Chapter 6

NONRESPONSE IN STRATIFIED SAMPLE SURVEYS: A MATHEMATICAL PROGRAMMING APPROACH
6.1 INTRODUCTION

In complex sample surveys where the evaluation of sampled units are practically difficult the surveyor may fail to measure some units selected in the sample for one reason or the other. In such cases the sampler gets an incomplete sample. In the sampling literature, failure to measure some of the units selected in the sample is termed as ‘nonresponse’. Cochran (1977) gave the following reasons for nonresponse.

(i) **Noncoverage**- due to the failure to locate or visit some units in the sample.

(ii) **Not-at-home**- due to the absence of the sampling unit from the given address.

(iii) **Unable to answer**- the respondent may not have the required information or may be unwilling to share it.

(iv) **The ‘hard core’**- persons who adamantly refuse to be interviewed.

In a questionnaire survey if a question is highly sensitive or personal the person may refuse to answer or may give evasive answer. This situation is covered in the “hard core”, that is, the reason (iv) stated above. To get response on such question the interviewer must encourage the truthful answers without revealing the identity of the person interviewed. Let $\pi_A$ denote the proportion of respondents belonging to a certain class ‘A’. By using a randomizing device Warner (1965) showed that $\pi_A$ can be estimated. Warner’s method is popularly known as “Randomized Response (RR) Method” in sampling literature. Many authors worked on RR methods. Some of them are Greenberg *et al.* (1969), Moors (1971), Mangat and Singh (1990), Mangat (1994), Singh *et al.* (2000). Kim and Warde (2004) used the Warner’s model in stratified sample surveys with nonresponse and obtained allocations under loglinear and nonlinear survey costs.
In this chapter the problem of optimum allocation in stratified random sampling when there are "hard cores" nonrespondents in the population is discussed. The problem is formulated as an All Integer Nonlinear Programming Problem (AINLPP). A numerical example is also presented and its solution is obtained by using optimization software LINGO (2001).


6.2 FORMULATION OF THE PROBLEM

In Warner's RR method either of the following two mutually exclusive and exhaustive questions is asked

(i) Are you a member of class A?
(ii) Are you not a member of class A?

Where A is certain attribute on the basis of which the population is to be classified. Each question requires a 'YES' or 'NO' response. By any randomizing device one of the questions is selected. The interviewer does not know which of the two questions is selected but does know the relative probabilities with which the two questions are selected. Let question (i) be selected with probability \( P \). Obviously the probability of selection of question (ii) will be \( 1-P \). Let in a random sample of size \( n \) the number of 'YES' answers be recorded as \( m \). Then the binomial estimate of the proportion \( \phi \) of the 'YES' answer is given by \( \hat{\phi} = \frac{m}{n} \). Assuming correct answers from the respondents Cochran (1977) gave the relation between \( \phi \) and \( \pi_A \) in the population as:

\[
\phi = (2P - 1) \pi_A + (1 - P) \quad ; \quad P \neq 0.5
\]  

(6.1)

The estimated value of \( \pi_A \) is

\[
\hat{\pi}_A = \left[ \frac{\hat{\phi} - (1 - P)}{(2P - 1)} \right] \quad ; \quad P \neq 0.5
\]

(6.2)

It can be seen that \( \hat{\pi}_A \) is the maximum likelihood estimate (MLE) of \( \pi_A \) with a
sampling variance

\[ V(\hat{\pi}_A) = \frac{\hat{\pi}_A(1-\hat{\pi}_A)}{n} + \frac{P(1-P)}{n(2P-1)^2}; P \neq 0.5 \]  \hspace{1cm} (6.3)

In a stratified population with \( L \) strata of sizes \( N_h; h=1,2,\ldots,L \) let a simple random sample of size \( n_h \) be obtained from the \( h^{th} \) stratum. Under the RR model for \( h^{th} \) stratum define:

- \( \pi_{Ah} \) : proportion of respondents who belong to certain class A
- \( \phi_h \) : proportion of 'YES' answers
- \( \hat{\phi}_h \) : binomial estimate of \( \phi_h \)
- \( P_{Ah} \) : probability of selection of question (i)

With the above definitions for, \( h^{th} \) stratum

\[ \phi_h = (2P_{Ah} - 1)\pi_{Ah} + (1 - P_{Ah}); P_{Ah} \neq 0.5, \]  \hspace{1cm} (6.4)

and

\[ V(\hat{\pi}_{Ah}) = \frac{\pi_{Ah}(1-\pi_{Ah})}{n_h} + \frac{P_{Ah}(1-P_{Ah})}{n_h(2P_{Ah}-1)^2}; P_{Ah} \neq 0.5 \]  \hspace{1cm} (6.5)

where \( \hat{\pi}_{Ah} \) is the MLE of \( \pi_{Ah} \).

If \( W_h = \frac{N_h}{N} \) denote the proportion of population units falling in the \( h^{th} \) stratum then

an unbiased estimate of \( \pi_A \) is given by

\[ \hat{\pi}_A = \sum_{h=1}^{L} W_h \hat{\pi}_{Ah} \]  \hspace{1cm} (6.6)

with a sampling variance

\[ V(\hat{\pi}_A) = \sum_{h=1}^{L} W_h^2 V(\hat{\pi}_{Ah}) \]

\[ = \sum_{h=1}^{L} W_h^2 \left[ \frac{\pi_{Ah}(1-\pi_{Ah})}{n_h} + \frac{P_{Ah}(1-P_{Ah})}{n_h(2P_{Ah}-1)^2} \right]; P_{Ah} \neq 0.5 \]  \hspace{1cm} (6.7)
where \( N = \sum_{h=1}^{L} N_h \) and \( N_h; h = 1, 2, ..., L \) is the strata sizes.

The interviewer have to approach the population units selected in the sample from each strata to get the answer of the two proposed questions under the RR model. Within each stratum the interviewer have to travel between selected units to contact them. This involves travelling cost, in addition to the usual overhead and measurement costs. Beardwood \textit{et al.} (1959) showed that the travelling cost between \( n \) randomly scattered points may be given as \( t\sqrt{n} \) where \( t \) is a constant. So that if the travel cost is substantial then instead of the usual linear cost function \( C = c_0 + \sum_{h=1}^{L} c_h n_h \), it would be advisable to use

\[
C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h}
\]  

(6.8)

as the cost function. Where \( c_0 = \) overhead cost, \( c_h = \) per unit cost of measurement in \( h^{th} \) stratum; \( h = 1, 2, ..., L \) and \( \sum_{h=1}^{L} t_h \sqrt{n_h} \) is the total cost involved in travelling within strata between units selected in the sample.

Under Warner's RR model, the problem of allocation for fixed total cost may be expressed as the following AINLPP.

\[
\text{Minimize} \quad V(\hat{x}_A) \\
\text{subject to} \quad \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
2 \leq n_h \leq N_h \\
n_h \text{ integers; } h = 1, 2, ..., L
\]

(6.9)

where \( C_0 = C - c_0 \) is the fixed budget.

\( t_h \sqrt{n_h} \) represent the travel cost within \( h^{th} \) stratum and \( V(\hat{x}_A) \) is given by (6.7).

The restrictions \( 2 \leq n_h \) and \( n_h \geq N_h \) in AINLPP (6.9) are placed to have estimates of
strata mean squares $S_h^2$ and to avoid oversampling, respectively. For practical implementation of the allocations the sampler needs their integer values therefore integer restrictions on $n_h$ are imposed.

Substituting $A_h = \left[ \pi_{th} (1 - \pi_{th}) + \frac{P_{th} (1 - P_{th})}{(2P_{th} - 1)^2} \right]_h = 1, 2, \ldots, L$

expression (6.7) for $V(\hat{\pi}_a)$ may be simplified as

$$V(\hat{\pi}_a) = \sum_{h=1}^{L} \frac{A_h W_h^2}{n_h}, \text{ or } V(\hat{\pi}_a) = \sum_{h=1}^{L} \frac{K_h}{n_h},$$

where $K_h = A_h W_h^2; h = 1, 2, \ldots, L$.

AINLPP (6.9) can now be stated in a more simpler form as:

Minimize $V(\hat{\pi}_a) = \sum_{h=1}^{L} \frac{K_h}{n_h}$

subject to $\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} f_h \sqrt{n_h} \leq C_0$ (6.10)

$$2 \leq n_h \leq N_h$$

$n_h$ integers, $h = 1, 2, \ldots, L$.

Using Lagrange Multipliers Technique (LMT) the AINLPP may be solved by taking equality in the constraint and ignoring the restrictions $2 \leq n_h \leq N_h$ and $n_h$ integers; $h = 1, 2, \ldots, L$. The noninteger solution may be rounded off to get integer allocations. If the rounded off values of $n_h$ satisfy the restrictions $2 \leq n_h \leq N_h$; $h = 1, 2, \ldots, L$ the AINLPP (6.10) is solved otherwise some integer nonlinear programming technique is to be used. For reasons given in Khan et al. (1997) 'rounding off' of the noninteger sample sizes is not always advisable because they may lead to infeasible or nonoptimum (or both) results. When the parameters $K_h, c_h, f_h, C_0$ and $N_h$ of the AINLPP (6.10) are known it can be solved by using an optimization software. In this chapter LINGO (2001) is used.

In the following section a numerical example is presented to illustrate the formulation of the problem for a given data. The solution to the formulated problem is obtained by
optimization software LINGO.

6.3 A NUMERICAL EXAMPLE

Parts of the following data for a stratified population with two strata are from Shabbir and Gupta (2005).

<table>
<thead>
<tr>
<th>Stratum</th>
<th>$N_h$</th>
<th>$W_h$</th>
<th>$\pi_{Ah}$</th>
<th>$P_{Ah}$</th>
<th>$c_h$</th>
<th>$t_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>700</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>19</td>
<td>15</td>
</tr>
</tbody>
</table>

It is assumed that the available budget $C=5500$ units including an overhead cost $c_0=500$ units. So that $C_0=5500-500=5000$ units. The strata sizes are assumed to be 300 and 700 respectively as given in Table 6.1. The computed values of $K_h, h = 1, 2$ are presented in Table 6.2.

<table>
<thead>
<tr>
<th>Stratum No.</th>
<th>$A_h$</th>
<th>$K_h = A_h W_h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.24</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>1.55</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Substituting the values of the parameters from Table 6.1 and 6.2, the AINLPP (6.10) becomes
Nonresponse in stratified surveys ...

Minimize \( F(n_h) = \frac{0.56}{n_1} + \frac{0.76}{n_2} \)
subject to \( 14n_1 + 19n_2 + 10\sqrt{n_1} + 15\sqrt{n_2} \leq 5000 \) \( (6.11) \)
\( 2 \leq n_1 \leq 300 \) and \( 2 \leq n_2 \leq 700; n_1, n_2 \) integers

Using optimization software LINGO we obtain the required optimum allocation as:
\( n_1 = 143 \) and \( n_2 = 142. \)
The total cost under this allocation is 4998.33 < 5000 units. The optimal value of the objective function \( V^*(\hat{\pi}_d) = 0.00926819. \)

6.4 CONCLUSION
This chapter gives a simple formulation for the problem of working out the optimum allocation in stratified sampling when the ‘hard core’ nonrespondents are there. No prior assumption has been made about any particular type of allocation. The cost function is nonlinear that includes the travelling cost also.