td>1.33</td>
<td>0.63</td>
<td>0.3</td>
<td>0.11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-0.59</td>
<td>-0.89</td>
</tr>
<tr>
<td>λ_2</td>
<td>0.29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>193</td>
<td>-3.3</td>
<td>-1.33</td>
<td>-0.63</td>
<td>-0.3</td>
<td>-0.11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.59</td>
<td>1.89</td>
<td>-3.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32.5</td>
<td>25.3</td>
<td>26.0</td>
<td>166</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
</tr>
</tbody>
</table>

TABLE 3.5.6

<table>
<thead>
<tr>
<th>Basic variable</th>
<th>C_i</th>
<th>X_i</th>
<th>λ_1</th>
<th>λ_2</th>
<th>λ_3</th>
<th>λ_4</th>
<th>λ_5</th>
<th>λ_6</th>
<th>λ_7</th>
<th>λ_8</th>
<th>λ_9</th>
<th>x_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ_1</td>
<td>143.869</td>
<td>0.48</td>
<td>3</td>
<td>1.03</td>
<td>0.33</td>
<td>0</td>
<td>-0.19</td>
<td>-0.3</td>
<td>1</td>
<td>0</td>
<td>-0.59</td>
<td>-0.89</td>
</tr>
<tr>
<td>λ_2</td>
<td>0.29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>193</td>
<td>-3</td>
<td>-1.03</td>
<td>-0.33</td>
<td>0</td>
<td>-0.19</td>
<td>-0.3</td>
<td>0</td>
<td>1</td>
<td>1.59</td>
<td>1.89</td>
<td>-3.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
</tr>
</tbody>
</table>

Since all the \( Z_i \) - \( C_i \) in the table (3.5.6) are negative, the optimum solution is obtained. The optimum solution is

\[
\lambda_{11} = 0, \lambda_{12} = 0, \lambda_{13} = 0, \lambda_{14} = 1, \lambda_{15} = 0, \lambda_{16} = 0
\]

and

\[
\lambda_{21} = 0.48, \lambda_{22} = 0.52, \lambda_{23} = 0, \lambda_{24} = 0 \text{ and } x_1 = 0
\]
a simplex code to solve the problem (3.4.3)-(3.4.6) with restricted basis entry rule.

3.6 THE MULTIVARIATE CASE:

The separable programming technique discussed in section 3.4 can also be applied when there are multiple characters under study.

Let the cost function in multivariate stratified sample for \( p \) characters be defined as

\[
C(n_h) = \sum_{h=1}^{L} \sum_{j=1}^{p} c_{hj} n_h + \sum_{h=1}^{L} \sum_{j=1}^{p} t_{hj} \sqrt{n_h} \quad \ldots \ldots \quad (3.6.1)
\]

where \( c_{hj} \) be the cost of measuring the \( j \)th character on a unit of the \( h \)th stratum, and \( t_{hj} \) be the travelling cost.

Let \( V_j \) be the variance of the \( j \)th characteristic which is given by

\[
V(\overline{y}_{jst}) = \sum_{h=1}^{L} \frac{w_h^2}{n_h} S_{hj}^2, \quad j=1,2,\ldots,p \quad \ldots \ldots \quad (3.6.2)
\]

where \( S_{hj} \) denote the stratum variance in the \( h \)th stratum with respect to the \( j \)th character.
The problem of optimum allocation of sample sizes to different strata can be stated as,

Minimize \[ C(n_h) = \sum_{h=1}^{L} \sum_{j=1}^{P} c_{hj} n_h + \sum_{h=1}^{L} \sum_{j=1}^{P} t_{hj} \sqrt{n_h} \] \hspace{1cm} (3.6.3)

Subject to,

\[ \sum_{h=1}^{L} \frac{w_h^2 S_{hj}^2}{n_h^2} \leq v_j, \quad j=1,2,\ldots,p \] \hspace{1cm} (3.6.4)

\[ 2 \leq n_h \leq N_h \] \hspace{1cm} (3.6.5)

With the substitution \( X_h = \sqrt{n_h} \) the nonlinear programming problem becomes

Minimize \[ C = \sum_{h=1}^{L} \sum_{j=1}^{P} c_{hj} X_h^2 + \sum_{h=1}^{L} \sum_{j=1}^{P} t_{hj} X_h \] \hspace{1cm} (3.6.6)

Subject to,

\[ \sum_{h=1}^{L} \frac{w_h^2 S_{hj}^2}{X_h^2} \leq v_j, \quad j=1,2,\ldots,p \] \hspace{1cm} (3.6.6)

\[ \sqrt{2} \leq X_h \leq \sqrt{N_h} \] \hspace{1cm} (3.6.7)
Both the objective function and the constraints are separable function, hence we apply the separable programming technique. The nonlinear objective and constraint functions are linearized by the method discussed in section 3.4. The approximated linear programming problem is thus,

Minimize  \[ Z = \sum_{n \in H} f_{hj} (x_n) + \sum_n \sum_j \lambda_{hrj} f_{hj} (a_{hrj}) \]

Subject to,

\[ \sum_{n \in H} g_{hj} (x_n) + \sum_n \sum_j \lambda_{hrj} g_{hrj} (a_{hrj}) \leq v_j \]
\[ \sum \lambda_{hrj} = 1 \]
\[ \lambda_{hrj} \geq 0 \]
\[ x_n \geq 0 \]

and the additional restrictions on \( \lambda_{hrj} \) as described elsewhere in this chapter.