INTERNATIONAL JOURNAL OF NATURAL AND APPLIED SCIENCES (IJNAS) VOL. 1, No. 1: pp. 90 - 96.

Response variance in double sampling for inclusion probabilities

Godwin A. Udofia[1]

ABSTRACT

The mathematical model developed by Hansen, Hurwitz and Bershad, for response errors under simple random sample design is extended to double sampling for inclusion probabilities. Under the double sample design and a linear response model, the total response variance is decomposed into components that reflect the different sources of response error. The effect of variation of the number of interviewers on the contribution of the simple response variances due to the various sources of response error is highlighted. A justification for proper identification of factors that involve correlated response deviations before the start of the survey and the need to determine the most efficient combination of the number of interviewers and the numbers of respondents are also given.

Keywords: Components of response variance, double sampling, inclusion probabilities, response bias, sample estimates.

INTRODUCTION

In recent studies of response variation in surveys, interest is centred on the total response variance, and on the relative sizes of the simple response variances and on the correlated components of the total response variance that reflect important sources of variation of the survey responses. After the pioneer work by Mahalanobis (1946), a number of methods for the study of response variation have been developed and discussed by survey statisticians. Notable among these are Hansen, Hurwitz, Marks and Mauldin (1951), Hansen, Hurwitz and Bershad (1961), Fellegi (1964), Tepping and Bolan (1973), Koch (1973) and Talukder (1975).

The paper by Hansen, Hurwitz, Marks and Mauldin (1951) states that where alternative methods of measurement of response exit, each with a different level of response bias, a combination of two of the methods in a double sampling design may prove more efficient than the use of a single method. The paper then gives an illustration of the use of double sampling with simple random sampling without replacement and under different essential survey conditions in both phases of the survey. Talukder (1975) also discusses the application of double sampling with simple random sampling without replacement in both phases to the study of response errors of surveys but under a modified linear response model.

It has been amply demonstrated [Raj (1954, 1958, 1968), Cochran (1977), Foreman and Brewer (1971)] that under certain conditions, sampling with probability proportional to size (πPS) gives a more precise result than equal probability sampling. In many such situations, the auxiliary information needed to compute the inclusion probabilities is not available. It is of interest therefore to see what the theory of response error measurement and control under linear response model would look like if method of double sampling for inclusion probabilities proportional to size were used to combine two methods of measurement. This paper is an attempt in this direction.

Following earlier studies referenced above, the following assumptions are made. There exists a population, π_1, of N respondents (which can be elementary units or clusters of units such as households), which fall into H identifiable groups, which may be geographic groupings, professional groupings, or different types of dwelling houses, for example;

1. There are N_h, $\sum_h N_h = N$, respondents in the h^{th} group or subset, G_h.
2. There exists also a population, π_2, of M interviewers who can be divided into H corresponding groups with M_h, $\sum_h M_h$ = M, interviewers in the h^{th} group or subset, Q_h, such that respondents within the h^{th} group (G_h) in π_1 can be interviewed only by interviewers in the h^{th} group (Q_h) within π_2.
3. The number of individuals in G_h that are available to be interviewed by the i^{th} individual in Q_h is $\overline{N}_h = \left[\frac{N_h}{M_h}\right]$
4. A survey can be repeated k times, k ≥ 1, [Hansen *et al.*, (1951), Raj (1968)] and all the repetitions may relate to the same time or to different periods in time.

Let X_{hijk} denote response obtained by interviewer i on unit j in stratum h in the k – th survey. Then X_{hijk} is a random variable (only one of the

Manuscript received by the Editor July 23, 2004; Revised manuscript accepted May 16, 2005.
[1]Department of Mathematics, Statistics and Computer Science, University of Calabar, P.M.B. 1115, Calabar, Cross River State, NIGERIA.

possible response, which could be obtained from the j^{th} unit).

ESSENTIAL SURVEY CONDITIONS

As noted by Hansen, Hurwitz and Bershad (1961), measurement errors may arise from different types of causes, and depend on the conditions under which the survey is taken. Some of these conditions such as the general political, economic and social situation at the time of the survey or rumours may be beyond the control or specification of the survey designer. Other conditions can be controlled so as to improve the quality of the survey results. These conditions are known as the "essential" survey conditions.

In general, the essential survey conditions are specified either explicitly or implicitly by the survey design. Those conditions that are usually explicitly specified by the survey design include subject of inquiry, method of obtaining information (interview, mail inquiry, direct observation), method of recording the information, and period of the survey. The other conditions that arise implicitly as a necessary consequence of the explicitly specified conditions include recall problem, condition of the labour market and the type of interviewer that can be obtained, compensation that can be paid to the interviewers, wording of the questions to be asked and sponsorship.

Hansen, Hurwitz, Marks and Mauldin (1951) conclude that the expected value of the response errors, and the random component of variation around that expected value, may be regarded as determined by the essential survey conditions. For the purpose of determining components of the response variance in double sampling we therefore consider the survey responses to be obtained under one of the following essential survey conditions as suggested by Hansen, Hurwitz and Madow (1953):

Essential Condition A:

These are the general specifications typical of large scale surveys such as personnel with secondary or high school certificate or below, low compensation rate, minimum training of field staff, low publicity, little or no probing, and little supervision.

Essential Condition B:

This is a set of more expensive and more efficient essential survey conditions than those in condition A above. These conditions include interviewers with higher qualifications, more experience and in more respectable positions in the society as well as higher compensation rate than in condition A.

Other specifications include highly effective interviewer training, probing (which may involve examination of records or use of available aids to memory), closer supervision of the interviewers, better controls than those under condition A, proper timing of the field operation, smaller assignment to each field staff and other devices that can produce better information than what can be obtained under condition A.

DOUBLE SAMPLING FOR INCLUSION PROBABILITIES AND MEASUREMENT OF RESPONSE BIAS.

We now consider the situation where the auxiliary information, (Z), free from response error, that is needed to compute the inclusion probabilities is not available. Hence we adopt a double sampling design in which n_q and m_q are fixed where n_q is the number of sampling units to be drawn from π_1 during the q^{th} phase and m_q is the number of interviewers to be drawn from π_2 during the q^{th} phase, q = 1, 2. The number of units n_{qh}, $n_q = \sum_h n_{qh}$, to be drawn during the q^{th} phase from G_h is determined as in stratified random sampling.

An initial sample, S_{1h}, of n_{1h} units is drawn by simple random sampling without replacement and independently from the h^{th} group (G_h) in π_1 out of which $\bar{n}_{1h} = \left[\frac{n_{1h}}{m_{1h}}\right]$, $n_1 = \sum_h \bar{n}_{1h} m_{1h}$, units are assigned at random to each interviewer in a sample of m_{1h} interviewers drawn at random and independently from the h^{th} group (Q_h) in π_2. The response, X_{hijk}, which is a value of the variable under condition A which is a set of general specifications or fixed rules typical of large scale surveys and described in detail by Hansen, Hurwitz, Marks and Mauldin (1951) is obtained in the k^{th} survey from the j^{th} unit in S_{1h} by the i^{th} interviewer drawn at random from Q_h. The value Z_{hj} of an auxiliary variable Z needed to compute the inclusion probabilities is also obtained for the j^{th} unit in S_{1h}. It is assumed that Z_{hj} can be obtained precisely by observation or otherwise. The case where the contrary is true can also be investigated and the results obtained in this paper can provide a useful reference for such a study.

From the n_{1h} units drawn from G_h during the first phase, we take a random sample of a fixed size, n_{2h}, n_{2h}, < n_{1h}, with replacement and by probability, $p_{hj} = Z_{hj} / Z_{1h}, Z_{1h} = \sum_{j=1}^{n_{1h}} Z_{hj}$. A second sample of m_{2h}, $m_{2h} < m_{1h}$, interviewers is also drawn at random from the initial sample of m_{1h} interviewers and $\bar{n}_{2h} = \left[\frac{n_{2h}}{m_{2h}}\right]$, $n_2 = \sum_h \bar{n}_{2h} m_{2h}$, interviewees are assigned at random to each of them. The discussion in this paper also applies to a situation where the second sample of m_{2h} interviewers is drawn from a population, π_3, say, that is different from π_2.

The i^{th}, i = 1, 2, …, m_{2h}, interviewer in the second – phase sample obtains information, y_{hij}, from the j^{th}, j = 1, 2, …, $\bar{n}_{2h}$, interviewee in a given assignment under a set, B, of more expensive and more efficient essential survey conditions than the set of general essential survey conditions that include highly effective interviewer training, probing (which may involve examination of records or careful measurement) and supervision of the interviewers and can produce accurate information. The information, y_{hij}, is the desired true value of the study variable which is independent of the interviewer and period of the survey but may differ from one individual respondent to another in π_1 i.e. a value free from all survey errors. Hence we drop i as a subscript and use y_{hj} instead of y_{hij}.

Notations:

For any W_{hijk}, $E_{hij}(W_{hijk}) = E(W_{hijk}|h,i,j)$ and in general

$Ea(W_{hijk}) = E(W_{hijk}|a)$ where a is any combination of the subscripts h,i,j,k.

Thus:

$$\bar{w}_{hij} = E_{hij}(w_{hijk}); \bar{w}_{hi}, = E_{hi}(w_{hijk}); \bar{w}_{h...} = E_h(w_{hijk})$$

The response model

We assume that X_{hijk} can be represented by the linear model [(Hansen *et al*. (1951), Raj (1968), Talukder (1975)]:

$$x_{hijk} = y_{hij} + \alpha_{hi} + \beta_{hj} + d_{hijk},$$

where; α_{hi} denotes bias from the i[th] interviewer in group h, β_{hj} denotes bias from respondent j in the i[th] interviewer assignment in stratum h and d_{hijk} denotes random response error for the j[th] respondent in group h interviewed by interviewer i on the k-th survey,

$E_{hij}(d_{hijk}) = 0$,

$V(d_{hijk}|h_{ij}) = \sigma^2(d)$, $\text{cov}(d_{hijk}, d_{hlwk}) = 0 =$

$= \text{cov}(d_{hijk}, d_{hljk})$; $i \neq \ell, j \neq w$

where: d_{hlwk} denotes random response error for the w[th] respondent, in group h interviewed by interviewer l, on the k[th] survey, and there is no correlation between possible pairs from $\{y_{hij}, \alpha_{hi}, \beta_{hj}, d_{hijk}\}$.

Under the above model, response bias is defined as

$$B = \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{j=1}^{\bar{N}_h} E_{hij}(X_{hijk} - Y_{hj}) = \sum_{h=1}^{H}\bar{N}_h\sum_{i=1}^{M_h}\alpha_{hi} + \sum_{h=1}^{H}M_h\sum_{j=1}^{\bar{N}_h}\beta_{hj}.$$

This is the algebraic sum of all biases [Kish (1965), p 518].

This can be written as

$$B = X - Y = \sum_{h=1}^{H} N_h(\bar{X}_{h...} - \bar{Y}_{h...}) \qquad \text{…(1)}$$

ESTIMATOR OF RESPONSE BIAS AND ITS SAMPLING VARIANCE

Under the above sample design, an unbiased estimator of response bias in (1) is given by

$$\hat{B} = \sum_{h=1}^{H}\frac{N_h}{n_{1h}n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{(x_{hijk} - y_{hj})}{P_{hj}}$$

where $\hat{B}$ denotes an estimator of B defined in equation (1) above.

By using theorems on conditional expectations and variances [Raj (1956, 1968)], the expected value of $\hat{B}$ is obtained as

$$E_2(\hat{B}|n_1) = E_2\sum_{h=1}^{H}\frac{N_h}{n_{1h}n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}E_{hij}\frac{(x_{hijk} - y_{hj})}{P_{hj}}$$

$$= \sum_{h=1}^{H}N_h(\bar{x}_{1h} - \bar{y}_{1h});\ \bar{x}_{1h} = \tfrac{1}{n_{1h}}\sum_{i=1}^{m_{1h}}\sum_{i=1}^{\bar{n}_{1h}}E_{hij}(x_{hijk}),\ \bar{y}_{1h} = \tfrac{1}{n_{1h}}\sum_{i=1}^{m_{1h}}\sum_{i=1}^{\bar{n}_{1h}}y_{hj}$$

and hence

$$E(\hat{B}) = E\{E(\hat{B}|n_1)\} = \sum_{h=1}^{H}N_h(\bar{X}_{h...} - \bar{Y}_{h...}) = X - Y = B$$

as in (1) since $E(\bar{x}_{1h}) = \bar{X}_{h...}$ and $E(\bar{y}_{1h}) = \bar{Y}_h$.

The conditional variance formula for $\hat{B}$ is

$$V(\hat{B}) = V_1E_2(\hat{B}) + E_1V_2(\hat{B}) \qquad \text{…(2)}$$

Now, $$V_1E_2(\hat{B}) = \sum_{h=1}^{H}N_h^2\left(\frac{1}{n_{1h}} - \frac{1}{N_h}\right)(S^2(x)_h + S^2(y)_h - 2S(xy)_h) \qquad \text{…(3)}$$

which can be written as:

$$V_1E_2(\hat{B}) = \sum_{h=1}^{H}N_h^2\left(\tfrac{1}{n_{1h}} - \tfrac{1}{N_h}\right)\tfrac{1}{M_h}\sum_{i=1}^{M_h}\left(1 - \tfrac{1}{\bar{N}_h}\right)\left\{S^2(x)_{hi} + S^2(y)_{hi} - 2\rho(xy)_{hi}S(x)_{hi}S(y)_{hi}\right\} \qquad \text{...(4)}$$

since

$$S^2(x)_h = \tfrac{1}{M_h\bar{N}_h}\sum_{i=1}^{M_h}\sum_{j=1}^{\bar{N}_h}E_{hij}(x_{hijk} - \bar{X}_{h...})^2$$

$$= \tfrac{1}{M_h}\sum_{i=1}^{M_h}\tfrac{1}{\bar{N}_h}\sum_{j=1}^{\bar{N}_h}E_{hij}(x_{hijk} - \bar{X}_{h...})^2 = \tfrac{1}{M_h}\sum_{i=1}^{M_h}\left(1 - \tfrac{1}{\bar{N}_h}\right)S^2(x)_{hi}.$$

where: $\left(1 - \tfrac{1}{\bar{N}_h}\right)S^2(x)_{hi} = \tfrac{1}{\bar{N}_h}\sum_{j=1}^{\bar{N}_h}E_{hij}(X_{hijk} - \bar{X}_{h...})^2$ and $S^2(y)_{hi}$ is defined similarly as $S^2(x)_{hi}$.

In a similar manner we define the covariance of X and Y for the subset of $\bar{N}_h$ individuals in group G_h that can be interviewed by the i[th] individual in group Q_h, $G_h \neq Q_h$, as;

$$\left(1 - \tfrac{1}{\bar{N}_h}\right)S(xy)_{hi} = \tfrac{1}{\bar{N}_h}\sum_{j=1}^{\bar{N}_h}E_{hij}(X_{hijk} - \bar{X}_{h...})(Y_{hj} - \bar{Y}_{h...}) = \left(1 - \tfrac{1}{\bar{N}_h}\right)\rho(xy)_{hi}S(x)_{hi}S(y)_{hi}$$

where; $\rho(xy)_{hi} = \frac{S(xy)_{hi}}{S(x)_{hi}S(y)_{hi}}$

$$E_1V_2(\hat{B}) = E_1V_2\left(\sum_{h=1}^{H}\frac{N_h}{n_{1h}n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk} - y_{hj}}{P_{hj}}\right)$$

$$= \sum_{h=1}^{H}\frac{N_h^2}{n_{1h}^2}\Bigg[E_1V_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk}}{P_{hj}}\right) + E_1V_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{y_{hj}}{P_{hj}}\right)$$

$$- 2E_1C_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk}}{P_{hj}}, \frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{y_{hj}}{P_{hj}},\right)\Bigg] \qquad \text{…(5)}$$

Now;

$$V_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk}}{P_{hj}}\right) = \frac{1}{n_{2h}}\sum_{i=1}^{m_{1h}}\sum_{j=1}^{\bar{n}_{1h}}P_{hi}E_{hij}\left(\frac{x_{hijk}}{P_{hj}} - X_{1h}\right)^2; X_{1h} = \sum_i\sum_j\sum_k X_{hijk}$$

Following Raj (1968) and Hansen *et al*. (1951), this can be calculated as:

$$V_2\left(\frac{1}{2}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk}}{P_{hj}}\right) = \frac{1}{n_{2h}}\sum_{i=1}^{m_{1h}}\sum_{i=1}^{\bar{n}_{1h}}\sum_{\ell>1}^{\bar{n}_{1h}}Z_{hj}Z_{h\ell}E_{hij,\ell}\left(\frac{X_{hijk}}{Z_{hj}} - \frac{X_{hi\ell k}}{Z_{h\ell}}\right)^2$$

By considering the $\bar{n}_{1h}$ units in an interviewer allotment as a random sample from N_h elements in stratum h, h = 1, 2, …, H, we obtain

$$E_1V_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hijk}}{P_{hj}}\right)=\frac{1}{n_{2h}}\frac{\bar{n}_{1h}(n_{1h}-1)}{N_h(N_h-1)}\sum_{i=1}^{M_h}\sum_{i=1}^{N_h}\sum_{\ell>j}^{N_h}Z_{hj}Z_{h\ell}E_{hij,\ell}\left(\frac{Z_{hijk}}{Z_{hj}}-\frac{Z_{hi\ell w}}{Z_{h\ell}}\right)^2$$

$$=\frac{1}{n_{2h}}\frac{\bar{n}_{1h}(\bar{n}_{1h}-1)}{N_h(N_h-1)}\sum_{i=1}^{M_h}\sigma_p^2(x)_{hi}$$

where; $\sigma^2(x)_{hi}=V(x)_{hi}=\sum_{j=1}^{N_h}\sum_{\ell=1}^{N_h}Z_{hj}Z_{h\ell}E_{hij,\ell}\left(\frac{X_{hijk}}{Z_{hj}}-\frac{X_{hi\ell w}}{Z_{h\ell}}\right)^2$

Similarly,

$$E_1V_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{y_{hj}}{P_{hj}}\right)=\frac{1}{n_{2h}}\frac{\bar{n}_{1h}(\bar{n}_{1h}-1)}{Nh(Nh-1)}\sum_{i=1}^{M_h}\sigma_p^2(y)_{hi} \quad \ldots(7)$$

Now from Tripathi (1973) :

$$C_p(x,y)=C_2\left(\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{x_{hij}}{P_{hj}},\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{y_{hj}}{P_{hj}}\right)=\frac{1}{n_{2h}}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}P_{hj}E_{hij}\left(\frac{X_{hijk}}{P_{hj}}-X_{1h}\right)\left(\frac{y_{hj}}{P_{hj}}-Y_{1h}\right)$$

$$=\frac{1}{n_{2h}^2}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{\bar{n}_{2h}}E_{hij}\left(\frac{x_{hijk}\,y_{hj}Z_h}{Z_{hj}}-x_{hijk}y_{hj}\right)$$

and hence,

$$E_1C_p(x,y)=\frac{n_{1h}(n_{1h}-1)}{n_{2h}N_h(N_h-1)}\sum_{i=1}^{M_h}\left[\sum_{j=1}^{Nh}E_{hij}(x_{hijk}y_{hj}/Z_{hj}(\sum_j Z_{hj})-\sum_{j=1}^{Nh}E_{hij}(x_{hijk})\sum_{j=1}^{Nh}E_{hij}(y_{hj})\right]$$

$$=\frac{n_h}{m_{1h}N_h}\frac{(n_{1h}-1)}{(N_h-1)}\sum_{i=1}^{M_h}Cov_p(x,y)_{hi}$$

$$=\frac{n_{1h}}{n_{2h}N_h}\frac{(n_{1h}-1)}{(N_h-1)}\sum_{i=1}^{M_h}\rho_p(x,y)_{hi}\sigma_p(x)_{hi}\sigma_p(y)_{hi} \quad \ldots(8)$$

where; $\rho_p(x,y)_{hi}=\frac{Cov_p(x,y)_{hi}}{\sigma_p(x)_{hi}\sigma_p(y)_{hi}}$

Substitution of (6), (7) and (8) in (5) gives the result

$$E_1V_2(\hat{B})=\sum_{h=1}^{H}\frac{N_h}{N_h-1}\frac{n_{1h}-1}{m_{1h}n_{2h}}\sum_{i=1}^{M_h}\left\{\sigma_p^2(x)_{hi}+\sigma_p^2(y)_{hi}-2\rho_p(xy)_{hi}\sigma_p(x)_{hi}\sigma_p(y)_{hi}\right\} \ldots (9)$$

which is in terms of response variances and covariances within the i^{th}, i=1, 2, …, M_h, interviewer assignment.

Substitution of (4) and (9) in (2) gives the total response variance for the sampling strategy as

$$V(\hat{B})=\sum_{h=1}^{H}N_h^2\left(\frac{1}{n_{1h}}-\frac{1}{N_h}\right)\frac{1}{M_h}\sum_{i=1}^{M_h}\left\{S^2(x)_{hi}+S^2(y)_{hi}-2\rho(xy)_{hi}S(x)_{hi}S(y)_{hi}\right\}+$$

$$+\sum_{h=1}^{H}\frac{Nh}{N_h-1}\frac{\bar{n}_{1h}-1}{n_{1h}n_{1h}n_{2h}}\sum_{i=1}^{M_h}\left\{\sigma_p^2(x)_{hi}+\sigma_p^2(y)_{hi}-2\rho_p(xy)_{hi}\sigma_p(x)_{hi}\sigma(y)_{hi}\right\} \quad \ldots(10)$$

A large positive value of $\rho_p(xy)$ will reduce the response variance i.e. the closer the response is to the true value, the smaller the response variance can be. A study of how such positive correlation occurs is therefore important. If $S^2(x)_{hi}=S^2(y)_{hi}$, $\rho(xy)_{hi} > 0$ and $\rho_p(xy)_{hi} > 0$ for all i, equation 10 reduces to;

$$V(\hat{B})=\sum_{h=1}^{H}N_h^2\left(\frac{1}{n_{1h}}-\frac{1}{N_h}\right)\frac{1}{M_h}\sum_{i=1}^{M_h}2S^2(x)_{hi}(1-\rho(x,y)_{hi})+\sum_{i=1}^{H}\frac{N_h}{N_h-1}\frac{n_{1h}-1}{m_{1h}n_{2h}}\sum_{i=1}^{M_h}2\sigma^2(x)_{hi}(1-\rho_p(x,y)_{hi})$$

It is obvious therefore that a concentration of a reasonable proportion of the survey resources on the reduction of the survey errors through a careful application of condition B can lead to a high gain in precision.

DECOMPOSITION OF TOTAL RESPONSE VARIANCE INTO SIMPLE RESPONSE VARIANCES

An essential aspect of the study of response variation is the decomposition of the total response variance into simple response variances that reflect the main sources of variation of the observed responses. Now

$$\bar{X}_{hij}=Y_{hj}+\alpha_{hi}+\beta_{hj},\bar{X}_{hi}=\bar{Y}_{h.}+\alpha_{hi}+\bar{\beta}_{h.}$$

and $\bar{X}_{h...}=\bar{Y}_{h.}+\bar{\alpha}_{h.}+\bar{\beta}_{h.}$.

Hence, $X_{hijk}-\bar{X}_{h...}=y_{hj}+r_{hi}+b_{hj}+d_{hijk}$ is the total response deviation for individual j within stratum h from the mean of all responses in that stratum, where

$$y_{hj}=Y_{hj}-\bar{Y}_{h...},r_{hi}=\alpha_{hi}-\bar{\alpha}_{h.},b_{hj}=\beta_{hj}-\bar{\beta}_{h.}$$

Following Hansen, Hurwitz and Bershad (1961) we define the following:

$X_{hijk}-\bar{X}_{hij.}=d_{hijk}$ is the random response error for individual j in the i^{th} interviewer allotment in stratum h.

$\bar{X}_{hij}-\bar{X}_{hi..}=y_{hj}+b_{hj}$ is the sampling deviation and

$\bar{X}_{hi..}-\bar{X}_{h...}=r_{hi}$ is the bias deviation in stratum h for interviewer i.

Given the above notations, $S^2(x)_h$ in (3) can be expressed in terms of the component simple response variances as

$$S^2(x)_h=\frac{1}{M_h\bar{N}_h}\sum_{i=1}^{N_h}\sum_{j=1}^{\bar{N}_h}E_{hij}(X_{hijk}-\bar{X}_{h...})^2=\frac{1}{M_h\bar{N}_h}\sum_{i=1}^{M_h}\sum_{j}^{\bar{N}_h}\left[(X_{hijk}-\bar{X}_{hij})+(\bar{X}_{hij.}-\bar{X}_{hi..})+(\bar{X}_{hi...}-\bar{X}_{h...})\right]^2$$

$$=\frac{1}{M_h}\sum_{i=1}^{M_h}\frac{1}{\bar{N}_h}\sum_{j=1}^{N_h}E_{hij}\left(d_{hijk}+y_{hj}+b_{hj}+r_{hi}\right)^2$$

$$=\frac{1}{M_h}\sum_{i=1}^{M_h}\left[\left(1-\frac{1}{\bar{N}_h}\right)S^2(d)_{hi}+\left(1-\frac{1}{\bar{N}_h}\right)S^2(y)_{hi}+\left(1-\frac{1}{\bar{N}_h}\right)S^2(\beta)_{hi}\right]+\left(1-\frac{1}{M_h}\right)S^2(\alpha)_h \quad \ldots(11)$$

where; $S^2(d)_{hi}=\frac{1}{\bar{N}_h-1}\sum_{j=1}^{\bar{N}_h}E_{hij}(d_{hijk}^2);S^2(y)_{hi}=\frac{1}{\bar{N}_h-1}\sum_{j=1}^{\bar{N}_h}y_{hj}^2$

$$S^2(b)_{hi}=\frac{1}{\bar{N}_h-1}\sum_{j=1}^{\bar{N}_h}b_{hj}^2;S^2\bar{(\alpha)}_h=\frac{1}{M_h-1}\sum_{i=1}^{M_h}r_{hi}^2$$

The cross-product terms vanish because of lack of association between different components of response deviation under the model. In a similar manner, we obtain

$$S^2(y)_h=\frac{1}{M_h}\sum_{i=1}^{M_h}\frac{N_h-1}{N_h}S^2(y)_{hi} \quad \ldots(12)$$

and $$S^2(xy)_h = \frac{1}{M_h}\sum_{i=1}^{M_h}\frac{\bar{N}_h-1}{N_h}S^2(y)_{hi} \qquad \ldots(13)$$

Substitution of (11), (12) and (13) in (3) gives the result

$$V_1E_2(\hat{B}) = \sum_{h=1}^{H}N_h^2\left(\frac{1}{n_{1h}}-\frac{1}{N_h}\right)\left\{\left(1-\frac{1}{\bar{N}_h}\right)\left[S^2(d)_{hi}+S^2(\beta)_{hi}\right]+\left(\frac{M_h-1}{M_h}\right)S^2(a)_h\right\} \qquad \ldots(14)$$

The second term on the right-hand-side of (2) is obtained by noting that

$$E_1V_2(\hat{B}) = \sum_{h=1}^{H}\frac{N_h^2}{n_{1h}^2n_{2h}^2}\left\{E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{x_{hijk}}{P_{hj}}\right)+E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)\right.$$
$$\left.-2\,E_1C_2\left(\sum_{h=1}^{H}\frac{N_h^2}{n_{1h}^2n_{2h}^2}\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{x_{hijk}}{P_{hj}},\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)\right\} \qquad \ldots(15)$$

All cross-product terms for h ≠ k vanish because of independence of selection within each stratum. In terms of the component notations in the above response model,

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{x_{hijk}}{P_{hj}}\right)=E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)+E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{r_{hi}}{P_{hj}}\right)+E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{b_{hi}}{P_{hj}}\right)+E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{d_{hijk}}{P_{hj}}\right) \ldots 16)$$

Now

$$E_1V_2(\hat{B})=\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hij}}{P_{hj}}\right)=n_{2h}E_1V_2\left(\frac{y_{hj}}{P_{hj}}\right)+n_{2h}(n_{2h}-1)E_1C_2\left(\frac{y_{hj}}{P_{hj}},\frac{y_{hl}}{P_{h\ell}}\right)\ldots (17)$$

$$V_2\left(\frac{y_{hj}}{P_{hj}}\right)=\sum_{i=1}^{m_{1h}}\sum_{j=1}^{n_{1h}}P_{hj}\left(\frac{y_{hj}}{P_{hj}}-y_h\right)^2=\sum_{i=1}^{m_{1h}}\sum_{j=1}^{n_{1h}}\sum_{j>\ell}^{n_{1h}}Z_{hj}Z_{h\ell}\left(\frac{y_{ij}}{Z_{hj}}-\frac{y_{hl}}{Z_{h\ell}}\right)^2,$$

and hence

$$E_1V_2\left(\frac{y_{hij}}{P_{hj}}\right)=\sum_{i=1}^{M_h}\sum_{j=1}^{N_h}\sum_{j>\ell}^{N_h}\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}Z_{hj}Z_{h\ell}\left(\frac{y_{hj}}{P_{hj}}-\frac{y_{hl}}{P_{h\ell}}\right)^2=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}\sum_{i=1}^{M_h}\sigma_p^2(y)_{hi} \qquad \ldots(18)$$

By equation (8)

$$E_1C_2\left(\frac{y_{hj}}{P_{hj}},\frac{y_{hl}}{P_{h\ell}}\right)$$
$$=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}\sum_{i=1}^{M_h}\rho_p(y_{hj},y_{h\ell})_{hi}\sigma_p^2(y_{hj})_{hi};\sigma_p(y_{hl})_{hi} \qquad \ldots(19)$$

where; $\rho_p(y_{hij},y_{hil})_{hi}=\dfrac{\operatorname{cov}_p(y_{hij},y_{hil})_{hi}}{\sigma_p(y_{hij})_{hi}}$

since $\sigma_p(y_{hj})_{hi} = \sigma_p(y_{hl})_{hi}$.

Substitution of (18) and (19) in (17) gives the result

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{i=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}n_{2h}\sum_{i=1}^{Mh}\left[1+(n_{2h}-1)\rho_p(y_{hj},y_{h\ell})_{hi}\right]\sigma_p^2(y)_{hi} \qquad \ldots(20)$$

Similarly,

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{i=1}^{n_{2h}}\frac{d_{hijk}}{P_{hj}}\right)=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}n_{2h}\sum_{i=1}^{M_h}\sigma_p^2(d)_{hi}\left[1+(n_{2h}-1)\rho_p(d_{hijk},d_{hi\ell k})_{hi}\right] \qquad \ldots(21)$$

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{i=1}^{n_{2h}}\frac{b_{hj}}{P_{hj}}\right)=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}n_{2h}\sum_{i=1}^{M_h}\sigma_p^2(\beta)_{hi}\left[1+(n_{2h}-1)\rho_p(\beta_{hj},\beta_\ell)_{hl}\right] \qquad \ldots(22)$$

We note that since the interviewers were not selected with πPS, $V_2\left(\sum_{j=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{r_{hi}}{P_{hj}}\right)$ reduces to

$$E_2(\alpha_{hi}-\bar{\alpha}_{h.})^2=V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\alpha_{hi}\right)=V_2\left(n_{2h}\sum_{i=1}^{m_{2h}}\alpha_{hi}\right)=n_{2h}^2\sum_{i=1}^{m_{2h}}V_2(\alpha_{hi})=n_{2h}n_{2h}V_2(\alpha_{hi})$$

since $\operatorname{cov}(\alpha_{hi},\alpha_{hk}) = 0$ for all i≠k under the assumption of independence of the different interviewers. However, it is possible for different interviewers working under the same supervisor to influence the bias of one another but such a situation is usually rare under efficient supervision and is therefore ignored. The above equation therefore reduces to

$$V_2\left(\sum_{j=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{\alpha_{hi}}{P_{hj}}\right)\alpha_{hi} = n_{2h}n_{2h}\frac{1}{m_{2h}-1}\sum_{i=1}^{m_{2h}}(\alpha_{hi}-\bar{\alpha}_{h.})_h^2 \text{ and hence}$$

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\alpha_{hi}\right)=n_{2h}n_{2h}\frac{1}{M_h-1}\sum_{i=1}^{m_{2h}}(\alpha_{hi}-\bar{\alpha}_{h.})_h^2=n_{2h}n_{2h}\sigma^2(\alpha)_h \qquad \ldots(23)$$

Substitution of (20), (21), (22) and (23) in (16) gives the result

$$E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{i=1}^{n_{2h}}\frac{x_{hijk}}{P_{hj}}\right)=\frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)}n_{2h}\sum_{i=j}^{M_h}\left\{\sigma_p^2(d)_{hi}\left[1+(n_{2h}-1)\rho_p(d_{hijk},d_{hi\ell k})_{hi}\right]+\right.$$
$$\left.\sigma^2(y)_{hi}\left[1+(n_{2h}-1)\rho_p(y_{hj}y_{h\ell})_{hi}\right]+\sigma_p^2(\beta)_{hi}\left[1+(n_{2h}-1)\rho_p(\beta_{hj};\beta_{h\ell})_{hi}\right]\right\}$$
$$+\;n_{2h}\bar{n}_{2h}\sigma^2(\alpha)_h \qquad \ldots(24)$$

Finally the covariance term in (15) is obtained in terms of the component factors of X_{hijk} as

$$C_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}+\alpha_{hi}+\beta_{hij}+d_{hijk}}{P_{hj}},\sum_{i=j}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)=V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right) \text{ so that}$$

$$E_1C_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{x_{hijk}}{P_{hj}},\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right)=E_1V_2\left(\sum_{i=1}^{m_{2h}}\sum_{j=1}^{n_{2h}}\frac{y_{hj}}{P_{hj}}\right) \ldots(25) \text{ as in (20).}$$

By substituting (20), (24) and (25) in (15) we obtain

$$E_1V_2(\hat{B})=\sum_{h=1}^{H}\frac{N_h}{N_h-1}\frac{n_{1h}-1}{m_{1h}n_{1h}n_{2h}}\sum_{i=1}^{M_h}\left\{\sigma_p^2(d)_{hi}\left[1+(n_{2h}-1)\rho_p(d_{hjk},d_{hi\ell k})_{hi}\right]+\right.$$
$$\left.\sigma_p^2(\beta)_{hi}[1+(n_{2h}-1)\rho_p(\beta_{hj},\beta_{i\ell})_{hi}\right\}+\sum_{h=1}^{H}\left(\frac{N_h}{n_{1h}}\right)^2\frac{\sigma^2(\alpha)_h}{m_{2h}} \qquad \ldots(26)$$

Finally, we substitute (14) and (26) in (2) and obtain, after rearrangement

$$V(\hat{B})=\sum_{h=1}^{H}N_h^2\left(\frac{1}{n_{1h}}-\frac{1}{N_h}\right)\frac{1}{M_h}\sum_{i=1}^{M_h}\left(1-\frac{1}{N_h}\right)\left[S^2(d)_{hi}+S^2(\beta)_{hi}\right]+$$
$$+\sum_{h=1}^{H}\frac{N_h}{N_h-1}\frac{n_{1h}-1}{m_{1h}n_{1h}n_{2h}}\sum_{i=1}^{M_h}\left\{\sigma_p^2(d)_{hi}\left[1+(n_{2h}-1)\rho_p(d_{hijk},d_{hi\ell k})_{hi}\right]+\sigma_p^2(\beta)_{hi}\left[1+(n_{2h}-1)\rho_p(\beta_{hj},\beta_{hi})_{hi}\right]\right\}+$$
$$+\sum_{h=1}^{H}\left(\frac{N_h}{n_{1h}}\right)^2\frac{\sigma^2(\alpha)_h}{m_{2h}} \qquad \ldots(27)$$

Since N_h is fixed, the last term which is the contribution from interviewer bias reduces as n_{1h} and m_{2h} increase.

ESTIMATE OF SAMPLING VARIANCE

Different ways of designing experiments to obtain approximate estimates of the response variance or of specified component simple response variances have been discussed in the references given in Section 1 of this paper. An estimate of the error term, d_{hijk}, as defined above, can be obtained by method of repeated measurement (interviewing) of the subsample of $\bar{n}_{2h}$ units drawn for the second phase of the survey from each of the initial interviewer assignment by another interviewer of the same quality as the initial interviewers and under the same essential condition A. To avoid additional training cost and alteration of survey condition, the $\bar{n}_{2h}$ units drawn from one initial interviewer assignment can be allocated at random to another interviewer under a different supervisor or leader.

Let X_{hij1} and X_{hij2} denote the two values of response (or measurement) from the jth unit of the ith interviewer assignment that fall in both the subsample of $\bar{n}_{1h}$ units and the subsample of $\bar{n}_{2h}$, units. Then an estimate of V $(\hat{B})$ in (27) is obtained from sample values as follows:

$$\hat{V}(\hat{B}) = \sum_{h=1}^{H} N_h^2\left(\tfrac{1}{n_{1h}} - \tfrac{1}{n_h}\right)\tfrac{1}{m_{1h}}\sum_{h=1}^{H}\left(1-\tfrac{1}{n_{1h}}\right)\left[\hat{S}^2(d)_{hi} + \hat{S}^2(\beta)_{hi}\right] +$$

$$+\sum_{h=1}^{H} \tfrac{1}{N_h-1}\,\tfrac{n_{1h}-1}{m_{1h}n_{1h}n_{2h}}\sum_{i=1}^{m_{2h}}\left\{\hat{\sigma}_p(d)_{hi}\left[1+(\bar{n}_{2h}-1)\hat{\rho}(d_{hijk}, d_{hi\ell k})\right] +\right.$$

$$\left.\hat{\sigma}_p^2(\beta)_{hi}[1+(\bar{n}_{2h}-1)\hat{\rho}(\beta_{hij}, \beta_{i\ell})_{hi}\right\} + \sum_{h=1}^{H}\left(\tfrac{N_h}{n_{1h}}\right)^2 \tfrac{\hat{\sigma}^2(\alpha)_h}{m_{2h}}$$

where; $\hat{S}^2(d)_{hi} = \frac{1}{\bar{n}_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\sum_{k=1}^{2}(d_{hijk} - \bar{d}_{2hj.})^2 = \frac{1}{\bar{n}_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\sum_{k=1}^{2} d_{hijk}^2$,

since $\bar{d}_{hj.} = 0$.

$$\hat{\sigma}_p^2(d)_{hi} = \frac{Z_{1h}^2}{2\bar{n}_{2h}(\bar{n}_{2h}-1)}\left[\sum_{j=1}^{\bar{n}_{2h}}\sum_{k=1}^{2}\frac{d_{hijk}^2}{Z_{hj}^2} - \frac{1}{2\bar{n}_{2h}}\left(\sum_{j=1}^{\bar{n}_{2h}}\sum_{k=1}^{2}\frac{d_{hijk}}{Z_{hj}}\right)^2\right],$$

$$\hat{\rho}_p(d_{hijk}, d_{hi\ell k})_{hi} = \frac{\hat{C}_p(d_{hijk}, d_{hi\ell w})_{hi}}{\sigma_p^2(d)_{hi}}$$

where

$$\hat{C}_p(d_{hijk}, d_{hi\ell w})_{hi} = \frac{1}{2\bar{n}_{2h}(\bar{n}_{2h}-1)}\sum_{j=1}^{\bar{n}_{2h}}\sum_{\ell\neq j}^{\bar{n}_{2h}} Z_{hj}Z_{h\ell}\left(\sum_{k=1}^{2}\frac{d_{hijk}}{Z_{hj}} - \frac{1}{2\bar{n}_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\sum_{k=1}^{2}\frac{d_{hijk}}{Z_{hj}}\right)\left(\sum_{w=1}^{2}\frac{d_{hi\ell k}}{Z_{h\ell}} - \frac{1}{2\bar{n}_{2h}}\sum_{l=1}^{\bar{n}_{2h}}\sum_{k=1}^{2}\frac{d_{hilwk}}{Z_{hl}}\right)_{hi}$$

$$\hat{b}_{hj} = \hat{\beta}_{hj} - \hat{\hat{\beta}}_{h.} = (\bar{X}_{hij} - \bar{X}_{hi}) - (Y_{hj} - \bar{Y}_h),$$

$$\hat{\sigma}_p^2(\beta)_{hi} = \frac{Z_{1h}^2}{2\bar{n}_{2h}(\bar{n}_{2h}-1)}\left[\sum_{j=1}^{\bar{n}_{2h}}\frac{b_{hj}^2}{Z_{hj}^2} - \frac{1}{\bar{n}_{2h}}\left(\sum_{j=1}^{\bar{n}_{2h}}\frac{b_{hj}^2}{Z_{hj}^2}\right)^2\right]_{hi};\ \hat{\rho}_p(b_{hj}, b_{h\ell}) = \frac{\hat{C}_p(b_{hj}, b_{hl})_{hi}}{\hat{\sigma}_p^2(\beta)_{hi}}$$

and

$$\hat{C}_p(\beta_{hj}, \beta_{hl})_{hi} = \frac{1}{2\bar{n}_{2h}(\bar{n}_{2h}-1)}\sum_{j=1}^{\bar{n}_{2h}}\sum_{\ell\neq j}^{\bar{n}_{2h}} Z_{hj}Z_{h\ell}\left(\frac{b_{hj}}{Z_{hj}} - \frac{1}{2\bar{n}_{2h}}\sum_{j=1}^{\bar{n}_{2h}}\frac{b_{hj}}{Z_{hj}}\right)\left(\frac{b_{hl}}{Z_{hj}} - \frac{1}{\bar{n}_{2h}}\sum_{l=1}^{\bar{n}_{2h}}\frac{b_{hl}}{Z_{hl}}\right)$$

and $\hat{\sigma}^2(\alpha)_{hi} = \frac{1}{m_{2h}-1}\sum_{i=1}^{m_{2h}}(\bar{x}_{2hi.} - \bar{x}_{2h...})^2$.

DISCUSSION

The components of the total response variance in (27) are the simple response variances $\sigma^2(\beta)$, $\sigma^2(\alpha)$, $\sigma^2(d)$ due respectively to response bias, interviewer bias, random response error, and the simple response variances due to the correlated factors represented by those terms that involve $(\bar{n}_{2h} - 1)\rho$. The significance of the contribution from correlated factors is extensively discussed by Hansen *et al.* (1961). Nevertheless, the fact that under the sampling strategy discussed here the correlated factors are multiplied by $\bar{n}_{2h} - 1$, a factor which depends on the average size of the interviewer assignment in the second phase only, is important.

The third term on the right-hand-side of (27) measures variation among the different interviewers, i.e. absence of uniformity among the interviewers, and indicates the need for uniformity in terms of qualification and experience of the interviewers as well as close supervision of the interviewers. An increase in m_{1h} tends to result in the reduction in the contribution from interviewer bias. For fixed n_{2h}, an increase in m_{2h} leads also to a reduction in $1/(n_{1h}\ m_{2h})$ and to a reduction in $\bar{n}_{2h} - 1$, a factor that multiplies contribution from the correlated factors. An increase in n_{1h} reduces the first term on the right-hand-side of (27) that is affected by bias of the respondents. The conclusion is that the interviewers assignments in both phases of the survey should be kept at a minimum.

REFERENCES

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons, New York.

Fellegi, J. P. (1964). Response variation and its estimation. *J. Amer. Stat. Assoc.*, 59: 1016 – 1041.

Foreman, E. K. and K. W. R. (1971). The efficient use of supplementary information in standard sampling procedures. *J. Roy. Stat. Soc.*, B. 333: 391 – 400.

Hansen, M.H., W. N. Hurwitz and W. G. Madow (1953). *Sampling survey methods and theory*. John Wiley & Sons, New York.

Hansen, M.H., W.N. Hurwitz and M.A. Bershad (1961). Measurement error in censuses and surveys. *Bull. Int. Stat. Inst.*, 38: 359 – 374.

Hansen, M. H. ; W. N. Hurwitz, E. S. Marks and W. F. Mauldin (1951). Response errors in surveys. *J. Amer. Stat. Assoc.*, 46:147-190.

Kish, L. (1962). Studies of Interviewer Variability for attitudinal Variables. *J. Amer. Stat. Assoc.*, 57:92 – 115.

Koch, G. G. (1973). An alternative approach to multivariate response error models for sample survey data with applications to estimators involving subclass means. *J. Amer. Stat. Assoc.*, 68:906 – 918.

Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the Indian statistical institute. *J. Roy. Stat. Soc.*, 109: 325 – 380.

Raj, D. (1954). On sampling with probabilities proportional to size. *Ganita*, 5: 175 – 182.

______ (1958): On the relative accuracy of some sampling techniques. *J. Amer. Stat. Assoc.*, 53:98 – 101.

______ (1965). Response bias in farmers' reports. *Sankhyā* B: 279 – 282.

______ (1968). *Sampling theory*. McGraw – Hill Books Co., New York.

Srikantan, K. S. (1977). Evaluation of social surveys: A case study of the Turkish social survey 1968. *Sankhyā*, 6. 89: 95 – 127.

Tripathi, T. P. (1973). Double sampling for inclusion probabilities and regression method of estimation. *J. Ind. Stat. Assoc.*, 10: 33 – 46.