# A MULTIPHASE SAMPLING PROCEDURE WHEN NONRESPONSE ARISES IN A MAIL SURVEY

BY

ROSAMMA MATHEW

*Ahmadū Bello University, Zaria, Nigeria*

(Received : January, 1982)

## SUMMARY

**A simple and applicable multiphase sampling procedure to solve the problem of non-response in mail surveys is proposed which is an extension of the call back method due to Hansen – Hurwitz. The unbiased estimator developed for the population parameter (mean or total) and the unbiased estimator of its variance are also simple. The empirical illustrations at the end of the paper, indicates that the proposed procedure is more efficient than the basic procedure due to Hansen—Hurwitz.**

## INTRODUCTION

Numerous procedures have been developed to reduce the effect of bias in the collected data due to non-response. In this paper a simple weighting procedure is proposed on the same lines as the one due to El-Badry [2]. The proposed method can also be treated as an extension of the basic method due to Hanson and Hurwitz [3].

Assume that the population under study is finite with $N$ distinguishable units. The population is further assumed to be conceptually partitioned into $(L+1)$ mutually exclusive and exhaustive strata. The $N_i$ (unknown) units belonging to the $i^{th}$ stratum consists of those who respond to the $(i-1)^{th}$ reminder. The $(L+1)^{th}$ stratum is comprised of the $N_{L_2}$ units which do not respond the $L$ reminders.

## 2. The Multiphase Sampling Procedure

A simple random sample of size $n$ is selected and questionnaires are sent by mail. Let $n_1$ of them respond and $n_{12}$ do not. Instead of taking a sub-sample of the non-respondents as proposed by El-Badry [2], we send reminders to all the non-respondents. Let $n_2$ of $n_{12}$ respond and $n_{22}$ do not. A second reminder is sent to all the $n_{22}$ and $n_3$ respond. This continues till we reach a stage when further reminders wouldn't help much. Let this be the $L^{th}$ stage. Now we have $n_{L_2}$ non-respondents, where

$$n_{L_2}=n-n_1-n_2-\ldots-n_L\,.$$

At this stage we use personal interview method. Since personal interview is much more expensive than mail survey we use subsampling at this stage. The last attempt is made on $n_{L+1}=\frac{n_{L_2}}{k}$ where $k \not> 1$. It is assumed that there is no non-response at this attempt. The estimater for the population mean is

$$\bar{x}_{mp}=\sum_{h=1}^{L}\frac{n_h\bar{x}_h}{n}+\frac{n_{L2}\,\bar{x}_{L+1}}{n} \qquad \ldots(1)$$

$$=\sum_{h=1}^{L}\omega_h\bar{x}_h+\frac{n_{L2}\,\bar{x}_{L+1}}{n}$$

where $\bar{x}_h=\sum_{j=1}^{n_h}\frac{x_{hj}}{n_h}$, $h=1, 2, \ldots(L+1)$.

It can easily be proved that $\bar{x}_{mp}$ is unbiased by theorem 12.1 in Cochran [1].

Similarly, the following expression for variance of the estimator $Var(\bar{x}_{mp})$ is worked out directly with the help of theorem 12.2 in Cochran [1].

$$Var(\bar{x}_{mp})=S^2\left(\frac{1}{n}-\frac{1}{N}\right)+\frac{W_{L+1}\,S^2{}_{L+1}}{n}(k-1) \qquad \ldots(2)$$

where $S^2$ is the population variance,

$W_{L+1}=\frac{N_{L_2}}{N}$ and $S^2_{L+1}$ is the variance of the $(L+1)^{th}$ stratum.

## 3. Estimation of Variance

Using the analysis of variance technique, $Var(\bar{x}_{mp})$ can be written as

$$Var(\bar{x}_{mp})=\sum_{1}^{L} W_h S_h^2\left(\frac{1}{n}-\frac{1}{N}\right)+W_{L+1}\, S_{L+1}\left(\frac{k}{n}-\frac{1}{N}\right)$$

$$+\frac{g}{nN}\sum_{1}^{L+1}(W_h-1)S_h^2+\frac{g}{n}\sum_{1}^{L+1}W_h(\bar{X}_h-\bar{X})^2$$

where, $g=\frac{N-n}{N-1}$ and $S_h^2$ is the variance of the $h^{th}$ stratum.

In mail survey, $\frac{n}{N}$ may not be negligible. However, $\frac{g}{nN}$ can be neglected in most applications. So $Var(\bar{x}_{mp})$ simplifies to

$$Var(\bar{x}_{mp})=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{1}^{L+1} W_h\, S_h^2+\frac{W_{L+1}\, S_{L+1}^2}{n}(k-1)$$

$$+\frac{g}{n}\sum_{1}^{L+1}W_h(\bar{X}_h-\bar{X})^2$$

If $\frac{1}{n}$ and $\frac{1}{N}$ are both negligible with respect to one, an unbiased sample estimator of $Var(\bar{x}_{mp})$ is

$$v(\bar{x}_{mp})=\frac{N-1}{N}\left[\sum_{1}^{L}\left\{\frac{n_h-1}{n_h}\left[\frac{1}{n_h-1}-\frac{1}{N-1}\right]\omega_h s_h^2\right\}\right.$$

$$\left.+\left\{\frac{n_{L+1}-1}{n-1}-\frac{n_{L+1}-k}{k(N-1)}\right\}\frac{\omega_{L+1}\, s_{L+1}^2\, k}{n_{L+1}}\right]$$

$$+\frac{N-n}{N(n-1)}\sum_{1}^{L+1}\omega_h(\bar{x}_h-\bar{x}_{mp})^2.$$

## 4. Optimum Allocation

The cost function taken is

$$\subset=c_1\{n+(n-n_1)+(n-n_1-n_2)+\ldots+(n-n_1,\ldots;n_{L-1})\}$$

$$+c_2(n_1+n_2+\ldots,+n_L)+\frac{c_3\, n_{L_2}}{k}.$$

where $c_1$ is the cost of mailing a questionnaire, $c_2$ is the cost of processing the results from a questionnaire and $c_3$ is the cost of an interview. Expected cost function is

$$E(C)=C^*=n\left[c_1\{L-\sum_{1}^{L-1}(L-h)W_h\}+c_2(1-W_{L+1})\right.$$
$$\left.+c_3\,\frac{W_{L+1}}{k}\right], \qquad \ldots(3)$$

Minimizing the product $C^*\left(V+\frac{S^2}{N}\right)$, we get

$$k^2_{opt}=\frac{c_3(S^2-W_{L+1}\,S^2_{L+1})}{S^2_{L+1}\,[c_1\{L-\sum_{1}^{L-1}(L-h)W_h\}+c_2(1-W_{L+1})]} \qquad \ldots(4)$$

Note that $k_{opt}$ does not depend on $n$. The optimum $n$ is obtained from the expected cost function for a given total cost. The resulting minimum variance is

$$V_{min}(\bar{x}_{mp})=\frac{1}{C^*}\,[\{S^2_L-W_{L+1}S^2_{L+1}\}+k_{opt}\,W_{L+1}\,S^2_{L+1}]-\frac{S^2}{N}.$$

## 5. EMPIRICAL ILLUSTRATION

A small scale survey was conducted using the proposed method. The population under study was the senior staff of Ahmadu Bello University, Zaria, Nigeria. The population size ($N$) was 1534. The parameter under study was the proportion of parents who would like to send their children to day-schools.

The initial sample size ($n$) was calculated for $(L+1)=3$; $c_1=0.10$; $c_2=0$; $c_3=3.00$; guessed values of $W_1$, $W_2$ and $W_3$ as 0.4, 0.2 and 0.4 respectively; and assuming $S^2=S^2_3$.

$$k^2_{opt}=11.25 \text{ from (4)}$$

$$i.e.\ k_{opt}=3.35.$$

Total fixed cost was 115.00.

$$115=n[.16+.36] \text{ from (3)}$$

$$\therefore \quad n=221.$$

In the two mail attempts, the following values of $n_h$ and $n_{h_2}$ were obtained.

$$\left\{\begin{matrix} n_1=124 \\ n_{12}=97 \end{matrix}\right\} \quad ; \quad \left\{\begin{matrix} n_2=41 \\ n_{22}=56 \end{matrix}\right\}$$

$$p_1=0.384 \qquad s_1^2=\frac{n_1\, p_1\, q_1}{n_1-1}=0.25$$

$$p_2=0.5 \qquad s_2^2=\frac{n_2\, p_2\, q_2}{n_2-1}=0.25$$

where $p_i$ is the parameter under study for the $i^{th}$ stratum.

At the third stage, the value of optimum $k$ was recalculated by using the values of $\omega_1=0.56$; $\omega_2=0.19$ and $\omega_3=0.25$,

$$k^2=\frac{c_3(1-\omega_3)}{c_1(2-\omega_1)}=15.62$$

*i e.* $\quad k=3.96.$

$$n_{L+1}=n_3=\frac{56}{3.96}=14$$

After contacting the 14 people personally we got $p_3=0.57$

Hence the estimate of the population parameter is

$$p_{mp}=\sum_1^3 \omega_h\, p_h=0.45$$

$$v(p_{mp})=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_1^3 \omega_h\, s_h^2+\frac{\omega_3\, s_3^2(k-1)}{n}$$

$$+\frac{N-n}{(N-1)n}\sum_1^3 (p_h-p_{mp})^2$$

$$=0.001+0.0008+0.0002$$

$$=0.00182$$

Slight increase in cost in the last column is due to the low values of $W_2$ and $W_3$. For higher values the cost would be lower. The table shows that the use of the proposed simple method would save us money and effort compared to the Hansen and Hurwitz method. Comparing with El-Badry's method, the cost of this survey may be a little higher, but the computation and guesswork involved is much less.

TABLE

**Comparison of Expected cost of the Proposed Procedure with the Hansen-Hurwitz Procedure**

| $W_1$ | *Expected cost (in dollars)* | | | |
|---|---|---|---|---|
| | *With Hansen and Hurwitz Method* | *with the proposed method* | | |
| | | *for 3 strata $W_2=0.1$* | *for 3 strata $W_2=0.2$* | *for 4 strata $W_2=W_3=0.1$* |
| 0.1 | 4110 | 3723 | 3203 | 3338 |
| 0.2 | 3560 | 3185 | 2694 | 2815 |
| 0.3 | 3034 | 2677 | 2222 | 2324 |
| 0.4 | 2545 | 2205 | 1791 | 1871 |
| 0.5 | 2096 | 1775 | 1404 | 1464 |
| 0.6 | 1690 | 1389 | 1062 | 1103 |
| 0.7 | 1327 | 1049 | 768 | 791 |

## 6. Conclusion

Since the cost of getting a response using reminders is not substantially higher than the cost of getting a response at the first attempt, much gain in efficiency cannot be expected through subsampling the non-respondents. Subsampling is used only at the last stage of this method where personal survey, which is more expensive than mail survey is used.

The variance of this method is given in terms of the response group parameters unlike that in El-Badry. Above all the variance can be estimated using a non-negative unbiased estimator. $k$ can be recalculated at the last stage by using the results obtained from the first $L$ stages. The only assumption required in the calculation of $k$ is $S_L^2=S_{L+1}^2$ and that $S^2$ is the pooled variance of $S_1^2,\ldots,S_L^2$.

If the information required for calculating $n$ is not available at the planning stage of the survey and if any value of $n$ is therefore applied, the procedure which employs $k$ will still be optimum in the sense that it will minimize the product cost $X$ variance. That is, the resulting estimate will have the least variance among all estimates that can be obtained for the expended cost or the least cost among all estimates that have the resulting variance.

This method can also be applied in a telephone survey where the first attempts are by phone and the last attempt by personal interview.

## ACKNOWLEDGEMENT

The author wishes to thank Dr. Kokan, Reader, Department of Mathematics, A.B.U., Zaria, NIGERIA, for his valuable advice, assistance and encouragement. Thanks are also due to the referees for their constructive suggestions.

## REFERENCES

1. Cochran, W.G. (1977) : Sampling Techniques, IIIrd edition, John Wiley & Sons, New York.

2. El-Badry, M.A. (1956) : A sampling procedure for mailed questionnairies. *J. Amer. Statist. Assoc.* **51**, 209-227.

3. Hansen, M.H. and Hurwitz, W.N. (1946) : The problem of non-response in sample surveys *J. Amer. Statist. Assoc.* **41**, 517-529.