Survey Methodology, June 1986
Vol. 12, No. 1, pp. 29–36
Statistics Canada

Some Optimality Results in the Presence of Nonresponse

V.P. GODAMBE and M.E. THOMPSON[1]

ABSTRACT

Using the optimal estimating functions for survey sampling estimation (Godambe and Thompson 1986), we obtain some optimality results for nonresponse situations in survey sampling.

KEYWORDS: Optimum estimating function; Nonresponse.

1. INTRODUCTION AND BACKGROUND

A typical survey sampling set-up consists of a survey population **P** of N labelled individuals i; $\mathbf{P} = \{i: i = 1, ..., N\}$. With each individual i is associated a real value y_i. The vector $\mathbf{y} = (y_1, ..., y_i, ..., y_N)$ is called the population vector. Any subset s of **P** is called a sample. Let $S = \{s\}$. Any probability distribution p on S is called a sampling design. A sample s is drawn using a sampling design p, and the values y_i: $i \epsilon s$ are ascertained through a survey.

Thus the data here are χ_s where

$$\chi_s = \{s, (i, y_i) : i \in s\}. \tag{1.1}$$

On the basis of the data x_s one tries to estimate a survey population parameter θ_N, that is a specified real function of the population vector **y**; $\theta_N = \theta_N(y)$.

In relation to the above estimation problem we assume a superpopulation model under which $y_1, ..., y_N$ are independent and for certain known covariate values x_i, $i = 1, ..., N$,

$$\epsilon(y_i - \theta x_i) = 0, \; i = 1, ..., N, \tag{1.2}$$

ϵ being the expectation with respect to the model. In the model (1.2), θ is the usual unknown regression parameter, the expectation being taken holding x_i fixed. The usual intercept term of the regression model is not mentioned in (1.2), for this term can often be eliminated by an appropriate stratification (Godambe 1982). Note the model (1.2) does not specify the variance function.

Following Godambe and Thompson (1986), for some *specified* numbers α_i, $i = 1, ..., N$, we define the survey population parameter θ_N as the solution of the equation

$$\tilde{g} = \sum_{i=1}^{N} (y_i - \theta x_i)\alpha_i = 0. \tag{1.3}$$

[1] V.P. Godambe and M.E. Thompson, Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada, N2L 3G1.

That is,

$$\theta_N = \sum_{i=1}^{N} y_i\alpha_i / \sum_{i=1}^{N} x_i\alpha_i. \tag{1.4}$$

The parameter θ_N is related to the model (1.2) through the equation

$$\epsilon\tilde{g} = 0. \tag{1.5}$$

Any real function h of the data χ_s in (1.1) and the parameter θ is called an *unbiased estimating function* for both the parameters θ_N and θ if

$$E(h - \tilde{g}) = 0 \text{ for all } \mathbf{y} \text{ and } \theta \tag{1.6}$$

'E' being the expectation under the sampling design p employed to draw the sample s. Because of (1.5) and (1.6) we say the solution of the equation

$$h(\chi_s, \theta) = 0,$$

for the given data χ_s, estimates both the parameters and θ and θ_N, given by (1.2) and (1.4) respectively. For the function $\tilde{g}$ in (1.4), under the sampling design p, let $H_{(p)}$ be the class of all unbiased estimating functions h. That is

$$H(p) = \{h\colon E(h - \tilde{g}) = 0 \text{ for all } \mathbf{y} \text{ and } \theta\}. \tag{1.7}$$

Now we say an *estimating function* $h^* \in H(p)$ is *optimum* if

$$\epsilon E(h^*)^2 \leq \epsilon E(h)^2, \text{ for all } h \in H(p) \tag{1.8}$$

(Godambe and Thompson 1986). Further, when the inequality (1.8) is satisfied,

$$h^* = 0 \tag{1.9}$$

is said to be the *optimum estimating equation* for estimating the parameter θ_N given by (1.3) and (1.4).

For the sampling design p, used to draw a sample s, let π_i, $i = 1, \ldots, N$ be the inclusion probabilities. That is

$$\pi_i = \sum_{s \ni i} p(s), \; i = 1, \ldots, N, \tag{1.10}$$

where $s \ni i$ indicates all samples s which include the individual i. We assume

$$\pi_i > 0, \; i = 1, \ldots, N. \tag{1.11}$$

Theorem 1.1. (Godambe and Thompson 1986). For any sampling design p satisfying (1.11), under the model (1.2), in the class of all unbiased estimating functions $H(p)$ in (1.7), the optimum h^*, that is h^* satisfying (1.8), is given by

$$h^* = \sum_{i \in s} (y_i - \theta x_i) \alpha_i / \pi_i, \tag{1.12}$$

π_i being the inclusion probability given by (1.10). Thus the optimum estimating equation here is

$$\sum_{i \in s} (y_i - \theta x_i) \alpha_i / \pi_i = 0. \tag{1.13}$$

The estimate $\hat{\theta}_s$ of the survey population parameter θ_N in (1.4) and the superpopulation parameter θ in (1.2) is given by

$$\hat{\theta}_s = \frac{\sum_{i \in s} y_i \alpha_i / \pi_i}{\sum_{i \in s} x_i \alpha_i / \pi_i} \cdot \tag{1.14}$$

This estimate was previously put forward by Brewer (1963) and Hájek (1971) on some "plausibility" considerations.

To explain the relationships of Theorem 1.1 above with earlier optimality results (e.g. Godambe 1982) we put $\alpha_i \equiv 1$ in (1.3) and therefore in (1.2). Further, we consider a superpopulation model obtained from (1.2) by letting $\theta = \theta_0$, a specified value. Now for any sampling design with inclusion probabilities π_i satisfying (1.11), in the class of all design unbiased estimates of θ_N (in (1.4) with $\alpha_i = 1$, $i = 1, \ldots, N$), the superpopulation expectation of the design variance is minimized for the estimate

$$e = \frac{1}{X} \left\{ \sum_{i \in s} \frac{y_i - \theta_0 x_i}{\pi_i} + \theta_0 \sum_{i=1}^{N} x_i \right\} \tag{1.15}$$

where $X = \Sigma_1^N x_i$. This "optimality" of the estimate e at $\theta = \theta_0$ carries over to all values of θ if the sampling design is such that

$$\text{Probability} \left\{ s: \left(\sum_{i \in s} \frac{x_i}{\pi_i} - \sum_{i=1}^{N} x_i \right) = 0 \right\} = 1. \tag{1.16}$$

Now when the sampling design satisfies condition (1.16), then $\hat{\theta}_s$ in (1.14) is equal to e in (1.15). Thus all the earlier optimality results are covered by Theorem 1.1, and it does a great deal more: in many situations, such as for designs with $\pi_i \propto x_i$, the condition (1.16) implies a *fixed sample size* design. In contrast the "optimality" in Theorem 1.1 holds regardless of the fixed sample size design condition. That is, the "optimality" is available for *random sample* size designs, which are common in the nonresponse situations discussed subsequently.

2. NONRESPONSE AND OPTIMALITY

Suppose a sample s is drawn from the survey population **P**, using a sampling design p. Suppose because of nonresponse the variate values y_i are available only for the subset $s' \subset s$; $s - s'$ are the non-respondents. Thus now the data instead of χ_s in (1.1) are

$$\chi_{s,s'} = (s, s', \{(i,y_i): i \in s'\}). \tag{2.1}$$

We may now consider two problems of estimation:

(I) If there were no nonresponse, that is if all the data χ_s in (1.1) where available, we would have estimated the survey population parameter θ_N in (1.4) by solving the optimum estimating equation given by (1.12), namely $h^* = 0$. When the hypothetical data χ_s are replaced by $\chi_{s,s'}$ in (2.1), one may try to estimate h^* with some function $h'(\chi_{s,s'})$. This is in line with a suggestion of Rubin (1976). Following (1.7) we define the class of unbiased estimating functions h' (for h^*, given the sample s) as

$$H'(p,.,s) = \{h': E(h' - h^*|s) = 0, \text{ for all } \mathbf{y} \ \& \ \theta\}; \tag{2.2}$$

the '.' in H' indicates that the class H' would be specified only after the *response mechanism* is specified. Again we define h'^* as the optimum estimating function in H' in (2.2), if $h'^* \in H'$ and if under the model (1.2), $\epsilon E(h'^*)^2 \leq \epsilon E(h'^*)^2$ for all $h' \in H'$.

(II) Alternatively we could try to estimate the survey population parameter θ_N directly, that is without estimating h^* as in (I) above, from the data $\chi_{s,s'}$. In line with (1.7) we define the class of unbiased estimating functions $h''(\chi_{s,s'})$:

$$H''(p,.) = \{h'': E(h'' - \bar{g}) = 0, \text{ for all } \mathbf{y} \ \& \ \theta\}; \tag{2.3}$$

as before the '.' in H'' indicates that the class H'', for its specification, requires the specification of the *response mechanism*. Again h''^* is called the *optimum estimating function* in H'' if $h''^* \in H''$ and if under (1.2), $\epsilon E(h''^*)^2 \leq \epsilon E(h'')^2$ for all estimating functions $h'' \in H''$.

In $H'(p,.,s)$ and $H''(p,.)$ of (2.2) and (2.3) we have left the response mechanism '.' unspecified. Now we specify it.

RESPONSE MECHANISM: If the individual 'i' of the survey population **P** were included in the sample s drawn,

$$\begin{array}{l}\text{'}i\text{' would respond with } \textit{known} \text{ probability } q_i \\ \text{and would fail to respond with probability } 1-q_i,\end{array} \tag{2.4}$$

$i = 1, \ldots, N$; we assume $q_i > 0$, $i = 1, \ldots, N$.

The response mechanism $\mathbf{q} = (q_1, \ldots, q_N)$ in (2.4) completely characterizes the class $H'(p,.,s)$ in (2,2) as $H'(p, \mathbf{q}, s)$ and $H''(p,.)$ in (2.3) as $H''(p, \mathbf{q})$.

The case (I) above is implemented by the following Theorem 2.1 and the remaining Theorems 2.2, 2.3 and 2.4 implement the case (II).

Theorem 2.1. For any sampling design p satisfying (1.11), and for any sample s, in the class of estimating functions $H'(p, \mathbf{q}, s)$ in (2.2) under the superpopulation model (1.2) $\epsilon E\{h')^2 | s\}$ is minimized for $h' = h'^*$ where

$$h'^* = \sum_{i \in s'} (y_i - \theta x_i)\alpha_i/\pi_i q_i; \tag{2.5}$$

that is h'^* is the optimum estimating function in $H'(p, \mathbf{q}, s)$. □

Proof. As was emphasized in Section 1, the optimality of h^* in (1.12) obtains even for *random sample size* designs and for any values of α_i, $i = 1, ..., N$ in (1.3). Thus the proof of Theorem 2.1 is accomplished by replacing, in Theorem 1.1, the population 'P' by 's' and α_i by α_i/π_i, $i \in s$ and noting that now the inclusion probabilities are q_i, $i \in s$. □

Theorem 2.2. Let $\bar{H}''$ be the subclass of H'' in (2.3) such that any estimating function $h''(\chi_{s,s'})$ in $\bar{H}''$ depends on (s,s') only through s'. Then for any sampling design p satisfying (1.11), in the class $\bar{H}''(p, \mathbf{q})$, under the superpopulation model (1.2), $\epsilon E\{(h'')^2\}$ is minimized for $h'' = h''^*$ where

$$h''^* = \sum_{i \epsilon s'} (y_i - \theta x_i)\alpha_i / \pi_i q_i; \tag{2.6}$$

that is h''^* is the optimum estimating function in $\bar{H}''(p, \mathbf{q})$. □

Proof. This follows directly from Theorem 1.1, by replacing in it s by s' and the inclusion probabilities by π_i by $\pi_i q_i$, $i = 1, ..., N$.

Theorem 2.3. The estimating function h''^* in (2.6) is the optimum estimating function in the entire class $H''(p, \mathbf{q})$ given by (2.3). That is the result of the Theorem 2.2 is valid without the restriction to the subclass $\bar{H}''$ of H''. □

Proof. For any given response probabilities $\mathbf{q}$ in (2.4) and the sampling design p, the statistic $(\{i, y_i\}: i \in s')$ is *sufficient* for the population vector $\mathbf{y}$. More specifically, referring to (1.1) and (2.1), we have the conditional probability Prob$(\chi_{s,s'} \mid \chi_{s'},\mathbf{y})$ independent of $\mathbf{y}$. Hence for any estimating function $h'' \in H''(p, \mathbf{q})$ in (2.3) we have the estimating function $E(h'' \mid \chi_{s'}) = \bar{h}'' \in \bar{H}''$ and $\epsilon E(\bar{h}'')^2 \leq \epsilon E(h'')^2$. This proves Theorem 2.3.

When $s \equiv s'$, that is when there are no nonrespondents, do we still estimate h^* by $h'^* = h''^*$? The obvious negative answer to this question is obtained, as shown by Godambe (1986), by an appropriate *conditioning*. The same reservation tends to be felt for cases where there are only a few nonrespondents, and again appropriate conditioning holds some promise of a resolution. In summary the formal optimality of $h'^* = h''$ suggests that it is useful, and is likely to give good estimation when nonresponse is considerable and the relative values of the q_i are known. However, it can clearly be improved upon in situations when nonresponse is rare; improved versions will have natural conditional interpretations. Appropriate conditioning becomes even more important in the case of unknown response probabilities, as will be seen next.

Now we assume that the survey population $\mathbf{P}$ is divided into k strata $\mathbf{P}_j$, of sizes N_j, $j = 1,..., k$. Further suppose that the response probabilities are constant within each stratum. That is

$$q_i = q^{(j)} \text{ for all } i \in \mathbf{P}_j; j = 1, ..., k. \tag{2.7}$$

Unlike in (2.4), where the response probalities were assumed to be known, now we assume that in (2.7), the response probabilities $q^{(j)}$, $j = 1, ..., k$ are *unknown*. Let p_0 denote the stratified sampling design, consisting of drawing from the stratum $\mathbf{P}_j$, a simple random sample (without replacement) of size n_j, $j = 1, ..., k$. Now as in (2.3) we define the class of unbiased estimating functions $h_1(\chi_{s,s'})$

$$H_i(p_0) = \{h_i\colon E(h_1 - \tilde{g}) = 0 \text{ for all } \mathbf{y}, \theta \text{ and } q^{(j)}, j = 1, ..., k\}, \tag{2.8}$$

where $q^{(j)}$ are as in (2.7). Let $s_j' = s' \cap \mathbf{P}_j$ and $|s_j'| = n_j'$, that is the size of the sample of respondents from the stratum $\mathbf{P}_j$, $j = 1, ..., k$.

Theorem 2.4. For the sampling design p_0, in the class of estimating functions $H_i(p_0)$ in (2.8), under the superpopulation model (1.2), ${}_\epsilon E(h_1^2)$ is minimized for $h_1 = h_1^*$ where

$$h_1^* = \sum_{j=1}^{k} \sum_{i \in s_j'} (y_i - \theta x_i)\alpha_i/(\frac{n_j'}{N_j}); \tag{2.9}$$

that is h_1^* is the optimum estimating function in $H_i(p_0)$.

Proof. The sampling distribution of the data $\chi_{s,s'}$ in (2.1) depends, in addition to the unknown population vector y, on the unknown (parameter) $q^{(j)}$, $j = 1, ..., k$. Now for every fixed $\mathbf{y}$, the statistic n_j', $j = 1, ..., k$ is *completely sufficient* for the parameter $q^{(j)}$, $j = 1, ..., k$. Hence for a fixed $\mathbf{y}$ and θ, in (2.8),

$$[E(h_1 - \tilde{g}) = 0, \text{ for all } q^{(j)}, j = 1, ..., k]$$

$$\Rightarrow E\{(h_1 - \tilde{g})|n_j', j = 1, ..., k\} = 0, \tag{2.10}$$

ignoring sets of '0' measure. Further, *conditional* on the number of respondents n_j' from the stratum P_j, the probability of $i \in s_j'$ is $(n_j/N_j)(n_j'/n_j) = (n_j'/N_j)$. Hence for any estimating function $h_1 \in H_1$ in (2.8) we have from Theorem 2.3.

$${}_\epsilon E((h_1^*)^2| n_j', j =1, ..., k\} \leq {}_\epsilon E\{(h_1)^2| n_j', j = 1, ..., k\}, \tag{2.11}$$

h_1^* being given by (2.9). Theorem 2.4 is proved by taking the expectations of both sides of (2.11) for the variations of n_j', $j = 1, ..., k$.

The optimum estimating function h_1^* in (2.9) has the following intuitive interpretation. If in (2.7), the response probabilities $q^{(j)}$, $j = 1, ..., k$ were *known*, by Theorem 2.3, the optimum estimating function, for the sampling design p_o, would be given by

$$h'' = \sum_{j=1}^{k} \sum_{i \in s_j'} (y_i = \theta x_i)\alpha_i/(\frac{n_j}{N_j} q^{(j)}).$$

Now when $q^{(j)}$ are unknown (which is the case in Theorem 2.4), we *estimate* them by (n_j'/n_j), $j = 1, ..., k$. Substituting these estimates for $q^{(j)}$ in h'' yields the estimating function h_1^* of (2.9).

These estimates obtained by solving the equations $h'^* = 0$, $h''^* = 0$ and $h_1^* = 0$ in (2.5), (2.6) and (2.9) respectively have previously been proposed, on plausibility considerations, by several authors. A good reference in this connection in Cassel et al. (1983). The assumption (2.4) of "response probabilities" seems to have evolved gradually in the literature. An interesting early reference in this connection is Hartley (1946).

3. OPTIMAL INCLUSION PROBABILITIES

It should be emphasized here that the "optimality" of the estimating function $h''*$ in (2.6) was established under the superpopulation model (1.2), which does *not* specify the variance function. However the specification of the variance function in the model (1.2) would be required to obtain the "optimal" inclusion probabilities. We assume

$$\epsilon(y_i - \theta x_i)^2 = \sigma^2 f(x_i),\ i = 1, \ldots, N, \tag{3.1}$$

where f is a *known* function of x, and σ^2 can be unknown. Now for the estimating function $h''*$ in (2.6), (3.1), we have

$$\epsilon E(h''*)^2 = \sum_{i=1}^{N} \frac{\epsilon(y_i - \theta x_i)^2 \alpha_i^2}{\pi_i q_i} = \sigma^2 \sum_{i=1}^{N} \frac{f(x_i)\alpha_i^2}{\pi_i q_i} \tag{3.2}$$

In (3.2), the response probabilities q_i as said in (2.4) are given (fixed) numbers. However, (a sampling design with) the optimal inclusion probabilities can be obtained by minimizing $\epsilon E(h''*)^2$ in (3.2) under a restriction, either (A) or (B).

$$(A)\colon \sum_{i=1}^{N} \pi_i = \text{constant},$$

$$(B)\colon \sum_{i=1}^{N} \pi_i q_i = \text{constant} \tag{3.3}$$

In (A) we hold the average size of the sample s fixed, for $E|s| = \Sigma_i^N \pi_i$. In (B) we hold fixed the average size of the effective sample s', for $E|s| = \Sigma_i^N \pi_i q_i$. Now since the q_i are fixed numbers we have for minimizing $\epsilon E(h''*)^2$ in (3.2), respectively,

$$(A)\colon \pi_i \propto \{\frac{f(x_i)}{q_i}\}^{1/2}\alpha_i,$$

$$(B)\colon \pi_i \propto \frac{(f(x_i))^{1/2}}{q_i}\alpha_i. \tag{3.4}$$

Denoting by n' the size of the effective sample s', that is $|s'| = n'$, we have from (B) in (3.4),

$$\pi_i = \frac{(f(x_i))^{1/2}\alpha_i}{\{\Sigma_1^N (f(x_i))^{1/2}\alpha_i\}} \frac{E(n')}{q_i},\ i = 1, \ldots, N. \tag{3.5}$$

Further for a fixed sample size design such that

$$\text{Probability } \{s\colon |s| \neq n\} = 0,$$

we have from (3.5).

$$\sum_{i=1}^{n} \pi_i = n = \sum_{i=1}^{N} \frac{(f(x_i))^{½}\alpha_i}{\{\Sigma_1^N (f(x_i))^{½}\alpha_i\}} \frac{1}{q_i} E(n'). \tag{3.6}$$

As a special case, when all the response probabilities q_i, $i = 1, ..., N$ are equal, $q_i = q$ say, $i = 1, ..., N$, in (3.6),

$$n = E(n')/q; \tag{3.7}$$

for instance if $q = 1/2$, the sample size of the (initial) sample s should be double the expectation of the effective sample (s') size!

Now we assume the survey population **P** to be divided into strata P_j, $= 1, ..., k$ so that the response probabilities in each stratum are constant, that is they satisfy (2.7). For a stratified sampling design consisting of drawing a sample of size n_j from the stratum $\mathbf{P}_j$, $j = 1, ... k$ we have from (3.5).

$$n_j = \frac{E(n')}{q^{(j)}} \frac{\sum_{i \in \mathbf{P}_j} (f(x_i))^{1/2}\alpha_i}{\sum_{i \in \mathbf{P}} (f(x_i))^{1/2}\alpha_i,}, j = 1, ..., k.$$

If $(f(x_i))^{½}\alpha_i$ are constant for $i = 1, ..., N$, it is clear from (3.8) that optimal allocation implies drawing a relatively larger sample from the stratum with smaller response probability. Actually in this situation

$$E(n_j') = E(n')/k$$

where n_j' is the size of the effective sample s_j' from the stratum $\mathbf{P}_j$, $j = 1, ..., k$.

REFERENCES

BREWER, K.R.W. (1963). Ratio estimation in finite populations: some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, 5, 93-105.

CASSEL, C.M., SARNDA, C.E., and WRETMAN, J.H. (1983). Some uses of statistical models in connection with the nonresponse problems. In *Incomplete Data in Sample Surveys*, Vol. 3, (Eds. W.G. Madow and Ingram Olkin), New York: Academic Press, 143-160.

GODAMBE, V.P. (1982). Estimation in survey sampling: Robustness and optimality. *Journal of the American Statistical Association*, 77, 393-403.

GODAMBE, V.P. (1986). Quasi-score function, quasi-observed Fisher information and conditioning in survey sampling (unpublished).

GODAMBE, V.P. and THOMPSON, M.E. (1986). parameters of superpopulation and survey population: Their relationships and estimation. *International Statistical Institute Review* (to Appear).

HAJEK, J. (1971). Contribution to discussion of paper by D. Basu. In *Foundations of Statistical Inference*, (Eds. V.P. Godambe and D.A. Sprott), Toronto: Holt, Rinehart and Winston, 236.

HARTLEY, H.O. (1946). Discussion of paper by F. Yates. *Journal of the Royal Statistical Society* Series A, 109, 37.

RUBIN, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-589.