STATISTICS IN TRANSITION-new series, October 2009
Vol. 10, No. 2, pp. 193—222

ESTIMATION OF NONSAMPLING VARIANCE COMPONENTS UNDER THE LINEAR MODEL APPROACH

Pulakesh Maiti[1]

ABSTRACT

The importance of nonsample or measurement errors has long been recognized. [for numerous references see e.g., the comprehensive papers by Mahalanobis (1946), Hansen et.al. (1961), Bailar and Dalenius (1970), Dalenius (1974)]. Attempts have been made for estimating components due to nonsampling errors. The work in this area starts developing **surveys, specifically_designed** to incorporate features which can facilitate the estimation of non sampling components such as **reinterviews and/or interpenetrating samples**. However most of the survey designs so far developed, though few, are very complex in nature [Fellegi (1964, 1974), Biemen et al. (1985), Folsom(1980), Nelson(1974)]. Here, a very simple survey design as well as a simple estimation procedure have been developed for the purpose of estimating **simple as well as correlated response variances**, namely **interviewer variance** and **superviser variance**.

Key words: Simple Response Variance, Correlated response variance, Measurement error.

1 Introduction

The various models developed for such errors have assumed that a survey record (a recorded content item) differs from **its true value** (Zarkovich 1966) by a systematic bias and various additive error contributions associated with various sources such as interviewers, supervisors, coders etc. These models indicate that the errors made by a specified error source (say a particular interviewer) are usually correlated. These correlated errors contribute to the additive components of the total mean square error of a survey estimate. As a result of these correlated components, the usual unbiased estimations of variances of estimators of total or mean appear to be negatively biased. The models also indicate that these biases can be eliminated or reduced, if estimates of correlated response variance are

[1] Indian Statistical Institute, Kolkata.

available. This necessitates the estimation of the components due to correlated variances.

1.1 Survey Measurements

We start with a set $U=\{U_1,U_2,...,U_N\}$ of N objects and a set $\{Y\}$of real numbers corresponding to the objects. Each object is assigned one and only one number and two objects may be assigned the same number [(Dalenius, (1974)].

Some of the essential conditions for having a **measurement** may be identified as follows:

(a) the value of the characteristic to be measured should be precisely defined for every population unit in a manner consonant with the users to which the data are to be put;
(b) for any given population unit, this value known as the **true value** should be unique and should exist;
(c) there should exist procedures for obtaining information on the true value and although this procedure may be costly and very difficult to use (Sukhatme and Sukhatme (1970);

1.2 Measurement Error

The approaches to define measurement error in surveys vary according to a particular researcher's view on **true values**. One approach considers the true value to exist on the survey condition, while the other takes a strict operational approach in relation to the survey condition [Hansen et al. (1951)].

Under the assumption that it is meaningful to talk about a true value Y_i of the study variable for the i^{th} population unit, then measurement error is defined as

$$(Y_i - y_i), \tag{1.1}$$

where y_i is observed measurement for the i^{th} individual using a specific measurement technique.

1.3 Three Distinct views on the Nature of Measurement Variability

Three different views in existence may be described as follows:

(a) Measurements are random variables having a mean and a finite variance [Hansen (1951); Raj (1956), (1968)]; (1.2)
(b) Measurements as random variables are generated by a conceptual sequence of repeated independent trials of a generating process [Hansen (1951)]; (1.3)

(c) A third point of view does not allow the variability at the elementary level, but assumes that the variability results from **interviewers** and **subsequent handling of data**; (1.4)

1.4 Nonsampling Bias and Nonsampling Variance

Nonsampling bias is a measure of the difference between the expected value of their repeated observations and the corresponding true value.

Nonsampling variance measures the variation of the observed values for fixed samples in **hypothetical repetitions** of the survey process, if it is agreed upon that the survey is conceptually repeatable under identical conditions. More precisely, it is assumed that a measurement derived has a well defined, though quite likely unknown, probability distribution.

Nonsampling variance has two components, namely (a) simple or uncorrelated response variance and correlated response variance.

a) **Simple response variance**: Uncorrelated responses are those that are not affected by the particular interviewer or superviser or coder or any other survey personnel who happen to be associated with a particular element of the sample.
b) **Correlated response variance**: In so far as individual interviewers have different average effects on their work loads, they introduce response errors which are correlated for all elements of the assignment included in the work load of the investigator. The correlated errors thus arising give rise to correlated response variance. The correlated response variance may be categorized as
   - interviewer variance;
   - superviser variance;
   - coder variance etc.

For different kinds of correlation of response deviation, one can refer to the paper by Fellegi (1964).

Another way of looking at correlated response variance, say, interviewer variance is that it results from bias effects that differ from one interviewer to the other.

2 General Measurement Model

A measurement model wants to specify the joint probability distribution of the measurement y_i for the i^{th} unit conditional on a samples "s" in sample surveys or for every unit i of the population U in case, a census or complete enumeration is conducted.

From a frequentist's point of view, given a particular sample selection procedure leading to a simple "s" and a specific measurement technique adopted, the process generates an observed value for every $i \in s$, and given independent observations many times on the same sample "s", a long series of data $\{y_i^t; t = 1,2,\ldots\}$ for each $i \in s$ is generated. The observed value of a specified element $i \in s$ would vary in a random fashion around a long term mean value μ_i and long term variance σ_i^2. These moments may or may not depend on the sample. The same thing applies to every $i \in U$, when complete enumeration is conducted.

2.1 Some Specific Error Models

2.1.1 Based on the views expressed in (1.1) and (1.2)

Non sampling error models are essential for understanding the effects of measurement errors on statistics and statistical inference. All such models developed assume that observed value differs from the true value by a systematic bias and additive error terms.

The basic model developed in the U.S. Bureau of census was first introduced by Hansen et al. (1951). The basic model assumes conceptually repeated trials and possesses the views on the nature of measurement variability expressed in (1.2) and (1.3). The measurement for the i^{th} unit at t^{th} trial, y_{it} was thus modeled a

$$y_{it} = Y_i + \beta_i + e_{it} \tag{2.1}$$

where β_i is a systematic bias and e_{it} is the variable error. Under repeat measurement for the same unit i, e_{it} is taken as a mean zero random error. Subsequent elaborations of the basic model was made by Hansen, Hurwitz, Bershad (1961), Hansen, Hurwitz and Pritzkar (1964).

2.1.2 Based on the views expressed in (1.3)

Let y_{ij} be an observation from a randomly selected population unit i which is thought of as the sum of two components, the true value Y_i and an error d_j; the error d_j may be attributed to the measurement processes (including the interviewer, questionnaire, the interviewer setting and so on).

In its most general form, the structure of the error d_j provides for essential correlations amongst the different measurement errors due to interviewers, supervisers, coders etc., or due to any other survey operators.

The observation collected from the i^{th} respondent by the j^{th} investigator i.e., y_{ij} may be modeled as

$$y_{ij} = Y_i + d_{ij}$$

$$y_{ij} = Y_i + \beta_j + e_{ij} \tag{2.2}$$

Y_i being the true value, β_j being the j^{th} operator bias and e_{ij}'s are elementary errors. β_j may be **fixed or random**. For random effects, β_j's constitute a random sample from an infinite population of operator effects having mean μ_b and variance σ_b^2. e_{ij}'s are random variables with mean 0, variance σ_e^2. The following covariance structure for d_{ij} and $d_{i'j'}$ may be mentioned as follows.

$$Cov\left(d_{ij}, d_{i'j'}\right) = \begin{cases} \sigma_b^2, \text{ for } j = j', i \neq i' \\ 0, \text{for } j \neq j', i \neq i \\ 0, \text{for } j \neq j', i = i' \\ \sigma_b^2 + \sigma_e^2 \text{ for } j = j', i = i' \end{cases} \tag{2.3}$$

Under the assumption of **fixed operator effects**, the covariance structure is modified by Letting $\sigma_b^2 = 0$. Another special case is the case of no-operator effects i.e., $\beta_j = 0$ for all j. This model is referred to as **uncorrelated model**.

2.2 Measurement Models taking care of a specific measurement Technique

2.2.1 Personal Interview Method

Measurements have been described as being realized under a model which specifies the joint distribution of y_i's .The model is specified in terms of its moments namely, μ_i, σ_i^2, σ_{ij}.

Introduction of the model did not require any specific measurement procedure. We now consider situations when data are collected by interviewers. They may introduce bias, variance and correlations in to the measurements (being reflected through σ_i^2, σ_{ij}). Such interviewer effects have been detected in many empirical studies.

Comprehensive and elaborate discussions on different kinds of interviewer settings and associated models taking interviewers effects into account are available in any standard text book.

3 Mean Square Error in the Presence of Combined Effects of Total Error

Decomposition and Linear model

The general models for variability in the literature are expressed either as **Mean Square Error Decomposition Model or Mixed Linear Models**. The net bias is assumed to be zero, so that the model deals only with variability. The major difference between the two models is that decomposition approach often has a component attributable to the interaction between sampling and measurement error, whereas the linear model approach omits this component. The linear model approach defines response variability about the true value. However, both the approaches merge on a specific occasion.

The Situation, When Both Variance Decomposition and Linear Model Approach Merge

The variance decomposition approach and the linear model formulation tend to merge, when the variance decomposition approach focuses on a particular source of error-mostly the error due to the interviewer. If, in the variance decomposition model with the interviewer setting, the interviewers influence in the response deviation independently, then the **additive model would be appropriate**. The additive model would not be applicable, if, in the hypothetical repition of the (original or repeat) survey, such factors as common training or supervision etc., have a correlating effect on the response deviations obtained by different enumerators/interviewers (Fellegi 1964). It has been observed by others also.

It may be mentioned that the model in (2.1) and that in (2.2) can be thought of as variance decomposion approach and linear model formulation respectively.

3.1 NonSampling Variances under: Mean Square Error Decomposition Model

Hansen, Hurwitz and Bershad Model (1961) decomposes the variance into three components namely, **(a) sampling variance, (b) measurement variance and (c) covariance between response and sampling deviation**. Their model has implications for the total survey design in so far it relates to the accuracy of survey results [Jabine and Tepping (1973)]. In our discussion we are mainly presenting the expressions for nonsampling variances.

Let $\hat{t}_\pi = \sum_{i \in S} y_i / \pi_i$ be an unbiased estimator for Y, the population total, then it may be shown that,

Sampling

$$\text{Variance(SV)} = \sum_{i \in U} \frac{(1-\pi_i)}{\pi_i} \mu_i^2 + \sum_{i \neq j} \sum_{\in U} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} \mu_i \, \mu_j \qquad (3.1A)$$

Measurement

$$\text{Variance (MV)} = \sum_{i,j} \sum_{\in U} \frac{\sigma_{ij}(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} + \sum_i \sum_j \sigma_{ij} \qquad (3.1B)$$

where,

$\mu_i, \sigma_i^2, \sigma_{ij}$ are model parameters and, (3.2)

π_i, π_{ij} the inclusion probabilities are sampling design parameters.

We have,

simple response variance $= \sum_{i \in U} \sigma_i^2 / \pi_i$ and (3.3A)

$$\text{correlated response variance} = \sum_{i \neq j} \sum_{\in U} \frac{\sigma_{ij}\, \pi_{ij}}{\pi_i \pi_j} \qquad (3.3B)$$

However, correlated response variance can alternatively be expressed as

$$\sum \sum \frac{\sigma_{ij}(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} + \sum \sum \sigma_{ij} \qquad (3.4)$$

3.2 Remarks

(1) Contrary to what generally accepted name suggests, the measurement variance depends both on measurement model and the sampling design $\left(\sigma_i^2, \sigma_j, \pi_i, \pi_{ij}\right)$. It would be therefore interesting to isolate a component that is unaffected by sampling i.e., **a term that would remain unaffected, even if the sampling were pushed to the ultimate limit of complete enumeration.**

(2) Therefore, under complete enumeration,

$$MV = \sum_{i \in U} \sigma_i^2 + \sum \sum_{i \neq j \in U} \sigma_{ij} \qquad (3.5)$$

3.2 Measurement Variances under Variance Decomposition Approach with different Measurement Models and Different Measurement Techniques

Here we present only the expressions for measurement variances under the following different situations.

3.2.1 Measurement variances without taking care of interviewer's effect:

a) Sampling Design: $SRS(N,n)$;

b) For model specification, we refer to the model specified by equation (2.1) of the basic model introduced by Hansen et al. (1951);

Let $\bar{y}_t$ be an estimator for population mean $\bar{Y}$. From equation (3.1) or otherwise also, it follows that,

$$MV(\bar{y}_t) = \frac{1}{n}\left[\sigma_d^2 + (n-1)\zeta\sigma_d^2\right], \text{ where,}$$

$$\sigma_d^2 = E_t\left(y_{it} - E_t(y_{it})\right)^2$$
$$\zeta\sigma_d^2 = E_t\left[\left(y_{it} - E_t(y_{it})\right)\left(y_{i't} - E_t(y_{i't})\right)\right] \quad (3.6)$$

This was originally derived by Hansen et al and also later by Bailar and Dalenius (1969).

3.2.2 Measurement variances with taking care of interviewer effect:

a) General Sampling Design: $[\pi:(N,n)]$;

b) For **Survey Design** with interviewer settings of deterministic as well as random assignments. We consider the following situation.

there is a fixed set of J interviewers labelled $j = 1,2,\ldots,J$, and prior to the survey, the population is partitioned in to j responding groups $U_1, U_2, \ldots, U_J$, so that each interviewer J is linked a unique or a number of groups according to the specified survey design (3.7)

For the estimator $\hat{t}_\pi$ of the population total, we have from equation (3.1A,B),

$$MV(\hat{t}_\pi) = \begin{cases} \sum_{j=1}^{J}\left(\sum_{i\in Uj} 1/\pi_i\right)v_j + \sum_{j=1}^{J}\left(\sum_{i\neq j\in Uj}\sum \pi_{ik}/\pi_i\pi_k\right)e_j v_j; & (3.8) \\ (v_\beta + v_e)\sum_{i\in U} 1/\pi_i + v_\beta \sum_{j=1}^{J}\left(\sum_{i\neq j\in Uj}\sum \pi_{ik}/\pi_i\pi_k\right), & (3.9) \end{cases}$$

where, (3.8) and (3.9) refer to deterministic and random assignments respectively and $\upsilon_j, \upsilon_\beta, \upsilon_e$ are model parameters.

3.3 Remark

Under complete enumeration, the expressions in (3.8), and (3.9) take the respective forms as

$$\sum_{j=1}^{J} N_j v_j + \sum_{j=1}^{J} N_j (N_j - 1)\ e_j v_j \tag{3.10}$$

$$v_\beta \sum_{j=1}^{J} N_j^2 + N v_e;$$

and N_j is the number of responding units in the j^{th} group U_j

3.3 A General Expression For the Measurement Variance of an Estimator taking Interviewer and Superviser Effects.

Let $y_{ijk}^{(t)}$ be a measurement made by the j^{th} investigator on the i^{th} respondent under the supervision of k^{th} superviser at the t^{th} trial $(i = 1,2,\ldots, I; j = 1,2,\ldots J \text{ and } k = 1,2,\ldots K)$;

Let us define the following indicator variables.

$$u_i = \begin{cases} 1, \text{if } i^{th} \text{ element is included into the sample;} \\ 0, \text{otherwise;} \end{cases} \tag{3.11}$$

$$v_j = \begin{cases} 1, \text{if } j^{th} \text{ interviewer is selected for the survey;} \\ 0, \text{otherwise;} \end{cases}$$

$$\delta_k = \begin{cases} 1, \text{if } j^{th} \text{ superviser is selected for supervising the job;} \\ 0, \text{otherwise;} \end{cases}$$

$$c_{ij} = \begin{cases} 1, \text{if the } i^{th} \text{ element is assigned to the } j^{th} \text{ interviewer;} \\ 0, \text{otherwise;} \end{cases}$$

$$\gamma_{k(i,j)} = \begin{cases} 1, \text{if the schedule filled up by the } j^{th} \text{ interviewer from the } i^{th} \text{ respondent} \\ \quad \text{is alloted to be } k^{th} \text{ supervisor for supervision;} \\ 0, \text{otherwise;} \end{cases}$$

Let $\bar{y}_t = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} u_i v_j c_{ij} \delta_k \gamma_{k(i,j)} / n$ be the estimator for $\bar{Y}$. (3.12)

Measurement variance, M.V. $(\bar{y}_t)$

$$= \left(\frac{1}{n}\right)^2 \sum_i \sum_j \sum_k E\left\{ u_i v_j c_{ij} \delta_k \gamma_{k(i,j)} E_t (y_{ijk}^{(t)} - y_{ijk})^2 \middle| u_i v_j c_{ij} \delta_k \gamma_{k(i,j)} = 1 \right\}$$

$$+ \sum_i \sum_{j \neq} \sum_{j'} \sum_k E\left\{ u_i v_j v_{j'} c_{ij} c_{ij'} \delta_k \gamma_{k(i,j)} \gamma_{k(i,j')} E_t \left(y_{ijk}^{(t)} - y_{ijk}\right)\left(y_{ij'k}^{(t)} - y_{ij'k}\right) \right|$$

$$\left| u_i v_j v_{j'} \delta_k c_{ij} c_{ij'} \gamma_{k(i,j)} \gamma_{k(i,j')} = 1 \right\}$$

$$+ \sum_i \sum_j \sum_{k \neq} \sum_{k'} E\left\{ u_i v_j c_{ij} \delta_k \delta_{k'} \gamma_{k(i,j)} \gamma_{k'(i,j)} E_t \left(y_{ijk}^{(t)} - y_{ijk}\right) \right.$$

$$\left. \left(y_{ijk'}^{(t)} - y_{ijk'}\right) \middle| u_i v_j \delta_k \delta_{k'} c_{ij} \gamma_{k(i,j)} \gamma_{k'(i,j)} = 1 \right\}$$

$$+ \sum_i \sum_{j \neq} \sum_{j'} \sum_{k \neq} \sum_{k'} E\left\{ u_i v_{j'} v_j c_{ij} c_{ij'} \delta_k \delta_{k'} \gamma_{k(i,j)} \gamma_{k'(i,j')} E_t (y_{ijk}^{(t)} - y_{ijk}) \right.$$

$$\left. y_{ij'k'}^{(t)} - y_{ij'k'}) \middle| u_i v_{j'} v_j c_{ij} c_{ij'} \delta_k \delta_{k'} \gamma_{k(i,j)} \gamma_{k'(i,j')} = 1 \right\}$$

$$+ \sum_{i \neq} \sum_{i'} \sum_j \sum_k E\left(u_i u_{i'} v_j c_{ij} c_{i'j} \delta_k \gamma_{k(i,j)} \gamma_{k(i',j)} \left\{ E_t \left(y_{ijk}^{(t)} - y_{ijk} \right)\left(y_{i'jk}^{(t)} - y_{i'jk} \right) \right| \right.$$

$$\left. u_i u_{i'} v_j c_{ij} c_{i'j} \delta_k \gamma_{k(i,j)} \gamma_{k(i',j)} = 1 \right\}$$

$$+\sum_{i\neq}\sum_{i'}\sum_{j\neq}\sum_{j'}\sum_{k} E\left(u_i u_{i'} v_j v_{j'} c_{ij} c_{i'j'} \delta_k \gamma_{k(i,j)} \gamma_{k(i',j')} \left\{ E_t \left(y_{ijk}^{(t)} - y_{ijk} \right)\left(y_{i'j'k}^{(t)} - y_{i'j'k} \right) \right| \right.$$

$$\left. u_i u_{i'} v_j v_{j'} c_{ij} c_{i'j'} \delta_k \gamma_{k(i,j)} \gamma_{k(i',j')} = 1 \right\} \tag{3.13}$$

$$+\sum_{i\neq}\sum_{i'}\sum_{j}\sum_{k}\sum_{k''} E\left(u_i u_{i'} v_j c_{ij} c_{i'j} \delta_k \delta_{k'} \gamma_{k(i,j)} \gamma_{k'(i',j)} \left\{ E_t \left(y_{ijk}^{(t)} - y_{ijk} \right)\left(y_{i'jk'}^{(t)} - y_{i'jk'} \right) \right| \right.$$

$$\left. u_i u_{i'} v_j c_{ij} c_{i'j} \delta_k \delta_{k'} \gamma_{k(i,j)} \gamma_{k'(i',j)} = 1 \right\}$$

$$+\sum_{i\neq}\sum_{i'}\sum_{j\neq} \sum_{j'}\sum_{k\neq}\sum_{k'} E\left(u_i u_{i'} v_j v_{j'} c_{ij} c_{i'j'} \gamma_{k(i,j)} \delta_k \delta_{k'} \gamma_{k'(i',j')} \right.$$

$$\left\{ E_t \left(y_{ijk}^{(t)} - y_{ijk} \right)\left(y_{i'j'k'}^{(t)} - y_{i'j'k'} \right) \right|$$

$$\left. u_i u_{i'} v_j v_{j'} \delta_k \delta_{k'} c_{ij} c_{i'j'} \gamma_{k(i,j)} \gamma_{k(i',j')} \gamma_{k'(i',j')}, \gamma_{k'(i,j)} = 1 \right\}$$

The above result is an extension of the similar type of result, obtained by Lessler, (1992).

3.4 Remark

1) Both under Census Bureau Model/Cochran Model and the model due to Raj (1968), simple as well as the correlated response/measurement variance of the sample mean takes the form of

$$SRV = \frac{1}{n}\frac{1}{JI}\sum_{j=1}^{J}\sum_{i=1}^{I} V_t(y_{ijt}) \tag{3.14}$$

$$\text{and } CRV = \frac{1}{n}\frac{(m-1)}{JI(1-1)}\sum_{j=1}^{J}\sum_{i\neq i'}^{I} Cov(y_{ijt}, y_{i'jt}) \tag{3.15}$$

where, m is the number of assignments for an investigator j and n is the sample size, y_{ijt} is the observation collected from the i^{th} respondent by the j^{th} investigator at the t^{th} trial and $V_t(y_{ijt}), Cov(y_{ijt}, y_{i'jt})$ are measurement variance and measurement co-variances.

2) Under Census Bureau Model, it has been assumed that a fixed number of investigators is available. No sampling of investigators is made, i.e., $\upsilon_j = 1$ for all j in our above setting.
3) Here only the covariance between the observations of a particular investigator was considered, but not that between investigators.

3.4 Linear Model in the Context of Variance Decomposition Approach

$$\text{Let } Y_{ijt} = Y_i + b_j + e_{ij} \tag{3.16}$$

where, Y_i, b_j and e_{ij} are mutually independent and that the b_j and e_{ij} arises respectively from an infinite population of interviewer effects and an infinite population of random effects. The b_j are independently distributed with $E_t(b_j) = 0$ and $V(d_j) = \sigma_b^2$; similarly, e_{ij} are independently distributed with $E_t(e_{ij}) = 0$ and $V(e_{ij}) = \sigma_e^2$.

It may be shown that

$$MV(\bar{y}_t) = \frac{(\sigma_b^2 + \sigma_e^2)}{n}[1 + (m-1)\zeta], \tag{3.17}$$

m being the number of elements assigned to an investigator and $\zeta = \sigma_b^2 / \left[\sigma_b^2 + \sigma_e^2\right]$;

It may be noted that $\left(\sigma_b^2 + \sigma_e^2\right)$ is the simple response variance and σ_b^2 is the correlated response variance.

4 Existing estimators of NonSampling Variance components

4.1 Need of Randomisation of the assignments and of repeat measurements

If the survey arrangement is such that each investigator is assigned to work in only one sample cluster, then the effect of interviewers will be completely confounded with the effect of clustering on the sample variance, and usual methods of estimating sampling variance will automatically include the correlated interviewer variance. By contrast, if all interviewers work as a team in each cluster or if the interviewer's work loads are distributed at random, the usual estimation of sampling variance would not include the effect of additional variability due to interviewers.

To make separate estimates of interviewer variance or other types of correlated variance, it is necessary to introduce some degree of randomization of interpenetration of work loads to the particular category of the survey personnel.

Repeat measurement technique has been advocated as a tool in measurement variance estimation. The measurement model using repeat measurement so far developed has one crucial assumption that all repeat measurements have to be uncorrelated with original measurements. There are several discussions on the implication of the assumption of lack of independence between the two surveys and change in the distribution of y_{it} [for example, see Hansen et al. (1964)].

4.2 Estimation of NonSampling Variance in the Variance Decomposition Approaches:

Different methods under this approach can be broadly categorized into one of the following categories.

a) Survey Design: Interpenetrated sample, but no repeat measurement: This method is primarily due to Mahalanobis (1946).

   Let $\bar{y}_{jt} = \sum_{i \in S_j} y_{ijt} \Big/ n(s_j)$, $j = 1,2,\ldots,J$ and $n(s_j) = m$;

   Let Between-interviewer mean Square(BIMS) be defined as

$$BIMS = \sum_{j=1}^{J} (\bar{y}_{jt} - \bar{y}_t)^2 /(J-1); \; \bar{y}_t = \sum_{j=1}^{J} \bar{y}_{jt} / J \qquad (4.1)$$

   Then, (*BIMS/J*), as expected, is an estimate of **total variance** i.e.,
   Thus, *E* (*BIMS/J*) = Sampling Variance + Measurement Variance.

b) **Survey Design: No interpenetration, but repeat measurements are available.**

   (b.1) Repeat measurements of the entire sample:

   Let, Mean Square within element (MSWE), Square Mean Deference (SMD) and Square of the difference Between Measures (BMWE) be defined as

$$MSWE = \frac{1}{2n} \sum_{m=1}^{2} \sum_{i=1}^{n} (y_{imt} - \bar{y}_{it})^2 \qquad (4.2)$$

$$SMD = \frac{1}{n}\left(\bar{y}_{1t} - \bar{y}_{2t}\right)^2 \tag{4.3}$$

$$BMWE = \frac{2}{n}\sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij}\left(y_{ij1t} - y_{ij2t}\right)^2 \tag{4.4}$$

It may be noted that the estimators in (4.2), (4.3) and (4.4) estimates simple response variance only.

(b. 2) Survey Design: Repeat measurements of the sub-sample

Let an original samples of n_s be drawn from a population with a sampling design $p(\cdot)$ having the inclusion probabilities π_i, π_{ij}; From S, a sub-sample of size $n_{r'}(< n_s)$ is drawn by SRSWOR. After two stages, we have the number of observations $n_s + n_{r'}$, as

$$\left\{y_i^{(1)}, i \in n(s)\right\} \text{ and } \left\{y_i^{(2)}, i \in n(r)\right\};$$

Let $Z_i = \left(y_i^{(1)} - y_i^{(2)}\right) for\ i \in S(n_{r'})$

Then, unbiased Estimate of SRV = $\left(\sum_{r} \frac{Z_i^2}{\pi_i}\right) \cdot \frac{n_s}{2n_{r'}}$, and

$$\text{unbiased Estimate of CRV} = \left(\sum_{r} \frac{Z_i Z_j}{\pi_i \pi_j}\right) \cdot \frac{n_s(n_s - 1)}{2n_{r'}(n_{r'} - 1)}; \tag{4.5}$$

It may be noted that (4.5) can be used to estimate SRV and CRV of (3.8) and (3.9).

However, in case of correlated response variance for other categories inclusive with interviewer's variance, (4.5) can not estimate separate correlated response variances due to all the categories.

c) **Survey Design: Methods that use a combination of interpenetrated samples and repeat measurements**.

Methods that use a combination of interpenetrated/ replicated samples and repeat measurements are reflected in Fellegi's work [(1964), (1974)]. Survey design developed by him with the use of both interpenetration and repeat measures is a very complex one. He, in his paper (1964) extended the model of Hansen, Hurwitz and Bershad (1961) to provide a frame work for joint application of two devices namely interpenetration and interviewer traditionally used to measure response variance. He built up some estimating equations of the parameters involved and these equations help one provide estimator for non-sampling variances, though biased. However, Fellegi (1974) came up with a relatively simple design compared to the previous one [Fellegi (1964)]. But this survey design is also not simple from the operational point of the experimental survey design.

4.3 Estimation of Non Sampling Variance in the linear Model frame work

Under linear model, we have $Y_{ijt} = Y_i + b_j + e_{ij}$ (4.7)

Let the statistic within interview Measurement Square (WIMS) be defined as

$$WIMS = \sum_{j=1}^{J} v_j \sum_{i=1}^{N} c_{ij} \left(y_{ijt} - \bar{y}_{it}\right) / \ (m-1)J,$$

υ_j, c_{ij} being defined as in (3.11), then under the assumption of fixed population of interviewers and $Y_{ij} = Y_i$; we have,

$$E(WIMS) = \sigma_e^2 + S_Y^2 \quad (4.8)$$

However, if one uses BMWE based on **repeat measurement** without the device of **interpenetration**, estimate of simple response variance, but not correlated response variance would be made possible, as one may observed that

$$E(BMWE) = \sigma_b^2 + \sigma_e^2 . \quad (4.9)$$

Hartley et al (1977) provided estimates of variance of **only elementary errors** under a two-stage sampling design using **linear model approach**. The linear model structure used by them was to capture the interviewer and Coders effect along with elementary errors. However they only estimated **variance of elementary errors**, by synthesis based method, which is a MINQUE estimate in component variance estimation problem.

It may be noted that, most of the available methods in estimating measurement variance fall under the category of **variance decomposition approach**; [For references, see papers by Koop (1974), Koch (1973), Nathan (1973), Chai (1971), Folsom (1980) etc.]. All the models developed so far are based on very complex survey designs.

Compared to estimation of non sampling variances under decomposition approach, the work of estimation following linear model formulation are not too many, except the early paper by Sukhatme and Seth (1952), followed by the work of Hartley and Rao (1978), Biemer (1978), Biemer's paper can be considered as an extension of the paper by Hartley and Rao (1978).

In the sections to follow, we have provided a simple estimation procedure for obtaining simple as well as separate correlated response variances following the linear model formulation.

5 Interactive Linear Model

Considered is the problem of simple response and correlated response variances, where a set of investigators are employed to extract responses from a set of respondents and additionally, a set supervisers are also employed to oversee the entire process and take corrective measures. Not withstanding the investigator and superviser bias, a **mixed effect model** has been developed which, in turn, has been used for the purpose of estimation of simple as well as correlated response variances, due to investigators and supervisors.

$$\text{Let } Y_{ijkt} = Y_i + b_j + S_k + e_{ijk} \tag{5.1}$$

where, Y_i, b_j, S_k being true value, investigator effect and superviser effect respectively. It is assumed that Y_i, b_j, S_k and e_{ijk} are mutually independent and that the b_j, S_k arise from an infinite population of random effects. The b_j's are independently distributed with $E(b_j) = 0$, and $V(b_j) = \sigma_b^2$; S_k'S are independently distributed with $E(s_k) = 0$ and $V(s_k) = \sigma_s^2$; Similarly, e_{ijk}'s are independently distributed with $E(e_{ijk}) = 0$ and $V(e_{ijk}) = \sigma_e^2$. In this case, there is no overall bias in the measurement process and

$$E_t\left(Y_{ijkt}\right) = Y_i \tag{5.2}$$

However, it may be noted that a bias could be introduced by letting either the expected value of b_j, s_k, e_{ijk} to be non-zero.

5.1 Interviewer Setting

We consider the same setting as in (3.7).

5.2 Survey Design with the help of a symmetric BIBD

Let J, the number of investigators be of the form $J = 4t + 3$, $t \geq 1$.

Then, we can have r, the number of responding groups to which is assigned an investigator and λ, the number of responding groups to which is assigned every pair of investigator for work of the form

$$\{J = 4t + 3; r = 2t + 1, \ \lambda = t\} \tag{5.3}$$

$$\text{and } \{J = 4t + 3; r = 2t + 2, \ \lambda = t\} \tag{5.4}$$

(complement of 5.3)

Series (5.3) and (5.4) exist and can be constructed easily be using Galois field (J), wherever, J is a prime or power of prime. The method of construction is due to R.C. Bose. However, for BIBD'S with different values of r in a given range, extensive tables are available, which may be consulted to construct the BIBDS (Raghava Rao, pp. 91–95).

The method of construction is bared on "difference sets". One will have the initial block as: $I = \left\{x^0, x^2, x^4, \ldots, x^{4t}\right\}$, where, x is a primitive root of GF (J = 4 t + 3), J being a prime or power of a prime.

5.2.1 Illustration Distribution of Seven investigators into Seven Responding Groups

Here $J = 7$ with $t = 1$ in $J = 4\ t + 3$ and $r = 3$, $\lambda = 1$; We have, $I = \left\{x^0, x^2, x^4\right\}$; Using $x = 3$, as a primitive root,

$$I = \left\{3^0, 3^2, 3^4\right\} = \{\ 1, 2, 4\ \}, \tag{5.5}$$

and the sets would be as:

$$[1, 2, 4], [2, 3, 5], [3, 4, 6], [\ 4, 5, 7], [5, 6, 1], [6, 7, 2], [7, 1, 3] \tag{5.6}$$

Table 5.1. Distribution of Interviewer Assignment into Responding Groups

Responding Groups Investigators	U_1	U_2	U_3	U_4	U_5	U_6	U_7
1	✔	✔		✔			
2		✔	✔		✔		
3			✔	✔		✔	
4				✔	✔		✔
5	✔				✔	✔	
6		✔				✔	✔
7	✔		✔				✔

5.2.2 Distribution of Investigator work to Supervisers

Let there be two supervisers S_1 and S_2 ; The investigators work are assigned randomly in to two supervisers according to the following lay out.

Table 5.2. Distribution of Investigator's work into two Supervisers

Investigators Supervisers	1	2	3	4	5	6	7
S_1	✔	✔	✔	✔			
S_2				✔	✔	✔	✔

5.3 Estimation of Simple and Correlated Response Variances

$$\left(\sigma_b^2, \sigma_s^2, \sigma_e^2 \text{ and } \sigma_b^2 + \sigma_s^2 + \sigma_e^2\right)$$

For simplicity of calculation. It has been assumed that each responding unit has only one respondent. However our method of estimation would remain unchanged, even if every $U_j\,(j = 1,2,\ldots,7)$ has more than one respondents. In that case, only the size of the data matrix would be larger.

5.3.1 Acquisition and modeling of Data

Following the methods of data collection and of supervision (Ref. Tables 5.1, 5.2) we would have 24 recorded and supervised values. Thus we shall have,

$$\underset{\sim}{Y}_{24\times1} = \left\{y_{ijk};\, i = 1,2,\ldots 7;\, j = 1,2,\ldots 7 \text{ and } k = 1,2\right\}$$

For example, data after collection and after supervision, we would have four observations from each of the respondents namely 4^{th}, 5^{th} and 7^{th}. In fact data from the 5^{th} respondent would read as follows:

$$\begin{pmatrix} y_{521} \\ y_{541} \\ y_{542} \\ y_{552} \end{pmatrix} = \begin{pmatrix} y_5 + b_2 + s_1 + e_{521} \\ y_5 + b_4 + s_1 + e_{541} \\ y_5 + b_4 + s_2 + e_{542} \\ y_5 + b_5 + s_2 + e_{552} \end{pmatrix} \qquad (5.7)$$

Similarly, there will be 3 observations each from 1^{st}, 2^{nd}, 3^{rd} and 6^{th} respondent, thus totaling to 24 observations.

5.3.2 Canonical Reduction of the Data

Theorem 5.2.1. The $\underset{\sim}{Y}$ can be partitioned as $\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim 7\times1}{U} \\ \underset{\sim 17\times1}{V} \end{pmatrix}$ with the dispersion matrix $\sum_{24\times24}$ which can be partitioned also as $\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$, where, $\sum_{11}$ and $\sum_{22}$ are the variance. Covariance matrices of $\underset{\sim}{U}$ and $\underline{V}$.

$$\underset{\sim}{U} = \begin{bmatrix} \frac{1}{\sqrt{3}}(Y_{111} + Y_{152} + Y_{172}) \\ \frac{1}{\sqrt{3}}(Y_{211} + Y_{221} + Y_{262}) \\ \frac{1}{\sqrt{3}}(Y_{321} + Y_{331} + Y_{372}) \\ \frac{1}{\sqrt{4}}(Y_{411} + Y_{431} + Y_{441} + Y_{442}) \\ \frac{1}{\sqrt{4}}(Y_{521} + Y_{541} + Y_{542} + Y_{552}) \\ \frac{1}{\sqrt{3}}(Y_{631} + Y_{652} + Y_{662}) \\ \frac{1}{\sqrt{4}}(Y_{741} + Y_{742} + Y_{762} + Y_{772}) \end{bmatrix};$$

$$\underset{\sim}{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} y_{111} - \frac{1}{\sqrt{2}} y_{152} \\ \frac{1}{\sqrt{6}} y_{111} + \frac{1}{\sqrt{6}} y_{152} - \frac{2}{\sqrt{6}} y_{172} \\ \frac{1}{\sqrt{2}} y_{211} - \frac{1}{\sqrt{2}} y_{221} \\ \frac{1}{\sqrt{6}} y_{211} + \frac{1}{\sqrt{6}} y_{221} - \frac{2}{\sqrt{6}} y_{262} \\ \frac{1}{\sqrt{2}} y_{321} - \frac{1}{\sqrt{2}} y_{331} \\ \frac{1}{\sqrt{6}} y_{321} + \frac{1}{\sqrt{6}} y_{331} - \frac{1}{\sqrt{6}} y_{372} \\ \frac{1}{\sqrt{2}} y_{411} - \frac{1}{\sqrt{2}} y_{431} \\ \frac{1}{\sqrt{6}} y_{411} + \frac{1}{\sqrt{6}} y_{431} - \frac{2}{\sqrt{6}} y_{441} \\ \frac{1}{\sqrt{12}} y_{411} + \frac{1}{\sqrt{12}} y_{431} + \frac{1}{\sqrt{12}} y_{441} - \frac{3}{\sqrt{12}} y_{422} \\ \frac{1}{\sqrt{2}} y_{521} - \frac{1}{\sqrt{2}} y_{541} \\ \frac{1}{\sqrt{6}} y_{521} + \frac{1}{\sqrt{6}} y_{541} - \frac{2}{\sqrt{6}} y_{542} \\ \frac{1}{\sqrt{12}} y_{521} + \frac{1}{\sqrt{12}} y_{541} + \frac{1}{\sqrt{12}} y_{542} - \frac{3}{\sqrt{12}} y_{552} \\ \frac{1}{\sqrt{2}} y_{631} - \frac{1}{\sqrt{2}} y_{652} \\ \frac{1}{\sqrt{6}} y_{631} + \frac{1}{\sqrt{6}} y_{652} - \frac{2}{\sqrt{6}} y_{662} \\ \frac{1}{\sqrt{2}} y_{741} - \frac{1}{\sqrt{2}} y_{742} \\ \frac{1}{\sqrt{6}} y_{741} + \frac{1}{\sqrt{6}} y_{742} - \frac{2}{\sqrt{6}} y_{762} \\ \frac{1}{\sqrt{12}} y_{741} + \frac{1}{\sqrt{12}} y_{742} + \frac{1}{\sqrt{12}} y_{762} - \frac{3}{\sqrt{12}} y_{772} \end{bmatrix}$$

Proof: The reduction follows immediately by Helmet transformation.

It may be shown that, $\sum_{11}$ Dispersion matrix of $\underset{\sim}{U}$ and $\sum_{22}$ the dispersion matrix of $\underset{\sim}{V}$ can be represented as

$$\sum_{11} = A_1' \sigma_b^2 + A_2' \sigma_s^2 + I \sigma_e^2$$

and (5.8)

$$\sum_{22} = A_1' \sigma_b^2 + A_2' \sigma_s^2 + I \sigma_e^2$$

where, A_1, A_2, A_2', A_2' are all symmetric matrices of real numbers. After calculation, the matrices A_1, A_2, A_2', A_2' in $\sum_{11}$ and $\sum_{22}$ have been found to be as given in the tables (5.3), (5.4) & (5.5).

5.4 Estimation of σ_b^2, σ_s^2 and σ_e^2

5.4.1 Estimation of σ_e^2

We have, $Y_{ijk} = Y_i + b_j + S_k + e_{ijk}$;

$$(i = 1,2,\ldots,I;\ j = 1,2,\ldots,J \text{ and } k = 1,2,\ldots,k)$$

$$= \bar{Y} + (Y_i - \bar{Y}) + b_j + S_k + e_{ijk} \quad (5.9)$$

$$= \bar{Y} + \alpha_i + b_j + S_k + e_{ijk} \quad \text{with } \Sigma\alpha_i = 0.$$

Let $SSE = \sum_i \sum_j \sum_k \left[y_{ijk} - y_{i\ldots} - y_{.j.} - y_{..k} + 2y_{\ldots}\right]^2$, then it can be shown that

$$E(SSE) = \left[IJK - \{(I-1) + (J-1) + (K-1)\} - 1\right]\sigma_e^2$$

$$\text{i.e., } E\left\{\left[SSE/\left[IJK - \{(I-1) + (J-1) + (K-1)\} - 1\right]\right]\right\} = \sigma_e^2$$

$$\text{Therefore, } \hat{\sigma}_e^2 = \left[SSE/\left[IJK - \{I-1) + (J-1) + (K-1)\} - 1\right] \quad (5.10)$$

5.4.2 Estimation of σ_e^2

Let ε_1 and λ_1 be the eigen vector corresponding to maximum eigen value λ_1 of the matrix A_1. Now, from (5.8), we have,

$$\varepsilon_1' \hat{\sum}_{22} \varepsilon_1 = \varepsilon_1' A_1 \varepsilon_1 \hat{\sigma}_b^2 + \varepsilon_1' A_2 \varepsilon_1 \hat{\sigma}_b^2 + \sigma_e^2 \qquad (5.11)$$

After calculaton, all eigen values appeared to be non negative, as they are expected. On computation, maximum eigen value λ_1 becomes 3.5 and $\varepsilon_1' A_2 \varepsilon_1$ becomes $-$ 7.85704 $e-005$ which is almost 0.

Thus, from (5.11), we have

$$\varepsilon_1' \hat{\sum}_{22} \varepsilon_1 - \hat{\sigma}_e^2 = 3.5 \hat{\sigma}_b^2 - .00007857047 \hat{\sigma}_s^2,$$

Thus, $$\hat{\sigma}_b^2 = \left[\varepsilon_1' \hat{\sum}_{22} \varepsilon_1 - \hat{\sigma}_e^2 \right] \Big/ 3.5 \qquad (5.12)$$

Table. 5.3 A_1. the Coefficient Matrix of σ_b^2 in $\sum 22$

1	0	0.5	.28867513	0	0	0.5	.28867513	-.204124	0	0	.612372	0.5	.288675	0	0	0
	1	.28867513	.166666	0	0	.28867513	.16666	.11785113	0	0	-.3535539	-.28867513	.16666	0	0	.70710678
		1	0	-0.5	-28867513	0.5	.28867513	.2041241659	-0.5	-.2886751	-.20412416	0	0	0	0	0
			1	.28867513	.166666	.28867513	.166666	.1178511309	.28867513	.1666	.11785133	0	.6666	0	.6666	-.23570226
				1	0	0.5	-.28867513	-.204124165	0.5	.28867513	.204124165	-0.5	-.28867513	0	0	0
					1	-.28867513	.16666	.17785113019	.28867513	.16666	.1178511309	.28867513	.16666	0	0	.7071067
						1	0	0	0	0	0	-0.5	-.28867513	0	0	0
							1	.70710678	.57735	.33333	-.47140452	.28867513	.1666	0	.66666	.471405
								0.5	.408248333	.2357023	-.333333	.204124166	.11785113	0	-.471405	-.16666
									1	.57735026	-204124165	0	0	0	-.5773503	-.4082483
										.33333	-.11785113	0	0	0	-.33333	-.2357023
											1.16666	.61237249	-.35355339	0	.4714045	.33333
												1	0	0	0	0
													1	0	.666666	-.23570226
														0	0	0
															1.33333	.235702
																1.16666

Table. 5.4 A_2, the Coefficient Matrix of σ_s^2 in $\sum 22$

1	.5773502	0	1.154705	0	1.154705	0	0	1.224745	0	1.154705	.8164966	1	.57735	1	.57735026	4082483
	.33333	0	.666666	0	.66666	0	0	.70710678	0	.6666	.4714045	.57735026	.33333	.57735026	.33333	.2357024
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			1.33333	0	1.3333	0	0	1.4142135	0	1.3333	.94280904	1.1547005	.66666	1.1547005	.666666	.4714045
				0	0	0	0	0	0	0	0	0	0	0	0	0
					1.33333	0	0	1.41421356	0	1.3333	.942809	1.1547005	.66666	1.1547005	.66666	.4714045
						0	0	0	0	0	0	0	0	0	0	0
							0	0	0	0	0	0	0	0	0	0
								1.5	0	1.41421356	1	1.2247449	.70710678	1.2247449	.707106678	0.5
									0	0	0	0	0	0	0	0
										1.33333	.942809	1.1547005	.66666	1.1547005	.66666	.47140452
											.666666	.8164966	.4714045	.8164966	.4140452	.333333
												1	.57735026	1	.57735026	.40824833
													.333333	.57735026	.33333	.23570266
														1	.57735026	.40824833
															.333333	.23570226
																.1666666

Table. 5.5. Coefficient Matrix of σ_b^2, σ_s^2 in $\sum_{11}$

$A_1' =$

1	.33333	.33333	.28867513	.28867513	.33333	.28867513
	1	.33333	.28867513	.28867513	.33333	.28867513
		1.5	.28867513	.28867513	.33333	.28867513
			1.5	1	.28867513	1
				1.5	.28867513	1
					1	.28867513
						1.5

$A_2' =$

1.66666	1.33333	1.33333	1.44337	1.732051	1.66666	2.02073
	1.66666	1.66666	2.02073	1.732051	1.33333	1.443375
		1.66666	2.02073	1.732051	1.33333	1.443375
			2.5	2	1.443375	1.5
				2	1.732051	2
					1.66666	2.02073
						2.5

5.4.3 Estimation of σ_s^2:

Method – I

Simple response variance, i.e., $\left(\sigma_b^2+\sigma_s^2+\sigma_e^2\right)$ can be estimated from the repeat measurements. This can be obtained through our survey design.

Hence, $\hat{\sigma}_S^2=\left(\sigma_b^2+\sigma_s^2+\sigma_e^2\right)-\hat{\sigma}_b^2-\hat{\sigma}_e^2$ (5.13)

Method – II

Let $\underset{\sim}{\varepsilon_2}$ and λ_2 be the eigen vector corresponding to maximum eigen value λ_2 of the matrix A_2. All the eigen values appeared to be non-negative as they are expected and the maximum eigen value is found to be 1.03332.

Now, from (5.8), we have ,

$$\varepsilon_2' \hat{\sum}_{22} \varepsilon_2 = \varepsilon_2' A_1 \varepsilon_2 \hat{\sigma}_b^2 + \varepsilon_2' A_2 \varepsilon_2 \hat{\sigma}_b^2 + \hat{\sigma}_e^2 \tag{5.14}$$

$$\varepsilon_2' \hat{\sum}_{22} \varepsilon -_2 \hat{\sigma}_e^2 = 1.650562751\, \hat{\sigma}_b^2 + 1.03332 \hat{\sigma}_s^2$$

Now, from (5.10), (5.12) and (5.13), $\hat{\sigma}_S^2$ can be obtained.

6 Estimation of Simple and Correlated Response variances associated with measurement process in estimating population total

Let y_{ijkt} be the observation collected from the i^{th} respondent by the j^{th} investigator and supervised by the k^{th} superviser at t^{th} trial following our survey design (Ref. Tables 5.1, 5.2). The investigator's and superviser's assignment rule has been defined earlier in section 5 and under the assumption of the model parameters,

$$\text{we have, } E_t\left(y_{ijkt}\right) = Y_i, \tag{6.1}$$

using the notation similar to those in (3.11),

$$\text{Let } C_{ij} = \begin{cases} 1, \text{ if } j^{th} \text{ investigator is assigned to the } i^{th} \text{ respondent} \\ 0, \text{ otherwise} \end{cases}$$

$$\text{and } C_{ij(k)} = \begin{cases} 1, \text{if } k^{th} \text{ superviser superviser the job to } j^{th} \text{ investigator} \\ 0, \text{otherwise} \end{cases}$$

$$\text{Let } \hat{Y} = \sum_i \sum_j \sum_k C_{ij}\ C_{ijk}\ y_{ijkt}$$

Then we have the following

Theorem 6.1.: $\hat{Y}$ is unbiased with simple response $\left(\sigma_b^2 + \sigma_s^2 + \sigma_e^2\right)$ and correlated response variances $\sigma_b^2, \sigma_s^2, \sigma_e^2$ (6.2)

Proof: $E(\hat{Y}) = E_R\ E_{t|R}\ (\hat{Y})$,

where t denotes the measurement variable and R stands for the random allocation of the investigators to the respondents. Therefore,

$$E(\hat{Y}) = \sum_{i}\sum_{j}\sum_{k} E_{t|R}\left(y_{ijkt} \middle| C_{ij}\ C_{ij(k)} = 1\right) E_R\left(C_{ij}\ C_{ij(k)} = 1\right)$$

$$= \sum_{i}\sum_{j}\sum_{k} Y_i \frac{1}{JK} = \sum_{i=1}^{I} Y_i \sum_{j}\sum_{k} \frac{1}{JK},$$

Since, $E\left(C_{ij} = 1\right) = \frac{1}{J}$ and $E\left(C_{ij(k)} = 1 \middle| C_{ij} = 1\right) = \frac{1}{K}$

Therefore, $E(\hat{Y}) = \sum_{i} Y_i = Y$ (6.3)

Now, measurement variance

$$\text{MV}(\hat{Y}) = E\left(\sum_{i}\sum_{j}\sum_{k} C_{ij}\, C_{ij(k)}\left(y_{ijkt} - E_t\left(y_{ijkt}\right)\right)\right)^2$$

Proceeding same as before in (3.13), it may be shown that the corresponding terms in (3.13) under this situation can be found to be

1st term = $I\left(\sigma_b^2 + \sigma_s^2 + \sigma_e^2\right)$

2nd term = $I\sigma_S^2$

3rd term = $I \cdot \sigma_S^2$

4th term = 0 (6.4)

5th term = $I(I-1)\left(\sigma_b^2 + \sigma_s^2\right)$

6th term = $I(I-1)\,\sigma_s^2$

7th term = $I(I-1)\,\sigma_b^2$

8th term = 0

Combining all the terms of (6.4), we have,

$$MV(\hat{Y}) = I^2\left(2\sigma_b^2 + 2\sigma_s^2 + \frac{\sigma_e^2}{I}\right) \qquad (6.5)$$

Let, $\zeta_1^* = \dfrac{\sigma_b^2}{\sigma_b^2 + \sigma_s^2 + \sigma_e^2}$ (6.6)

$$\zeta_2^* = \frac{\sigma_s^2}{\sigma_b^2 + \sigma_s^2 + \sigma_e^2}$$

and $\zeta_3^* = \dfrac{\sigma_e^2}{\sigma_b^2 + \sigma_s^2 + \sigma_e^2}$

then, $MV(\hat{Y}) = I^2 \left(\sigma_b^2 + \sigma_s^2 + \sigma_e^2\right) \left[2\varsigma_1^* + 2\varsigma_2^* + \dfrac{\varsigma_e^*}{I}\right]$ (6.7)

and $MV(\bar{Y}) = 2 \left(\sigma_b^2 + \sigma_s^2 + \sigma_e^2\right) \left(\varsigma_1^* + \varsigma_2^*\right) + \dfrac{\sigma_e^2}{I}$;

The form of the expressions in (6.7) are similar to that of Hansen, Hurwitz and Bershad (1961).

Following the methodology already discussed in the section in the section 5, $\sigma_b^2, \sigma_s^2, \sigma_e^2$ can be estimated.

7 Some Concluding Remarks and Discussions

1. Symmetric BIBD was needed to be constructed for this specific interview setting, where the number of responding groups into which the population was partitioned was equal to the number of interviewers. However, the design need not necessarily be always that of a symmetric BIBD. Depending on the interviewer setting, it could be non-symmetric BIBD also. In fact, intensive tables are available for construction of symmetric and non-symmetric BIBD for any specied value of J.
2. The proposed survey design is very simple to construct and to operate compared to the earlier methods, where both **interpenetration** and **repeat measurement techniques** have been used [(Fellegi, (1964), (1974); Beimer (1985)]. In fact, from the user point of view, he can simply consult the available tables for constructing the BIBDS for allocation of the interviewer assignment, without having in depth knowledge on Galois field etc. Our proposed method is a very user friendly one.
3. The method of estimation procedure also involves easy computation. It requires calculation of only eigen values and eigen vectors and for that standard computer package is readily available. The method is very flexible also in the sense that it can accommodate any number of effects arising from different categories of survey personnel from survey management group.
4. For illustration purpose, we have taken J = 7 and number of respondents from each U_j to be one. However, our estimation procedure would remain

unchanged, even if we have more number of investigators and more number of respondents from each responding group. Only the size of the data matrix will be larger, but the estimation procedure would remain unaltered.

REFERENCES

BAILAR, BARBARA A. and TORE DALENIUS (1969): Estimating the Response Variance Components of the U.S. Bureau of the Census Survey Model, Sankhȳa , 31B, 341–360.

BIEMER, PAUL P. and S. LYNNE STokes (1985): Optimal Design of Interviewer Variance Experiments in Complex Surveys, JASA, 80, 158–166.

CHAI, JOHN J (1971): Correlated measurement errors and the least Square Estimators of the Regression Coefficient, JASA, 66, 478–483.

DALENIUS TORE (1974): The Ends and Means of Total Survey Design. Stockholm: The University of Stockholm.

FELLEGI, IVAN P. (1964): Response variation and estimation, JASA, 59, 1016–1041.

------------------- (1974): An Improved method of Estimating the Correlated Response Variance, JASA, 69, 496–501.

FOLSOM, RALPH E. Jr. (1980): U-Statistics estimation of Variance Components for unequal probability samples with non additive Interviewer and response errors, American Statistical Association 1980 proceedings of the Survey Research Methods Section, 137–142.

HANSEN, MORRIS H. WILLIAM N. HURWITZ, ETES. MARKS, and W. PARKER MAULDIN (1951): Response Errors in Survey, JASA, 46, 147–190.

---------------------------- and MAX A. BERSHAD (1961): Measurement Errors in Censuses and Surveys, Bulletin of the International Statistical Institute, 38, 359–374.

HANSEN, MORRIS H, WILLIAM N. HURWITZ and LEAN PRITZKER (1964): The Estimation and Interpretation of Gross differences and the simple response variance: In C. R. Rao with D.B. Lahiri, K.R. Nair, P. Pant and S.S. Shrikhande eds. Contributions to Statistics Presented to Professor P.C. Mahalanobis on the Occasion of his 70th Birthday, Oxford, England: Pergamen Calcutta Statistical Publishing society, 111–136.

HARTLEY, H.O. and RAO, J.N.K. (1978): The Estimation of Non sampling Variance Components in Sample Surveys, N. Krishnan Namboodiri, ed., Survey Sampling and Measurement, New York; Academic.

JABINE, J.B. and B.J. TEPPING (1973): Controlling the quality of occupation and industry data. Bulletin of the International Statistical Institute, 45 (3), 36–389.

KOOP, J.C. (1974): Notes for a unified theory of Estimation for Sample Surveys taking into account Response Errors, Metrika 21, 19-39.

KOCH GORY G. (1973): An alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to estimators involving Subclass Means, JASA, 68, 906 – 913.

LESSLER, JUDITH T. and KALSBEEK, WILLIAM, D. (1992): Non sampling Errors in Surveys. John Wiley and Sons. Inc.

MAHALANOBIS, P.C. (1946): Recent Experiments in Statistical Sampling in the Indian Statistical Institute, JRSS, 109, 327-328.

NATHAN, GAD (1973): Response Errors Based on Different Samples, Sankhya, 35 A, 205–220.

NELSON, FORREST D. (1977): Censored Regression Models with unobserved Stochastic Censoring Thresholds, Journal of Econometrics, 6, 581 592.

RAGHAVA RAO, D. (1971): Constructions and Combinatorial Problems in Design of Experiments. John Wiley and Sons. Inc.

RAJ DES (1956): Some Estimators in sampling with varying probabilities without replacement, JASA, 51, 269–284.

RAJ DES (1968): Sampling Theory, New York: Mc Graw Hill.

SUKHATME, PANDURANG VASUDEO and G.R. SETH (1952): Non sampling Errors in Surveys, Journal of the Indian Society of Agricultural Statistics, 4, 5–41.