INTERNATIONAL JOURNAL OF NATURAL AND APPLIED SCIENCES (IJNAS), VOL. 2, NO. 1(2007); P. 97 – 103, 1 TABLE.

Estimation in two-stage sampling of equal clusters with subsampling the nonrespondents

G. A. Udofia[1]

ABSTRACT

We consider estimation of the mean and variance in two-stage sampling of equal clusters with large sizes. The subgroups of respondents in the different clusters are assumed to be equal. An extension of the sample mean to the nonoresponse group is obtained by combining means of subsamples of nonrespondents in each of the selected clusters. The estimation procedure is illustrated with data from a survey of fishing communities in Cross River State, Nigeria.

INTRODUCTION

Nonresponse or failure to measure some units is a common problem in social surveys. It causes sample bias or failure of a sample to represent a survey population adequately enough to enable reliable inference to be made from the survey data. Since nonresponse is such a frequently occurring phenomenon of social surveys that cannot be avoided, it is necessary to develop method of obtaining reliable estimates of a population parameter in the presence of nonresponse. This paper is a contribution in this direction.

Hansen and Hurwitz (1946) first developed a mathematical model for estimation of population mean with subsampling nonrespondents as a solution to the nonresponse problem. The paper considers a population that is evenly distributed with respect to the variable of interest and accordingly assumes a simple random sample design. Under the same simple random sample design, Birnbaum and Sirken (1950) as well as Cochran, Mosteller and Tukey (1954) extend the model proposed and discussed by Hansen and Hurwitz (1946) to estimation of proportion. Little (1983) presents a general framework for data with nonresponse. Sarndal and Swensson (1985) introduce the use of unequal probabilities of selection in the first phase and subsampling the nonrespondents after post-stratification. Rao (1986) extends the procedure discussed by Hansen and Hurwitz (1946) to ratio estimation. Udofia (2004) extends the procedure developed by Rao (1986) to small area estimation. Udofia (2005) also extends the procedure developed by Hansen and Hurwitz (1946) to estimation in single-stage sampling and with subsampling the nonrespondents.

In some practical sample survey situations, cluster sizes, though equal, may be so large that an investigator cannot include all the units of each selected cluster in his final sample.

This paper therefore discusses the procedure of estimation in two-stage sampling of large equal clusters where nonrespondents are present.

Under Hansen and Hurwitz (1946) estimation procedure, the population under study is supposed to consist of a response stratum of size N_1 and a nonresponse stratum of size $N_2 = N-N_1$. A random sample of size n is drawn without replacement. During the survey, n_1, say, units respond and the remaining $n_2 = n - n_1$ units fail to respond. The n_1 respondents constitute a random subsample from the response stratum and the n_2 nonrespondents constitute a random subsample from the nonresponse stratum. Hansen and Hurwitz (1946) suggest drawing a random subsample of $m = n_2/k$, $k \geq 1$, from the n_2 nonrespondents in the sample and assuming that all the m units respond.

Let Y_i denote the value of the study variable, Y, for element i. Let $\bar{Y}_1 = \left(\sum_{i=1}^{N_1} Y_i \right) \Big/ N_1$, and $S_1^2 = \sum_{i=1}^{N_1} \left(Y_i - \bar{Y}_1 \right)^2 \Big/ \left(N_1 - 1 \right)$ respectively denote the mean and variance for the response stratum. Let $\bar{Y}_2 = \left(\sum_{j=1}^{N_2} Y_j \right) \Big/ N_2$, and $S_2^2 = \sum_{j=1}^{N_2} \left(Y_j - \bar{Y}_2 \right)^2 \Big/ \left(N_2 - 1 \right)$ be simarlarly defined for the nonresponse stratum. The population mean can be written as

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \frac{1}{N} \left(\sum_{i=1}^{N_1} Y_i + \sum_{j=1}^{N_2} Y_j \right) = W_1 \bar{Y}_1 + W_2 \bar{Y}_2$$

where $W_1 = N_1/N$ and $W_2 = N_2/N$.

Let $\bar{y}_1 = \left(\sum_{i=1}^{n_1} y_i \right) \Big/ n_1$ denote the mean for the response subsample and

Manuscript received by the Editor August 10, 2005; revised manuscript accepted October 20, 2006.
[1]Department of Mathematics/Statistics & Computer Science, University of Calabar, Calabar, Nigeria.

$\bar{y}_2 = \left(\sum_{j=1}^{n_2} y_j \right) \Big/ n_2$ denote the mean for the nonresponse subsample. If the study variable is also measured on all the n_2 units the total sample mean is calculated as $\bar{y} = \left(\sum_{i=1}^{n} y_i \right) \Big/ n = w_1 \bar{y}_1 + w_2 \bar{y}_2$ where $w_1 = n_1/n$ and $w_2 = n_2/n$. In the presence of nonresponse, the sample mean is extended to the nonrespose group by calculating the sample mean as $\hat{\bar{Y}}_{HH} = w_1 \bar{y}_1 + w_2 \bar{y}_2$.

This estimator is unbiased for $\bar{Y}$ with variance $V\left(\hat{\bar{Y}}_{HH}\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right) S^2 + W_2 \frac{k-1}{n} S_2^2$ where $S^2 = \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \Big/ (N-1)$. An unbiased estimator of this variance is

$$\hat{V}\left(\hat{\bar{Y}}_{HH}\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\left[\frac{(n_1 - 1)s_1^2 + k(m-1)s_{2m}^2}{n-1}\right] +$$

$$+ \frac{1}{n}\left(1 - \frac{n}{N}\right)\left[\frac{n_1\left(\bar{y}_1 - \hat{\bar{Y}}_{HH}\right)^2 + n_2\left(\bar{y}_{2m} - \hat{\bar{Y}}_{HH}\right)^2}{n-1}\right]$$

$$+ \frac{(N-1)w_2(k-1)s_{2m}^2}{N(n-1)}$$

where $s_1^2 = \sum_{i=1}^{n_1} \left(y_i - \bar{y}_1\right)^2 \Big/ (n-1)$, $s_{2m}^2 = \sum_{s=1}^{m} \left(y_s - \bar{y}_{2m}\right)^2$ and

$$\bar{y}_{2m} = \left(\sum_{s=1}^{m} y_s \right) \Big/ m .$$

Udofia (2005) considers a finite population that consists of N clusters of M units of analysis each. In each of these clusters, M_1 units would respond and the other $M_2 = M - M_1$ units would not respond. Under the sample design, a sample of n clusters is drawn at random and without replacement and all the M units of each selected cluster are included in the sample. Assume that during the survey in each selected cluster, all the M_1 units from the response substratum respond while all the M_2 units from the nonresponse substratum fail to respond. Draw a sample of $m_2 = M_2/k$, $k > 1$, at random and without replacement from the M_2 nonrespondents in each selected cluster and revisit all the m_2 units. Assume that all the m_2 units respond during the revisit.

Let $\bar{Y}_{\cdot 1} = \frac{1}{NM_1} \sum_{i=1}^{N} \sum_{j=1}^{M_1} Y_{ij}$, $\bar{Y}_{\cdot 2} = \frac{1}{NM_2} \sum_{i=1}^{N} \sum_{\ell=1}^{M_2} Y_{i\ell}$, $W_1 = M_1/M$, $W_2 = M_2/M$,

$\bar{y}_1 = \frac{1}{nM_1} \sum_{i=1}^{n} \sum_{j=1}^{M_1} Y_{ij}$, $\bar{y}_{i m_2} = \frac{1}{m_2} \sum_{s=1}^{m_2} y_{is}$, $\bar{y}_{\cdot m_2} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{i m_2}$,

$S_{i w m_2}^2 = \frac{1}{m_2 - 1} \sum_{s=1}^{m_2} \left(y_{is} - \bar{y}_{i m_2}\right)^2$; $s_{y1B}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_{i1}^2 - \frac{y_1^2}{n} \right)$

$s_{yBm_2}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_{i m_2}^2 - \frac{y_{\cdot 2 m_2}^2}{n} \right)$, $y_{i1} = \sum_{j=1}^{M_1} y_{ij}$, $y_1 = \sum_{i=1}^{n} y_{i1}$,

$y_{i m_2} = \sum_{s=1}^{m_2} y_{is}$, $y_{\cdot 2 m_2} = \sum_{i=1}^{n} y_{i m_2}$. An unbiased estimator of the population mean $\bar{Y} = W_1 \bar{Y}_{\cdot 1} + W_2 \bar{Y}_{\cdot 2}$ is $\bar{y}_{cl}^* = W_1 \bar{y}_{\cdot 1} + W_2 \bar{y}_{\cdot m_2}$.

Udofia (2005) gives the variance of this estimator as

$$V\left(\bar{y}_{cl}^*\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\left[W_1^2 S_{1B}^2 + W_2^2 S_{2B}^2\right] +$$

$$\frac{1}{n}\left(1 - \frac{n}{N}\right)\left[\frac{W_1^2 N\left(\bar{Y}_{\cdot 1} - \bar{Y}\right)^2 + W_2^2 N\left(\bar{Y}_{\cdot 2} - \bar{Y}\right)^2}{N-1}\right] +$$

$$+ \frac{W_2(k-1)}{NM_2 - 1}\left[\frac{N(M_2 - 1)S_{2w}^2}{nM} + \frac{W_2(N-1)S_{2B}^2}{n}\right]$$

and its estimator as

$$\hat{V}\left(\bar{y}_{cl}^*\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\frac{1}{M^2}\left(s_{y1B}^2 + k s_{yBm_2}^2\right) + \frac{1}{n}\left(1 - \frac{n}{N}\right)\left[\frac{W_1^2 n\left(\bar{y}_{\cdot 1} - \bar{y}_{cl}^*\right)^2 + W_2^2 n\left(\bar{y}_{\cdot m_2} - \bar{y}_{cl}^*\right)^2}{n-1}\right]$$

$$+ \frac{W_2 k(k-1)}{NM_2 - 1}\left[\frac{N(m_2 - 1)\frac{1}{n}\sum_{i=1}^{n} S_{i w m_2}^2 + (N-1)S_{yBm_2}^2 \Big/ m_2}{N-1}\right] \quad (1)$$

EXTENSION OF HANSEN AND HURWITZ (1946) ESTIMATION PROCEDURE

We consider a finite population of the type discussed by Udofia (2005). Under the sampling strategy discussed in that paper, a random sample of n clusters is drawn without replacement from N clusters and all the M units of each selected cluster are included in the sample. We shall assume that the size M of each selected cluster is so large that, under given time and budget constraints, an investigator cannot include all units of each selected cluster in the sample.

In that case, draw a sample of m units at random and without replacement from each selected cluster. Assume that during the survey in each selected cluster, m_1 of the sample units respond and the other $m_2 = m - m_1$ units fail to respond. Select $m_{22} = m_2/k$, $k > 1$, units from m_2 non-respondents at random and without replacement and revisit all the m_{22} units in the subsample under improved survey conditions (more experienced, more qualified, more respectable field staff as well as better timing of visits, improved publicity and approach) that guarantee complete response. Assume that under such improved conditions all the m_{22} units in each selected cluster respond.

Estimation in two-stage sampling of equal clusters

Notations

Let Y_{ij} denote the value of the study variable, Y, for unit j in cluster i. The population mean (the estimand) is

$$\bar{Y} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij} = \frac{1}{NM}\sum_{i=1}^{N}\left(M_1\bar{Y}_{i1}+M_2\bar{Y}_{i2}\right) = W_1\bar{Y}_{.1} + W_2\bar{Y}_{.2};$$

where $W_1 = M_1/M$; $W_2 = M_2/M$ and $\bar{Y}_i = \frac{1}{M}\sum_{j=1}^{M} Y_{ij}$ is the mean of Y for cluster i.

Then $\bar{Y}_{i1} = \frac{1}{M_1}\sum_{j=1}^{M_1} Y_{ij}$ is the mean of Y for the part of the response stratum in cluster i;

$\bar{Y}_{i2} = \frac{1}{M_2}\sum_{\ell=1}^{M_2} Y_{i\ell}$ is the mean of Y for the part of the non response stratum in cluster i.

$$\bar{Y}_1 = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{i1} = \frac{1}{NM_1}\sum_{i=1}^{N}\sum_{j=1}^{M_1} Y_{ij}$$

$\bar{Y}_2 = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{2i} = \frac{1}{NM_2}\sum_{i=1}^{N}\sum_{\ell=1}^{M_2} Y_{i\ell}$ is the mean of Y for the nonresponse stratum.

$$S_{1B}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i1}-\bar{Y}_{.1}\right)^2;\quad S_{2B}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i2}-\bar{Y}_{.2}\right)^2;$$

$$\bar{y}_{i\,m_1} = \frac{1}{m_1}\sum_{j=1}^{m_1} y_{ij};\quad \bar{y}_{.\,m_1} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i\,m_1};\quad \bar{y}_{i\,m_2} = \frac{1}{m_2}\sum_{\ell=1}^{m_2} y_{i\ell}$$

$$\bar{y}_{.\,m_2} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i\,m_2};\quad \bar{y}_{i\,m_{22}} = \frac{1}{m_{22}}\sum_{s=1}^{m_{22}} y_{is};\quad \bar{y}_{.\,m_{22}} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i\,m_{22}}.$$

The other notations will be defined as the need arises.

Estimation of Mean

The sample mean as an estimator of the population mean is

$$\bar{y}_{cl} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{im}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{i=1}^{m_1} y_{ij} + \sum_{\ell=1}^{m_2} y_{i\ell}}{m}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(w_1\bar{y}_{i\,m_1} + w_2\bar{y}_{i\,m_2}\right)$$

$$= w_1\bar{y}_{.\,m_1} + w_2\bar{y}_{.\,m_2} \quad \ldots \quad (2)$$

where $w_1 = m_1/m$ and $w_2 = m_2/m$. Since $y_{.\,m_2}$ cannot be calculated because of lack of information for all the m_2 units, an unbiased estimator of $\bar{Y}$ is $\bar{y}_{cl\cdot 2} = w_1\bar{y}_{.\,m_1} + w_2\,y_{.\,m_{22}}$

We note that $E_1\left(\bar{y}_{.\,m_1}\right) = \bar{Y}_{.1}$, $E\left(\bar{y}_{.\,m_2}\right) = \bar{Y}_{.2}$, $E_2\left(\bar{y}_{.\,m_{22}}\right) = \bar{y}_{.\,m_2}$, and hence $E\left(\bar{y}_{.\,m_{22}}\right) = E_1\left[E_2\left(\bar{y}_{.\,m_{22}}\right)\right] = \bar{Y}_{.2}$. Thus,

$$E\left(\bar{y}_{cl\cdot 2}\right) = E_1\left[E_2\left(\bar{y}_{cl\cdot 2}\right)\right] = E_1\left[w_1E_2\left(\bar{y}_{.\,m_1}\right) + w_2E_2\left(\bar{y}_{.\,m_{22}}\right)\right]$$

$$= E_1\left(w_1\bar{y}_{.\,m_1} + w_2\bar{y}_{.\,m_2}\right) = E_1\left(\bar{y}_{cl}\right) \text{ by (2)}$$

$$= \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij} = \bar{Y}$$

This proves the unbiasedness of $\bar{y}_{cl\cdot 2}$ for $\bar{Y}$.

The variance of $\bar{y}_{cl\cdot 2}$ is obtained from the conditional variance formula [Hansen, Hurwitz and Madow (1953), p.65]

$$V\left(\bar{y}_{cl\cdot 2}\right) = V_1E_2\left(\bar{y}_{cl\cdot 2}\right) + E_1V_2\left(\bar{y}_{cl\cdot 2}\right) \quad (3)$$

Now $E_2\left(\bar{y}_{cl\cdot 2}\right) = w_1\bar{y}_{.\,m_1} + w_2\bar{y}_{.\,m_2} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i.}$ where $\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}$.

Hence

$$V_1E_2\left(\bar{y}_{cl\cdot 2}\right) = V_1\left(\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i\right) = \frac{N-n}{Nn}S_B^2 \quad (4)$$

where

$$S_B^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i.}-\bar{Y}\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{1}{M}\sum_{j=1}^{M}\left(Y_{ij}-\bar{Y}\right)\right]^2$$

$$= \frac{1}{M^2(N-1)}\left[\sum_{i=1}^{N} M_1^2\left(\bar{Y}_{i1}-\bar{Y}\right)^2 + \sum_{i=1}^{N} M_2^2\left(\bar{Y}_{i2}-\bar{Y}\right)^2\right]$$

$$= \frac{1}{M^2(N-1)}\left[\sum_{i=1}^{N} M_1^2\left(\bar{Y}_{i1}-\bar{Y}_{.1}\right)^2 + \sum_{i=1}^{N} M_2^2\left(\bar{Y}_{i2}-\bar{Y}_{.2}\right)^2\right] +$$

$$\frac{1}{M^2(N-1)}\left[\sum_{i=1}^{N} M_1^2\left(\bar{Y}_{.1}-\bar{Y}\right)^2 + \sum_{i=1}^{N} M_2^2\left(\bar{Y}_{.2}-\bar{Y}\right)^2\right]$$

$$= W_1^2S_{1B}^2 + W_2^2S_{2B}^2 + \frac{W_1^2N\left(\bar{Y}_{.1}-\bar{Y}\right)^2 + W_2^2N\left(\bar{Y}_{.2}-\bar{Y}\right)^2}{N-1} \quad (5)$$

Substitution of (5) in (4) gives the result

$$V_1E_2\left(\bar{y}_{cl\cdot 2}\right) = \frac{1}{n}\left(1-\frac{n}{N}\right)\left[W_1^2S_{1B}^2 + W_2^2S_{2B}^2\right] +$$

$$\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{W_1^2N\left(\bar{Y}_{.1}-\bar{Y}\right)^2 + W_2^2N\left(\bar{Y}_{.2}-\bar{Y}\right)^2}{N-1}\right] \quad (6)$$

Also

$$V_2\left(\bar{y}_{cl\cdot 2}\right) = V_2\left(w_1\bar{y}_{.\,m_1} + w_2\bar{y}_{.\,m_{22}}\right) = w_2^2V_2\left(\bar{y}_{.\,m_{22}}\right)$$

$$= \frac{m_2^2}{m^2}\,\frac{nm_2-nm_{22}}{nm_2nm_{22}}\,\frac{1}{nm_2-1}\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\ell}-\bar{y}_{.\,m_2}\right)^2$$

$$= \frac{m_2}{m^2}\,\frac{m_2-m_{22}}{nm_{22}}s_{n\,m_2}^2 = \frac{m_2}{m^2}\,\frac{m_{22}(k-1)}{nm_{22}}s_{n\,m_2}^2$$

$$= \frac{w_2(k-1)}{nm}s_{n\,m_2}^2 \quad (7)$$

Where $s^2_{n\,m_2} = \frac{1}{nm_2-1}\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\ell}-\bar{y}_{\cdot m_2}\right)^2$,

$E_1V_2\left(\bar{y}_{cl\cdot 2}\right)=\frac{E_1(m_2)}{m^2}\frac{k-1}{n}E_1\left(s^2_{n\,m_2}\right).$

By Cochran (1963 p.369) $E_1(m_2)=m\frac{NM_2}{NM}=mW_2$ and hence

$$E_1V_2\left(\bar{y}_{cl\cdot 2}\right)=\frac{mW_2(k-1)}{m^2n}S_2^2=\frac{W_2(k-1)}{nm}S_2^2 \qquad (8)$$

where $S_2^2=\frac{1}{NM_2-1}\sum_{i=1}^{N}\sum_{\ell=1}^{M_2}\left(Y_{i\ell}-\bar{Y}_{\cdot 2}\right)^2$.

Substituting of (6) and (8) in (3) gives the result

$$V\left(\bar{y}_{cl\cdot 2}\right) = \frac{1}{n}\left(1-\frac{n}{N}\right)\left[W_1^2S_{1B}^2+W_2^2S_{2B}^2\right]+$$

$$+\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{W_1^2N\left(\bar{Y}_{\cdot 1}-\bar{Y}\right)^2+W_2^2N\left(\bar{Y}_{\cdot 2}-\bar{Y}\right)^2}{N-1}\right]+\frac{W_2(k-1)}{nm}S_2^2 \qquad (9)$$

But

$$S_2^2 = \frac{1}{NM_2-1}\sum_{i=1}^{N}\sum_{\ell=1}^{M_2}\left(Y_{i\ell}-\bar{Y}_{\cdot 2}\right)^2 = \frac{1}{NM_2-1}\sum_{i=1}^{N}\sum_{\ell=1}^{M_2}\left[\left(Y_{i\ell}-\bar{Y}_{i2}\right)+\left(\bar{Y}_{i2}-\bar{Y}_{\cdot 2}\right)\right]^2$$

$$=\frac{1}{NM_2-1}\left[\sum_{i=1}^{N}\sum_{\ell=1}^{M_2}\left(Y_{i\ell}-\bar{Y}_{i2}\right)^2+M_2\sum_{i=1}^{N}\left(\bar{Y}_{i\ell}-\bar{Y}_{\cdot 2}\right)^2\right]$$

$$=\frac{1}{NM_2-1}\left[(M_2-1)\sum_{i=1}^{N}S_{i2W}^2+M_2\sum_{i=1}^{N}\left(\bar{Y}_{i2}-\bar{Y}_{\cdot 2}\right)^2\right]$$

$$=\frac{M_2-1}{NM_2-1}\sum_{i=1}^{N}S_{i2W}^2+\frac{M_2\sum_{i=1}^{N}\left(Y_{i\ell}-\bar{Y}_{\cdot 2}\right)^2}{NM_2-1} \qquad (10)$$

$$=\frac{N(M_2-1)}{NM_2-1}S_{\cdot 2W}^2+\frac{M_2(N-1)}{NM_2-1}S_{2B}^2 \qquad (11)$$

where $S_{12W}^2=\frac{1}{M_2-1}\sum_{\ell=1}^{M_2}\left(Y_{i\ell}-\bar{Y}_{i2}\right)^2$

and $S_{\cdot 2W}^2=\frac{1}{N(M_2-1)}\sum_{i=1}^{N}\sum_{\ell=1}^{M_2}\left(Y_{i\ell}-\bar{Y}_{i2}\right)^2$.

Substitution of (10) or (11) in (9) gives the result

$$V\left(\bar{y}_{cl\cdot 2}\right) = \frac{1}{n}\left(1-\frac{n}{N}\right)\left[W_1^2S_{1B}^2+W_2^2S_{2B}^2\right]+$$

$$\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{W_1^2N\left(\bar{Y}_{\cdot 1}-\bar{Y}\right)^2+W_2^2N\left(\bar{Y}_{\cdot 2}-\bar{Y}\right)^2}{N-1}\right]+$$

$$+\frac{W_2(k-1)}{NM_2-1}\left[\frac{(M_2-1)\sum_{i=1}^{N}S_{i2W}^2+M_2\sum_{i=1}^{N}\left(\bar{Y}_{i2}-\bar{Y}_{\cdot 2}\right)^2}{n\,m}\right] \qquad (12)$$

$$=\frac{1}{n}\left(1-\frac{n}{N}\right)\left[W_1^2S_{1B}^2+W_2^2S_{2B}^2\right]+$$

$$\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{W_1^2N\left(\bar{Y}_{\cdot 1}-\bar{Y}\right)^2+W_2^2N\left(\bar{Y}_{\cdot 2}-\bar{Y}\right)^2}{N-1}\right]+$$

$$+\frac{W_2(k-1)}{NM_2-1}\left[\frac{N(M_2-1)S_{\cdot 2W}^2+M_2(N-1)S_{2B}^2}{n\,m}\right] \qquad (13)$$

This result depends positively on the variance of subgroups of respondents in the different clusters, the variance of subgroups of the nonrespondents in those clusters, the variance within the nonresponse stratum, the difference between the response stratum and the population as a whole as well as the difference between the nonresponse stratum and the population as a whole and hence the difference between the response stratum and the nonresponse stratum with respect to the variable under study.

ESTIMATION OF VARIANCE OF THE ESTIMATOR

To obtain an unbiased estimator of $V_1E_2\left(\bar{y}_{cl\cdot 2}\right)$ in (6) we only need an unbiased estimator of S_B^2 which is $s_B^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{y}_{i\cdot}-\bar{y}\right)^2$. Since $\bar{y}$ cannot be calculated in the presence of nonresponse, we use $\bar{y}_{cl\cdot 2}$ [Rao(1986)] and obtain

$$s_B^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{y}_{i\cdot}-\bar{y}_{cl\cdot 2}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}\left(y_{ij}-\bar{y}_{cl\cdot 2}\right)\right]^2$$

$$=\frac{1}{m^2(n-1)}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m_1}\left(y_{ij}-\bar{y}_{cl\cdot 2}\right)+\left(y_{i\ell}-\bar{y}_{cl\cdot 2}\right)\right]^2$$

Now

$$\sum_{j=1}^{m_1}\left(y_{ij}-\bar{y}_{cl\cdot 2}\right)=\sum_{j=1}^{m_1}\left[\left(y_{ij}-\bar{y}_{i1}\right)+\left(\bar{y}_{i1}-\bar{y}_{cl\cdot 2}\right)\right]=m_1\left(\bar{y}_{i1}-\bar{y}_{cl\cdot 2}\right)$$

$$\sum_{\ell=1}^{m_2}\left(y_{i\ell}-\bar{y}_{cl\cdot 2}\right)=\sum_{\ell=1}^{m_2}\left[\left(y_{i\ell}-\bar{y}_{i2}\right)+\left(\bar{y}_{i2}-\bar{y}_{cl\cdot 2}\right)\right]=m_2\left(\bar{y}_{i2}-\bar{y}_{cl\cdot 2}\right)$$

and hence

$$S_B^2 = \frac{1}{m^2(n-1)}\left[\sum_{i=1}^{n}m_1^2\left(\bar{y}_{i1}-\bar{y}_{cl\cdot 2}\right)^2+\sum_{i=1}^{n}m_2^2\left(\bar{y}_{i2}-\bar{y}_{cl\cdot 2}\right)^2\right] \qquad (14)$$

But

$$\sum_{i=1}^{n}m_1^2\left(\bar{y}_{i1}-\bar{y}_{cl\cdot 2}\right)^2 = \sum_{i=1}^{n}m_1^2\left[\left(\bar{y}_{i1}-\bar{y}_{\cdot 1}\right)+\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)\right]^2;$$

$\bar{y}_{\cdot 1}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i1}$

$$=\sum_{i=1}^{n}m_1^2\left(\bar{y}_{i1}-\bar{y}_{\cdot 1}\right)^2+\sum_{i=1}^{n}m_1^2\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2 \qquad (15)$$

Similarly,

$$\sum_{i=1}^{n}m_2^2\left(\bar{y}_{i2}-\bar{y}_{cl\cdot 2}\right)^2=\sum_{i=1}^{n}m_2^2\left(\bar{y}_{i2}-\bar{y}_{\cdot 2}\right)^2+\sum_{i=1}^{n}m_2^2\left(\bar{y}_{\cdot 2}-\bar{y}_{cl\cdot 2}\right)^2;$$

$\bar{y}_{\cdot 2}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i2}$

$$=\sum_{i=1}^{n}m_2^2\left[\frac{1}{m_2}\sum_{\ell=1}^{m_2}\left(y_{i\ell}-\bar{y}_{\cdot 2}\right)\right]^2+\sum_{i=1}^{n}m_2^2\left(\bar{y}_{\cdot 2}-\bar{y}_{cl\cdot 2}\right)^2$$

From Chochran (1977), p.372 we express m_2 in terms of m_{22} as $m_2 = km_{22}$ and hence

$$\sum_{i=1}^{n} m_2^2\left(\bar{y}_{i\,2}-\bar{y}_{cl\cdot 2}\right)^2=\sum_{i=1}^{n}\left[\sum_{t=1}^{k}\sum_{s=1}^{m_{22}} y_{i\,s}-km_{22}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{km_{22}}\sum_{t=1}^{k}\sum_{s=1}^{m_{22}} y_{i\,s}\right)\right]^2$$

$$+\sum_{i=1}^{n} k^2 m_{22}^2\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_{22}}\sum_{s=1}^{m_{22}}\left(y_{i\,s}-\bar{y}_{cl\cdot 2}\right)\right]^2$$

$$=\sum_{i=1}^{n}\left[k\sum_{s=1}^{m_{22}} y_{i\,s}-k\frac{\sum_{i=1}^{n}\sum_{s=1}^{m_{22}} y_{i\,s}}{n}\right]^2+k^2 m_{22}^2\sum_{i=1}^{n}\left[\frac{1}{m_{22}n}\sum_{i=1}^{n}\sum_{s=1}^{m_{22}}\left(y_{i\,s}-\bar{y}_{cl\cdot 2}\right)\right]^2$$

$$=k^2\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{\sum_{i=1}^{n} y_{i\,m_{22}}}{n}\right)^2+nk^2 m_{22}^2\left(\bar{y}_{\cdot\, m_{22}}-\bar{y}_{cl\cdot 2}\right)^2 \qquad (16)$$

Substitution of (15) and (16) in (14) gives the result

$$s_B^2=\frac{1}{m^2(n-1)}\left[\sum_{i=1}^{n} m_1^2\left(\bar{y}_{i\,1}-\bar{y}_{\cdot 1}\right)+nm_1^2\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2+\right.$$

$$\left.+k^2\left\{\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2+nm_{22}^2\left(\bar{y}_{\cdot\, m_{22}}-\bar{y}_{cl\cdot 2}\right)^2\right\}\right]$$

$$=\frac{m_1^2}{m^2}\frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{y}_{i\,1}-\bar{y}_{\cdot 1}\right)^2+\frac{k^2}{m^2}\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2+$$

$$+\frac{m_1^2}{m^2}\frac{n}{n-1}\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2+\frac{k^2 m_{22}^2}{m^2}\frac{n}{n-1}\left(\bar{y}_{\cdot\, m_{22}}-\bar{y}_{cl\cdot 2}\right)^2$$

$$=w_1^2 s_{1B}^2+k^2\frac{s_{y\,B\,m}^2}{m^2}+\frac{w_1^2 n\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2+w_2^2 n\left(\bar{y}_{\cdot\, m_{22}}-\bar{y}_{cl\cdot 2}\right)^2}{n-1} \qquad (17)$$

where $s_{y\,B\,m_{22}}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2$ is the variance between subsamples of nonrespondents from the selected clusters. Substitution of (17) for S_B^2 in (4) gives the following estimator of $V_1E_2\left(\bar{y}_{cl\cdot 2}\right)$

$$\hat{V}_1E_2\left(\bar{y}_{cl\cdot 2}\right)=\frac{1}{n}\left(1-\frac{n}{N}\right)\left[w_1^2 s_{1B}^2+k^2\frac{s_{y\,B\,m_{22}}^2}{m^2}\right]+$$

$$+\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{w_1^2 n\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2+w_2^2 n\left(\bar{y}_{\cdot\, m_{22}}-\bar{y}_{cl\cdot 2}\right)^2}{n-1}\right] \qquad (18)$$

To obtain an unbiased estimator of $E_1V_2\left(\bar{y}_{cl\cdot 2}\right)$ in (8), we only need to express $S_{n\,m_2}^2$ in (7) in terms of information from the m_{22} units that responded later.

Thus

$$\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\,\ell}-\bar{y}_{\cdot\, m_2}\right)^2=\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left[\left(y_{i\,\ell}-\bar{y}_{i\,m_2}\right)+\left(\bar{y}_{i\,m_2}-\bar{y}_{\cdot\, m_2}\right)\right]^2$$

$$=\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\,\ell}-\bar{y}_{i\,m_2}\right)^2+m_2\sum_{i=1}^{n}\left(\bar{y}_{i\,m_2}-\bar{y}_{\cdot\, m_2}\right)^2 \ldots \qquad (19)$$

and

$$\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\,\ell}-\bar{y}_{i\,m_2}\right)^2=\sum_{i=1}^{n}\sum_{t=1}^{k}\sum_{s=1}^{m_{22}}\left(y_{i\,s}-\bar{y}_{i\,m_{22}}\right)^2=k\sum_{i=1}^{n}\sum_{s=1}^{m_{22}}\left(y_{i\,s}-\bar{y}_{i\,m_{22}}\right)^2$$

$$=k\left(m_{22}-1\right)\sum_{i=1}^{n} s_{i\,w\,m_{22}}^2 \qquad (20)$$

where $s_{i\,w\,m_{22}}^2=\frac{1}{m_{22}-1}\sum_{s=1}^{m_{22}}\left(y_{i\,s}-\bar{y}_{i\,m_{22}}\right)^2$, and

$$m_2\sum_{i=1}^{n}\left(\bar{y}_{i\,m_2}-\bar{y}_{\cdot\, m_2}\right)^2=\frac{m_2}{m_{22}^2}\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2 \qquad (21)$$

Substitution of (20) and (21) in (19) gives the results

$$\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\,\ell}-\bar{y}_{\cdot\, m_2}\right)^2=k(m_{22}-1)+\sum_{i=1}^{n} s_{i\,w\,m_{22}}^2+\frac{k}{m_{22}}\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2$$

and hence

$$s_{n\,m_2}^2=\frac{1}{nm_2-1}\sum_{i=1}^{n}\sum_{\ell=1}^{m_2}\left(y_{i\,\ell}-\bar{y}_{\cdot\, m_2}\right)^2$$

$$=\frac{k(m_{22}-1)\sum_{i=1}^{n} s_{i\,w\,m_{22}}^2+\frac{k}{m_{22}}\sum_{i=1}^{n}\left(y_{i\,m_{22}}-\frac{y_{\cdot\, m_{22}}}{n}\right)^2}{n\,m_2-1}$$

$$=\frac{nk(m_{22}-1)s_{\cdot\,w\,m_{22}}^2+k(n-1)s_{y\,B\,m_{22}}^2\big/m_{22}}{n\,m_2-1} \qquad (22)$$

Substitution of (22) in (7) gives the result

$$E_1\hat{V}_2\left(\bar{y}_{cl\cdot 2}\right)=\frac{w_2k(k-1)}{n\,m}\left[\frac{(m_{22}-1)s_{w\,m_{22}}^2+(n-1)s_{y\,B\,m_{22}}^2\big/m_{22}}{n\,m_2-1}\right] \qquad (23)$$

A combination of (18) and (23) under (3) gives the following estimator of $V\left(\bar{y}_{cl\cdot 2}\right)$

$$\hat{V}\left(\bar{y}_{cl\cdot 2}\right)=\frac{1}{n}\left(1-\frac{n}{N}\right)\left[w_1^2 s_{1B}^2+k^2\frac{s_{y\,B\,m_{22}}^2}{m^2}\right]+$$

$$+\frac{1}{n}\left(1-\frac{n}{N}\right)\left[\frac{w_1^2 n\left(\bar{y}_{\cdot 1}-\bar{y}_{cl\cdot 2}\right)^2+w_2^2 n\left(\bar{y}_{\cdot m_{22}}-\bar{y}_{cl\cdot 2}\right)^2}{n-1}\right]+$$

$$+\frac{w_2 k(k-1)}{n\,m}\left[\frac{n(m_{22}-1)s^2_{\cdot w\,m_{22}}+(n-1)s^2_{y\,B\,m_{22}}\Big/m_{22}}{n\,m_2-1}\right]$$

APPLICATION

A study of average value of fishing tools per fisherfolk in the coastal region of Cross River State of Nigeria in 1993 was based on the above design. The population studied consisted of 24 clusters of 100 households each with a large subpopulation of non-native fisherfolks. In each cluster, 70 (M_1) of the households were native and would respond and the other 30 (M_2) were non-native households and would not respond.

A random sample of 8 clusters was drawn without replacement. Then 60 households were drawn at random and without replacement from each selected cluster by selecting 42 units (m_1) from the response subpopulation and 18 units (m_2) from the nonresponse subpopulation. During the survey, all the m_1 = 42 units responded on the study variable but the other m_2 = 18 units did not. It was obvious that apart from the order by the union of "non-native fisherfolks" to its members not to answer survey questions that related to their fishing operations, these nonrespondents were also afraid of property tax on their fishing tools.

A sample of 9 (50%) units was drawn at random and without replacement from the 18 nonrespondents in each cluster. All the 9 units were reinterviewed by specially trained persons who tried to emphasize the benefits, such as assistance from micro-credit scheme, which response to the survey questions could bring to the respondents. All the 9 fisherfolks in the subsample in each cluster responded.

A summary of the sample observations in thousands of Naira is given in Table 1.

Table 1. Summary of a result in fishing tools (in thousands of naira) in cross river state, Nigeria, in 1993

Sample Cluster	$\sum_{j=1}^{m_1} y_{ij}$	$\bar{y}_{i1}$	$y_{i\,m_{22}}$	$\bar{y}_{i\,m_{22}}$	$s^2_{i\,w\,m_{22}}$
1	591	14.0	293	32.6	83.37
2	867	20.6	349	38.8	71.21
3	780	18.6	344	38.2	202.15
4	797	19.0	343	38.7	99.51
5	772	18.4	374	41.6	140.75
6	765	18.2	433	48.1	67.85
7	845	20.1	366	40.7	136.76
8	736	17.5	381	42.3	83.39
Total	**6153**	**146.4**	**2888**	**321.0**	**884.99**

$\bar{y}_{i1}$ = 18.3, $\bar{y}_{i\,m_{22}}$ = 40.1, w_1 = 0.7, w_2 = 0.3; $\bar{y}_{cl\cdot 2}$ = (18.3 × 0.7) + (40.1 × 0.3) = 24.84;

$s^2_{1B}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{y}_{i1}-\bar{y}_{\cdot 1}\right)^2=\frac{28.36}{7}=4.04$; $s^2_{y\,B\,m_{22}}=\frac{11{,}004}{8}=157200$; $s^2_{i\,w\,m_{22}}$ = 884.99/8 = 110.62

$$\frac{1}{n}\left(1-\frac{n}{N}\right)\left[w_1^2 s^2_{1B}+k^2\frac{s^2_{y\,B\,m_{22}}}{m^2}\right]=0.3107;$$

$$\frac{w_2 k(k-1)}{nm_2-1}\left[\frac{n(m_{22}-1)s^2_{\cdot w\,m_{22}}+(n-1)s^2_{y\,B\,m_{22}}\Big/m_{22}}{n\,m}\right]=0.0553$$

$$\hat{V}\left(\bar{y}_{cl\cdot 2}\right) = 0.3107 + 4.1814 + 4.1814 + 0.0553 = 4.5474$$

DISCUSSION AND CONCLUSION

The components of variance of the estimator under the sampling design discussed in section 4 of this paper are (a) variance between subgroups of respondents located in the different clusters, (b) variance between subgroups of nonrespondents also located in the different clusters, (c) variance within the nonresponse stratum within the survey population, and (d) the difference between the response and the nonresponse strata. Since these various factors of variation are outside the control of any survey designer, the variance of an estimator under the sample design can be minimized by increasing the sample sizes (n, m and m_{22}) at all the stages of sampling as high as possible.

REFERENCES

Birnbaum, Z.W. and Sirken, M. G. (1950). Bias due to nonavailability in sampling surveys. *Jour. Amer. Stat. Assoc.*, 45: 98-111.

Cochran, W. G. (1977). *Sampling Techniques*. John Wiley and Sons. Inc., New York.

Cochran, W. G., Mosteller, F and Tukey, J. W. (1954). Statistical *Problems of the Kinsey Report*. American Statitistical Association, Washington D. C.

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of nonresponse in sample surveys. *Jour. Amer. Stat. Assoc*., 41:517-529.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Survey Sample Methods and Theory*: John Wiley and Sons, Inc., New York.

Little, R. J. A. (1983). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77: 237-350.

Rao, Poduri S. R. S. (1986). Ratio estimation with subsampling the nonrespondents. *Survey Methodology*, 12: 217 – 230.

Sarndal, E. E. and Swensson, B. (1985). Incoporating nonresponse modeling in a general randomization theory approach. *Bull. Int. Stat. Inst.*, 45th session, LI – 3, 15.2:1-16.

Udofia, G. A. (2004). Ratio estimation for small domains with subsampling the nonrespondents: An application of Rao strategy. *Statistics in Transition* :713-724.

Udofia, G. A. (2005). Estimation in single-stage sampling and with equal subsamples of the nonrespondents: *Global Journal of Mathematical Sciences*.4: 53-64.