SURVEY DESIGNS WHICH EMPLOY DOUBLE SAMPLING SCHEMES FOR ELIMINATING MEASUREMENT PROCESS BIAS

Judith T. Lessler, Research Triangle Institute

INTRODUCTION

The present paper describes some survey designs which employ double sampling schemes in order to reduce the measurement errors associated with estimates from a survey. Suppose one has available for use in a survey two measurement processes: a cheap-faulty measurement process and an expensive-accurate measurment process. If the net bias associated with the faulty measurement process is large, it may be advantageous to use a double sampling scheme for eliminating the measurement process bias.

The concepts employed are based upon the Census Bureau model for measurement errors [1,2] which may be briefly described as follows: There exists a population of N individuals. For each individual in the population, one wishes to measure their values on a set of p characteristics. For a particular measurement process, the measurement obtained for the i-th individual at the t-th trial of the survey is $\underset{\sim}{Y}_{it}$. The subscript t indexes a series of repeatable trials of the measurement process (i.e. of the census or survey in question).

One can define the expected value over trials of the measurement for the i-th individual:

$$E_t \{\underset{\sim}{Y}_{it} | U_i=1\} = \underset{\sim}{Y}_i \quad , \tag{1}$$

where U_i is an indicator random variable denoting presence in the sample.

If we denote the "true" or actual values for the i-th individual as $\underset{\sim}{X}_i$, then the expected measurement for the i-th individual may not be equal to the actual or "true" values for that individual. That is $\underset{\sim}{X}_i \neq \underset{\sim}{Y}_i$ and there may be a net bias in the measurement process. For example, if one wishes to estimate the population mean $\bar{\underset{\sim}{X}} = \frac{1}{N} \sum_{i=1}^{N} \underset{\sim}{X}_i$, there is a net bias in the measurement process if $\bar{\underset{\sim}{X}} \neq \bar{\underset{\sim}{Y}}$, where

$$\bar{\underset{\sim}{Y}} = \frac{1}{N} \sum_{i=1}^{N} \underset{\sim}{Y}_i \ .$$

We will term a measurement process which measures $\underset{\sim}{Y}_{it}$ as a faulty measurement process. For a simple random sample of size n, the estimate of the population mean using the faulty measurement process is

$$\bar{\underset{\sim}{y}}_t = \frac{1}{n} \sum_{i=1}^{N} U_i \underset{\sim}{Y}_{it} \quad . \tag{2}$$

Now, $E(\bar{\underset{\sim}{y}}_t) = \bar{\underset{\sim}{Y}}$; and, assuming that there is no interaction between the sampling errors and measurement errors, i.e.,

$$E_t \{\underset{\sim}{Y}_i | U_i = U_{i'} = 1\} = \underset{\sim}{Y}_{ii'} = \underset{\sim}{Y}_i \tag{3}$$

(see Koch [2]) then the mean square error matrix of $\bar{\underset{\sim}{y}}_t$ is

$$\begin{aligned} MSE(\bar{\underset{\sim}{y}}_t) = \frac{1}{n} & (\underset{\sim}{SMV}) + (n-1)(\underset{\sim}{CMV}) \\ & + \frac{1}{n}\left(\frac{N-n}{N}\right) \underset{\sim}{BV} + \underset{\sim}{BT} + \underset{\sim}{TV} \\ & + \underset{\sim}{B} \ . \end{aligned} \tag{4}$$

The first term is the measurement variance (or response variance), $\underset{\sim}{MV}$ and consists of two terms: $\underset{\sim}{SMV}$ the simple measurement variance and $\underset{\sim}{CMV}$, the correlated measurement variance. This term arises because the measurements obtained for the i-th individual are not the same from trial to trial of the survey.

The second term is the sampling variance, $\underset{\sim}{SV}$, and is due to the variability of the $\underset{\sim}{Y}_i$ around $\bar{\underset{\sim}{Y}}$. The sampling variance is composed of three components: $\underset{\sim}{BV}$, the sampling variance of the individual bias terms, $\underset{\sim}{B}_i = \underset{\sim}{Y}_i - \underset{\sim}{X}_i$ around the net bias $\bar{\underset{\sim}{B}}$; $\underset{\sim}{TV}$, the sampling variance of the true values; and $\underset{\sim}{BT}$, the interaction between the individual bias terms and the true values.

The third term $\underset{\sim}{B}$ is due to the net bias in the measurement process.

THE DOUBLE SAMPLING SCHEME IN GENERAL

The general double sampling scheme (DSS) is as follows: the survey is conducted in two phases. In phase 1, an initial sample is drawn and faulty measurements are obtained. In phase 2, a subsample fo the original sample is drawn and one of two schemes are employed:

1. Repeat measurement for each individual in the subsample are made using the faulty measurement process and the accurate values are obtained for each individual in the subsample. This scheme allows one to simultaneously estimate components of measurement error and eliminate the measurement bias. Assuming srs of size n and n_1 at the two phases of the survey, the unbiased estimate of $\bar{\underset{\sim}{X}}$ is

$$\bar{\underset{\sim}{z}}_t = \bar{\underset{\sim}{y}}_{1t} - (\bar{\underset{\sim}{y}}_{2t} - \bar{\underset{\sim}{x}}_2) \quad . \tag{5}$$

If we assume that the measurements obtained at the two phases of the survey with the faulty measurement process are independent, the variance-covariance matrix of $\bar{\underset{\sim}{z}}_t$ is given by

$$\begin{aligned} \underset{\sim}{V} = & \frac{1}{n}\{(\underset{\sim}{SMV}) + (n-1)(\underset{\sim}{CMV})\} \\ & + \frac{1}{n_1}\{(\underset{\sim}{SMV}) + (n_1-1)(\underset{\sim}{CMV})\} \\ & + \frac{n-n_1}{nn_1}\{\underset{\sim}{BV}\} \\ & + \frac{1}{n}\left(\frac{N-n}{N}\right)\{\underset{\sim}{TV}\} \ . \end{aligned} \tag{6}$$

2. One may measure only the accurate values in the subsample which allows one to eliminate the measurement process bias but not estimate the components or error. In this case, an unbiased estimator for $\bar{\underset{\sim}{X}}$ is

$$\bar{\underset{\sim}{w}}_t = \bar{\underset{\sim}{y}}_t - (\bar{\underset{\sim}{y}}_{st} - \bar{\underset{\sim}{x}}_s) \quad , \tag{7}$$

and

$$\underset{\sim}{V} = \frac{n-n_1}{nn_1} \{(\underset{\sim\sim\sim}{SMV}) - (\underset{\sim\sim\sim}{CMV})\}$$
$$\frac{n-n_1}{nn_1} \quad \{(\underset{\sim\sim}{BV})\} + \frac{1}{n}\left(\frac{N-n}{N}\right) \quad \{\underset{\sim\sim}{TV}\} \ . \tag{8}$$

Greater detail concerning the DSS may be found in Lessler [3].

SPECIFIC SURVEY DESIGNS WHICH EMPLOY DSS's

The above DSS may be adapted to a variety of survey situations from the simple to the complex. The focus of the first scheme is to allow one to estimate the components of the mean square error of estimates using faulty measurement processes and to form estimates which are free of the net bias associated with the faulty measurement process. This is particularly important for the evaluation of ongoing surveys in which one wishes to estimate components of error in order that future adjustments can be made in the survey procedures that will accomplish a reduction of these errors. In addition, it would be useful for pilot studies of surveys in which alternative measurement processes including alternative questionnaires, types of interviewers, and other procedures are to be evaluated.

Two specific survey designs which employ the first form of the DSS are illustrated in the following:

A. A self-enumeration survey employing simple random sampling.

An original sample is drawn and the faulty measurements are obtained. In a second phase of the survey, the accurate values are obtained for members of the original sample as well as remeasurements with the faulty measurement process. For example, the faulty measurement process might be a mail survey in which individuals were queried as to certain demographic characteristics and their bank balance, income, health expenses, etc. A subsample is drawn and remeasurements obtained by the mail survey. In addition, record checks are done to obtain the accurate values.

The following model for the faulty measurements is assumed. Here, α indicates the phase of the survey:

$$Y_{i\alpha t} = X_i + B_{i\alpha t} \quad . \tag{9}$$

We assume the following:

1. $$E_t(Y_{i\alpha t}) = Y_i \quad , \alpha = 1, 2 \tag{10}$$

2. $$E_t\,(B_{i\alpha t}) = B_i \quad , \alpha = 1, 2 \tag{11}$$

3. The measurement process is equally variable from trial to trial so that

$$E_t\{(B_{i\alpha t} - B_i)^2\} = \gamma_i^{\ 2} \qquad \text{for } \alpha = 1, 2. \tag{12}$$

4. The measurement process for each individual in the sample is statistically independent of that for any other individual and is statistically independent between phases for a particular individual, i.e.,

$$\mathrm{Cov}(B_{i\alpha t}, B_{i'\alpha t}) = \mathrm{Cov}(B_{i\alpha t}, B_{i\alpha' t}) = 0$$

for $i \neq i'$, $\alpha \neq \alpha'$.

The variance of $\bar{z}_t$ in terms of this specific model is

$$V(\bar{z}_t) = \frac{n+n_1}{nn_1}\{S_\gamma^{\ 2}\} + \frac{n-n_1}{nn_1}\ \{S_B^{\ 2}\} + \frac{1}{n}\left(\frac{N-n}{N}\right) \quad \{S^2\} \ ; \tag{13}$$

where

$$S_\gamma^{\ 2} = \frac{1}{N}\sum_{i=1}^{N} \gamma_i^{\ 2} \ ;$$

$$S_B^{\ 2} = \frac{1}{N-1}\sum_{i=1}^{N} (B_i - \bar{B})^2 \ ;$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \bar{X})^2 \qquad .$$

Letting U_i and V_i be the indicator random variables indicating presence in the sample and subsample respectively, we have the following sample estimators for the above variance components.

1. The subsample variance of the true values $s_x^{\ 2}$,

$$s_x^{\ 2} = \frac{1}{n_1 - 1}\sum_{i=1}^{N} [V_i(X_i - \bar{x})]^2 \quad .$$

$$E(s_x^{\ 2}) = S^2 \quad .$$

2. The subsample variance of the bias terms,

$$s_B^{\ 2} = \frac{1}{n_1 - 1}\sum_{i=1}^{N} \{V_i[Y_{i2t} - X_i) - (\bar{y}_{2t} - \bar{x}_2)]\}^2 \ .$$

$$E(s_B^{\ 2}) = S_\gamma^{\ 2} + S_B^{\ 2} \quad .$$

3. The between-phase within individual sum of squares s_w^2,

$$s_w^2 = \sum_{i=1}^{N} U_i V_i (Y_{i1t}-Y_{i2t})^2 .$$

$$E(\frac{s_w^2}{2_{n1}}) = s_Y^2 .$$

Thus we have the following set of estimators:

$$\hat{S}^2 = s_x^2 ,$$

$$\hat{S}^2 = \frac{s_w^2}{2n_1} ,$$

$$\hat{S}_B^2 = s_B^2 - \frac{s_w^2}{2n_1} .$$

B. Survey Design Using Interviewers

In addition to the population of N individuals, let there be a fixed population of B interviewers, indexed by the subscript j, available for use in the survey. A simple random sample of size n is drawn for the first phase of the survey. Each individual in the sample is assigned at random to one of the interviewers. The interviewer structure is characterized by indicator random variables C_{ij} where

$$C_{ij} = \begin{cases} 1 & \text{if the i-th individual is assigned to the j-th interviewer} \\ 0 & \text{otherwise.} \end{cases}$$

For simplicity, we assume n = Br. The subsample and interviewer structure at the second phase is characterized by indicator random variables V_i and D_{ij} respectively. $Y_{ij\alpha t}$ is the measurement obtained for the i-th individual by the j-th interviewer at the α-th phase of the t-trial of the survey process. A specific model for the faulty measurements is:

$$Y_{ij\alpha t} = \bar{X}+H_i+\bar{B}+L_i+Q_j+(IQ)_{ij}+Z_{j\alpha t}+R_{ij\alpha t} \tag{14}$$

where the effects in the model are defined using the following,

$$E_t(Y_{ij\alpha t}) = Y_{ij} ,$$

$$Y_i = \frac{1}{N} \sum_{j=1}^{B} Y_{ij} ,$$

$$\bar{Y}_{.j} = \frac{1}{N} \sum_{i=1}^{N} Y_{ij} ,$$

$$\bar{Y}_{.j\alpha t} = \frac{1}{N} \sum_{i=1}^{N} Y_{ij\alpha t} ,$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \frac{1}{NB} \sum_{i=1}^{N} \sum_{j=1}^{B} Y_{ij} ,$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i ,$$

$$B_i = Y_i - X_i ,$$

$$\bar{B} = \frac{1}{N} \sum_{i=1}^{N} B_i ,$$

which gives

$$H_i = X_i - \bar{X} ,$$

$$L_i = (Y_i-X_i) - \bar{B} = B_i - \bar{B} ,$$

$$I_i = H_i + L_i = Y_i - \bar{X} - \bar{B} = Y_i - \bar{Y} ,$$

$$Q_j = \bar{Y}_{.j} - \bar{Y} ,$$

$$(IQ)_{ij} = Y_{ij} - I_i - Q_j - \bar{Y} ,$$

$$Z_{i\alpha t} = \frac{1}{N} \sum_{i-1}^{N} (Y_{ij\alpha t}-Y_{ij}) = \bar{Y}_{.j\alpha t} - Y_{.j} ,$$

and

$$R_{ij\alpha t} = Y_{ij\alpha t} - Y_{ij} - Z_{j\alpha t} .$$

Let

$$E_t \{Z_{j\alpha t}^2\} = \xi_j^2 \tag{15}$$

and

$$E_t \{R_{ij\alpha t}^2\} = \eta_{ij}^2 . \tag{16}$$

If we assume that the measurement process associated with each interviewer is statistically independent of that associated with any other interviewer and is statistically independent from phase to phase for the same interviewer, but may be correlated within interviewers at a particular phase, then the variance of $\bar{z}_t$ in terms of this specific model is

$$V(\bar{z}_t) = \frac{2}{B} \{S_\xi^2\} + \frac{1}{nn_1} (\frac{n(N-r_1)+n_1(N-r)}{N-1}) \{S_\eta^2\} + \frac{n+n_1}{nn_1}\{S_{IQ}^2\} + \frac{n-n_1}{nn_1} \{S_B^2\} + \frac{1}{n}(\frac{N-n}{N}) \{S^2\}. \tag{17}$$

Each variance component is defined as follows:

$$S_\xi^2 = \{\frac{1}{B}\sum_{j=1}^{B} \xi_j^2\}$$

= Interviewer random effect variance

$$S_\eta^2 = \{\frac{1}{NB}\sum_{i=1}^{N}\sum_{j=1}^{B} \eta_{ij}^2\}$$

= Interviewer individual interaction random effect variance

$$S_{IQ}^2 = \{\frac{1}{B(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{B} (IQ)_{ij}^2\}$$

= Interviewer individual interaction fixed effect variance

$$S^2 = \{\frac{1}{N-1}\sum_{i=1}^{N} H_i^2\}$$

= Sampling variance

$$S_B^2 = \{\frac{1}{N-1}\sum_{L=1}^{N} L_i^2\}$$

= Sampling variance of the bias terms.

Given the above assumptions, the following set of estimators allows one to estimate each of these components.

1. $s_x^2 = \hat{S}^2$

2. Between-phase within interviewer within individual sum of squares, BPWII,

$$BPWII, = \sum_{i=1}^{N}\sum_{j=1}^{B} U_i V_i C_{ij} D_{ij} (Y_{ij1t} - Y_{ij2t})^2 .$$

$$E(BPWII) = 2r_1(S_\xi^2 + S_\eta^2) \quad .$$

3. Between phase within interviewer between individual sum of squares, BPWIBI,

$$BPWIBI = \sum_{i\neq i'}^{N}\sum_{j=1}^{B} U_i U_{i'} V_i V_{i'} C_{ij} C_{i'j} D_{ij} D_{i'j} \times$$

$$[(Y_{ij1t} - Y_{i'j1t}) - Y_{ij2t} - Y_{i'2t})]^2 .$$

$$E(BPWIBI) = \frac{r_1(r_1-1)\ (r-1)\ 4N}{(n-1)\ (N-1)} \quad S_\eta^2 \quad .$$

4. Within phase within interviewer between individual sum of squares, WPWIBI,

$$WPWIBI = \sum_{i\neq i'}^{N}\sum_{j=1}^{B} U_i U_{i'} C_{ij} C_{i'j} [Y_{ij1t} - Y_{i'j1t}]^2 .$$

$$E(WPWIBI) = 2\ Br(r-1)\ \{S^2 + S_B^2 + S_{IQ}^2 + \frac{N}{N-1} S_\eta^2\}.$$

5. Between phase between interviewer between individual sum of squares, BPBIBI,

$$BPBIBI = \sum_{i\neq i'}^{N}\sum_{j\neq j'}^{B} U_i U_{i'} V_i V_{i'} C_{ij} C_{i'j'} D_{ij} D_{i'j'} \times$$

$$[(Y_{ij1t} - Y_{i'j1t}) - (Y_{ij'2t} - Y_{i'j'2t})]^2 \quad .$$

$$E(BPBIBI) = \frac{4(r-1)r_1(r_1-1)}{n-1} BS_{IQ}^2 + \frac{(B-1)N}{N-1} S_\eta^2 \quad .$$

Thus we have the following set of estimators.

$$\hat{S}^2 = s_x^2 \quad ,$$

$$\hat{S}_\eta^2 = \frac{\dfrac{BPWIBI}{r_1(r_1-1)(r-1)4N}}{(n-1)(N-1)} \quad ,$$

$$\hat{S}_\xi^2 = \frac{BPWII}{2r_1} - \hat{S}_\eta^2 \quad ,$$

$$\hat{S}_{IQ}^2 = \frac{\dfrac{BPBIBI}{4(r-1)r_1(r_1-1)}}{(n-1)\ B} - (\frac{B-1}{B})\ (\frac{N}{N-1})\ \hat{S}_\eta^2 ,$$

and

$$\hat{S}_B^2 = \frac{WPWIBI}{2Br(r-1)} - \hat{S}^2 - \hat{S}_{IQ}^2 - \frac{N}{N-1}\ \hat{S}_\eta^2 \ .$$

The second form of the DSS is not directed at getting estimates of the components of error but, rather, at eliminating the net bias associated with the faulty measurement process. In addition, it should be noted that the correlated measurement variance makes a negative contribution to the variance of the estimate. This component, when present, is thought by many to make the largest contribution to the MSE of estimates using the faulty measurement process. As example of a survey which may employ the second form of the DSS is as follows:

C. A Multistage Cluster Sampling Design for the National Medical Care Expenditure Survey

RTI is in the process of designing a National Medical Care Expenditure Survey. The original specifications called for conducting household interviews in which the medical care expenditures of each member of the household are collected along with other data. Following this, the medical care provider and third party payors (TPP) were to be visited and the actual cost of the medical care was to be obtained. The obtaining of provider data and TPP data is an expensive procedure and cost savings could result if provider data are collected on a subsample basis and these data used to correct for the biases in the entire household interview data. The proposed plan for the sur

An original sample of households is drawn and interview values obtained for each individual in the household. A subsample of the households and individuals within households is drawn and the accurate values obtained for each individual in the subsample. The survey design is as follows:

Y_{ijkt} = The interview value obtained for the k-th individual, in the j-th household, of the i-th cluster, at the t-th trial of the interview process.

X_{ijk} = the accurate value obtained for the k-th individual, in the j-th household, of the i-th cluster.

M= number of clusters in population

$\bar{N}$ = N= average number of households in a cluster (or N_i if cluster sizes are not equal)

$\bar{L}$ = L= average number of individuals in a household (or L_{ij} if household sizes are not equal).

In this case, we are not interested in estimating the components of error and are only interested in eliminating the bias.

Assuming srs at all phases of sampling with an original sample of clusters of size m, an original household sample within each cluster of size n, and corresponding subsample sizes of m_1, and n_1 and within household subsamples of size ℓ_1, we have the following:

The estimator is

$$\bar{w}_t = \bar{y}_t - \bar{y}_{st} + \bar{x}_s \quad . \tag{18}$$

$$V(\bar{w}_t) = MV + BV + TV \; . \tag{19}$$

$$TV = \frac{1}{m}\left(\frac{M-m}{M}\right)\left\{\frac{1}{M-1}\sum_{i=1}^{M} Q_i^{2}\right\} + \frac{1}{mn}\left(\frac{N-n}{N}\right)\left\{\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij}^{2}\right\}$$

$$BV = \frac{m-m_1}{mm_1}\left\{\frac{1}{M-1}\sum_{i=1}^{M} K_i^{2}\right\} + \left[\frac{1}{m_1n_1}\left(\frac{N-n_1}{N}\right) - \frac{1}{mn}\left(\frac{N-n}{N}\right)\right]\left\{\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} A_{ij}^{2}\right\} + \left[\frac{1}{m_1n_1\ell_1}\left(\frac{L-\ell_1}{L}\right)\right]\frac{1}{NM(L-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L} E_{ijk}^{2}\}$$

$$MV = \frac{m-m_1}{mm_1}\left\{\frac{1}{M}\sum_{i-1}^{M}\xi_i^{2} - \frac{1}{M(M-1)}\sum_{i\neq i'}^{M}\xi_{ij'}\right. + \left[\frac{1}{m_1n_1}\left(\frac{N-n_1}{N}\right) - \frac{1}{mn}\left(\frac{N-n}{N}\right)\right]\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\eta_{ij}^{2} + \frac{1}{m_1n_1\ell_1}\,\frac{L-\ell_1}{L}\,\frac{1}{MN(L-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L}\rho_{ijk}^{2}$$

The previous components are derived from the following model for the Y_{ijkt}:

$$Y_{ijkt} = \bar{X} + \bar{B} + Q_i + H_{ij} + L_{ijk} + K_i + A_{ij} + E_{ijk} + F_{it} + G_{ijt} + R_{ijkt} \quad .$$

Q_i = cluster effect = $\bar{X}_i - X$.

H_{ij} = household effect = $\bar{X}_{ij} - \bar{X}_i$.

L_{ijk} = individual effect = $X_{ijk} - \bar{X}_{ij}$.

K_i = bias effect for cluster i = $\bar{B}_i - \bar{B}$.

A_{ij} = bias effect for household = $\bar{B}_{ij} - \bar{B}_i$.

E_{ijk} = bias effect for individual = $B_{ijk} - \bar{B}_{ij}$.

$F_{it} = \bar{Y}_{it} - \bar{Y}_i$; $E_t\{F_{it}\} = 0$; $E_t\{F_{it}^{2}\} = \xi_i^{2}$.

$G_{ijt} = \bar{Y}_{ijt} - \bar{Y}_{ij} - F_{it}$; $E_t\{G_{ijt}\} = 0$; $E_t\{G_{ijt}^{2}\} = \eta_{ij}^{2}$.

$R_{ijkt} = Y_{ijkt} - Y_{ijk} - G_{ijt} - F_{it}$; $E_t(R_{ijkt}) = 0$;

$E_t(R_{ijkt}^{2}) = \rho_{ijk}^{2}$.

We will let the various components be denoted

$$S_Q^{2} = \frac{1}{M-1}\sum_{i=1}^{M} Q_i^{2},\quad S_H^{2} = \frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij}^{2}, \text{ etc.}$$

In order to have a cost effective design for the survey, it is necessary to determine the optimum sizes of the sample and subsample. To do this one must have estimates for the following set of variance components:

(1) $S_\rho^{2} + S_E^{2}$

(2) $S_\eta^{2} + S_A^{2}$

(3) $S_\xi^{2} - S_{\xi\xi'} + S_K^{2}$

(4) S_Q^{2}

(5) S_K^{2} .

Suppose pilot study data is available in which interview values and accurate values are available for each individual in the pilot study. Then the above components can be estimated using the following sums of square and sums of products.

Summary of expected values of sums of squares and sums of products

	SSX	SSY	SXY
Between Clusters	$(m-1)\ S_Q^2 + (\frac{m-1}{n})(\frac{N-n}{N})\ S_H^2$	$(m-1)\ \{S_\xi^2 - S_{\xi\xi'} + S_Q^2 + S_K^2 + 2S_{QK}\} + (\frac{m-1}{n})(\frac{N-n}{N})\{S_\eta^2 + S_H^2 + S_A^2 + 2S_{HA}\}$	$(m-1)\ \{S_Q^2 + S_{QK}\} + \frac{m-1}{n}(\frac{N-n}{N})\ \{S_H^2 + S_{HA}\}$
Between Households	$m(n-1)\ S_H^2$	$m(n-1)\ \{S_\eta^2 + S_H^2 + S_A^2 + 2S_{HA}\}$	$m(n-1)\ \{S_H^2 + S_{HA}\}$
Between Individuals	$mn(L-1)\ S_L^2$	$mn(L-1)\ \{S_\rho^2 + S_L^2 + S_E^2 + 2S_{LE}\}$	$mn(L-1)\ \{S_L^2 + S_{LE}\}$

OPTIMUM ALLOCATION TO THE SAMPLE AND SUBSAMPLE

In each of the above cases, the variance of the sample statistics may be written in terms of the following simple expression

$$V = \sum_{k=1}^{K} \frac{V_K^2}{\ell_K} + V_o \tag{21}$$

where V_K^2 and ℓ_K are the variance and cost components associated with the k-th design level. In addition, simple linear cost functions may be used of the form

$$C = \sum_{k=1}^{K} C_K \ell_K + D_o \quad . \tag{22}$$

Thus, optimum allocation to sample and subsample for fixed cost may be obtained using the usual solution to the above linear forms. In addition, overall multipurpose allocations may be derived using a procedure proposed by Kish in 1974 [4].

REFERENCES

[1] Hansen, M.H., Hurwitz, W.N. and Bershad, M.A., "Measurement Errors in Censuses and Surveys," Bulletin of the Intervational Statistical Institute, 38, Part II (1962), 359-74.

[2] Koch, Gary G., " An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimators Involving Subclass Means," Journal of the American Statistical Association, 68 (December 1973), 906-913.

[3] Lessler, Judith T., A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys, Institute of Statistics Mimeo Series No. 949, University of North Carolina, August 1974.

[4] Kish, Leslie, "Optimal and Proximal Multipurpose Allocation" Proceedings of the Social Statistics Section, American Statistical Association, 1974, pp 111-118.