OPTIMAL ALLOCATION TO CONTROL QUESTIONNAIRE DESIGN VARIANCE IN SAMPLE SURVEYS

Gad Nathan, Hebrew University
Monroe G. Sirken, National Center for Health Statistics

1. Introduction

This is a sequel to an earlier paper by Nathan and Sirken (1986) which presented a random effects model for estimating the response errors due to questionnaire design (QD) in sample surveys. Traditionally, QD errors have been investigated principally for their biasing effects. The QD model, on the other hand, proposes to measure the QD effects in terms of response variance. This change in perspective is compatible with recent findings of Marquis, Marquis and Polich (1986), who after studying the direction and size of response biases in a number of sensitive topic surveys, concluded that the distribution of biases appeared to center on zero. In this paper, we present design strategies for controlling QD variance in sample surveys, and for minimizing the joint effects of QD and sampling variance.

The proposed questionnaire design response error model is based on the classical test theory of Lord and Novick (1968), and the response error models of Hansen, Hurwitz, and Bershad (1961). It decomposes the total variance of a sample survey estimate into the QD variance and a residual or sampling variance. The QD variance is defined as the variance of the expected responses over a universe of exchangeable QD versions which has the following properties:

Property 1 - the universe contains two or more interchangeable versions of the questionnaire, differing in design but asking for the same information.

Property 2 - the expected value of responses over all QD's in the universe is the true value.

The situations favoring the construction of QD universes with both these properties has yet to be determined. Certainly it would would be desirable and possibly feasible to do so in situations where it is known or suspected that particular QD features are likely to produce systematic cognitive errors in the respondents' answers. For example, responses to questions that provide respondents with initial values to construct numeric answers are usually biased toward the initial values - the so-called anchoring phenomenon. And responses to questions with multiple response categories are often biased toward the categories at the beginning and end of the list - the so-called primacy and recency effects.

In situations such as these, in which responses to any single QD option is likely to be biased, it certainly would be feasible to construct universes of exchangeable QD options and possibly in such a manner that the expected value over the QD universe would be the true value. If, for example, a question elicited response bias due to the anchoring effect, a QD universe could be constructed from the QD options having different starting points. And if the question was subject to response bias due to primacy and recency effects, the QD universe would contain QD options with different orderings of the response categories.

Whether or not unbiased QD universes could be constructed in this manner is a matter worth exploring. For example, it might be feasible to construct a "symmetric" universe of QD's such that the expected value of responses over all exchangeable QD's averaged out the response biases associated with each of the QD versions in the universe.

The possibility of constructing "symmetric" universes was illustrated by an experiment by Monsees and Massey (1979) which tested the effect of the order of the income categories on the income distribution in a national telephone survey. Their split panel survey tested three QD versions that varied by the income category that was asked first. Each QD version yielded an income distribution that was biased toward the initial income category. However, the income distributions that were based on the combined responses to multiple QD versions appeared to be essentially unbiased.

A better understanding of the cognitive processes that lead to biased judgments would improve the prospects of constructing unbiased QD universes, and more importantly, might ultimately provide a scientific base to the art of designing questionnaires. Bradburn, Rips and Shevell (1987) describe how the results from cognitive psychology can be useful in understanding and controlling survey errors. Cognitive pyschologists, Tversky and Kahneman (1974), for example, describe a number of cognitive "heuristics" that lead to biased judgments, and Brown, Rips, and Shevell (1985) describe the cognitive processes that lead to biases in judging the dates of past events. To promote and advance interdisciplinary research on the cognitive aspects of survey methodology, the National Center for Health Statistics (1986) recently established a National Laboratory for Collaborative Research in Cognition and Survey Measurement which is being jointly supported by the Center and the National Science Foundation.

The proposed QD response error model may be useful even in the absence of information about the unbiasedness of the QD universe. In these situations QD variance could serve as a measure of response sensitivity to variations in alternative QD versions. When large, the estimated QD variance would serve as a warning signal of response instability. When small, it would provide a degree of reassurance about the stability of response. The use of the QD variance in this way might be particularly appropriate when the survey dealt with subjective phenomena such as attitude surveys where external validity measures would not be available.

The application of the proposed QD model implies a split panel survey design in which a sample of "k" QDs, is selected at random from a

universe of exchangeable QDs, and then randomized over a sample of "n" reporting units. Where split panel features can be incorporated into the main survey design, the model provides the means for routinely estimating QD variance from data collected entirely as a by-product of the survey itself without resorting to any non-survey data sources. Furthermore, as we will indicate in this paper, the proposed model provides design strategies, (1) for controlling QD variance by varying "k", the number of QD versions selected from the QD universe, and
(2) for minimizing the joint effects of sampling and QD variances for fixed survey costs by selecting k_{opt}, the optimum number of QD options, and n_{opt}, the optimum sample size.

In the next section the basic model is extended to domain proportions since, as originally presented, it applied only to proportions of the entire population. The remainder of this paper considers the problem of optimizing survey resources allocation to control the joint effects of sampling errors and QD errors. Equations are derived in Section 3 for k_{opt} and n_{opt}, and an application is presented in Section 4.

2. The Basic Model

The random effects model, proposed by Nathan and Sirken (1986) for estimates of proportions for the whole population, is extended, as follows, to the estimation of domain proportions. We consider a single dichotomous variable, Y, defined over a finite population of size N and a domain of interest of size R_0N.
A simple random sample (with replacement) of size n is selected from the whole population and the number of units in the domain of study included in the sample is denoted n_0 (a random variable).

A hypothetical infinite or large universe of QD's, whose effects are considered exchangeable, is assumed, from which k QD's are selected at random. Let Y_{ij} represent the response elicited by the i-th QD (i=1,...,k) for the j-th sample unit, belonging to the domain of study (j=1,..., $,n_0$). The basic model is:

$$Y_{ij} = P_i + e_{ij} = P_0 + D_i + e_{ij}$$

$$(i=1,\dots, k;\ j=1,\dots, n_0), \qquad (1)$$

where P_i are i.i.d. with $E(P_i) = P_0$ (the proportion of positive responses in the domain) and $V(P_i) = V(D_i) = \sigma_D^2$, so that $Y_{ij}|P_i \sim B(1,P_i)$ and $E(e_{ij}) = 0$; $V(e_{ij}) = P_0(1-P_0) - \sigma_D^2 = \sigma_E^2$.

Thus the total variance of Y_{ij} can be broken down as follows

$$V(Y_{ij}) = P_0\ (1-P_0) = \sigma_D^2 + \sigma_E^2, \qquad (2)$$

where σ_D^2 is the QD variance and σ_E^2 is the residual error variance.

The k QD's are assumed to be allocated at random to the n sample units so that m=n/k sample units are allocated to each QD (for simplification, m is assumed to be integral). The number of units allocated to the i-th QD and belonging to the domain of study, m_i, is a random variable with the binomial distribution, $m_i \sim B(m, R_0)$. Although m_i takes the value zero with positive probability (approximately $\exp(-mR_0)$ if R_0 is small), the probability of any given QD not being represented in the domain is extremely small (less than 1 in 10,000 if the expected number of units per option in the domain, mR_0, is not less than ten). In the following, we consider expectations involving m_i to be conditional, given that m_i is positive.

Estimators of P_0 are based on the observed proportions, $\hat{P}_i$, of positive responses, out of the responses for the m_i units belonging to the domain and allocated to the i-th QD. Then it is easily seen that each $\hat{P}_i$ is an unbiased estimate of P_0 with variance:

$$V(\hat{P}_1) = \sigma_D^2 + [P_0(1-P_0) - \sigma_D^2]\ E(1/m_i), \qquad (3)$$

where

$$E(1/m_i) = 1/(mR_0) + (1-R\ )/(m^2R_0^2) + 0(m^{-3}). \qquad (4)$$

Obviously $\hat{P} = (1/k)\ \sum_i \hat{P}_i$ is an unbiased estimator of P_0 with approximate variance - to order $0(n^{-3})$ - of:

$$V(\hat{P}) \doteq P_0(1-P_0)\ \{\delta/k + (1-\delta)\ [nR_0 + k(1-R_0)]/(nR_0)^2\}\ . \qquad (5)$$

where $\delta = \sigma_D^2/P_0(1-P_0)$.

For all practical purposes, the approximation to order n^{-2} obtained by dropping the term $k(1-R_0)/(nR_0)^2$ can be used giving

$$V(\hat{P}) \doteq P_0\ (1-P_0)\ [1 + (mR_0-1)\delta]/(nR_0)\ . \qquad (6)$$

Since there is some fluctuation in the values of m_i and some of them may be zero, an alternative estimator, which weights the values of $\hat{P}_i$ by m_i, may be considered:

$$P^* = (\sum m_i\hat{P}_i)/(\sum_i m_i) \qquad (7)$$

This estimator, which basically disregards the allocation to QD's, has the advantage that for its expectation to exist, it is sufficient to condition on $\sum_i m_i \neq 0$ and the probability that $\sum_i m_i = 0$ is negligible. The variance of the weighted estimator, to order n^{-2}, is:

$$V(\hat{P}^*) \doteq P\,(1-P)\,\frac{1+(m-1)R_0\delta}{1+(n-1)R_0} \quad . \qquad (8)$$

This will be somewhat smaller than the variance of $\hat{P}$ if δ is small. In particular $V(\hat{P}^*) < V(\hat{P})$ if and only if $\delta < [(n-m)R_0+1]^{-1}$. For most practical sets of conditions the differences between $V(\hat{P}^*)$ and $V(\hat{P})$ are very small.

3. Optimal Allocation

Both the variance of $\hat{P}$, (5) or (6), and that of $\hat{P}^*$, (8), are increasing functions of k - the number of different QD's used. Although there are practical limitations on the number of QD's which can be used in the same survey, a range of feasible values of k may be considered. The cost of the survey will, in general, increase as a function of k. The increase could be both in the fixed cost component - reflecting the added costs of additional questionnaire types, training manuals, etc. - and in the cost per unit - reflecting increases in costs of training, field-control, editing, etc. for each unit included in the survey. A reasonable linear cost function could be:

$$C = C_1 n + C_2 nk + C_3 k, \qquad (9)$$

where C_1 is the unit cost, independent of the number of QD's, C_2 is the increase in unit cost due to an additional QD and C_3 is the fixed cost component per QD. The minimization of the variance expression (5), subject to (9), requires the solution of fifth degree equations. However, the second order approximation, (6), is minimized, subject to (9), for the values:

$$k_{opt} = \frac{C/C_3}{1+\sqrt{(1-\delta)/(R_0\delta)}\;\sqrt{C_1/C_3+(C_2/C_3)(C/C_3)}}$$

$$n_{opt} = \frac{C-C_3 k_{opt}}{C_1+C_2 k_{opt}} \qquad (10)$$

Since the variances of $\hat{P}$ and of $\hat{P}^*$ are very close, the value of k_{opt} (which in any case has to be rounded off to the nearest integer) will also approximately optimize the variance of the weighted estimator, (8).

4. Empirical Example

In Nathan and Sirken (1986) the basic model, with respect to estimation for the whole population, was applied to data from a pretest for the 1986 National Health Interview, conducted by the Bureau of the Census for the National Center for Health Statistics in the fall of 1985. Estimates of δ were obtained for several questions from a dental health supplement with respect to four QD's (two questionnaire versions and two interview types). Positive values for δ of 0.012 for a question on the use of a fluoride mouth rinse and of 0.007-0.008 for questions on dental visits during the past two weeks and on the use of supplementary fluoride products were obtained.

In order to apply the results of the previous sections, parameters of the cost function (9) were required. Rough "guestimates" of the parameters for the NHIS were obtained as follows:

$C = 5{,}000{,}000$

$C_1 = 100$

$C_2 = 0$

$C_3 = 50{,}000$

Using these values the optimal values of k and of n, from (10), are given in Table 1, for R_0 = .001, .005, .01 and for δ = .0075 and .012 (the approximate values obtained for the NHIS). The QD effect, defined as the ratio of the variance of the estimate when using k QD's, to that of the estimate based on a single QD, is given for the optimal k, with respect to $\hat{P}$ and $\hat{P}^*$ and for k=2 and 10, with respect to $\hat{P}$.

From the table it can be seen that considerable reduction in variance can be achieved for larger domain proportions. The differences between the unweighted and weighted estimators are negligible (for the optimal value of k). The sensitivity to different values of k can be judged from the last two columns. For all combinations considered, if the number of QD's differs from the optimal value by one unit, the change in variance is less than one in a thousand. The sensitivity to changes in the parameters is also small. For instance, for the domain proportion R_0 = .001, values of δ between .0068 and .0095 give the same optimal value of k. For R_0 = .001 and δ = .0075 values of C/C_3 between 89 and 105 and values of C_1/C_3 between .0016 and .0023 result in the same value for k_{opt}.

Table 1: Optimal Number of QD's, k_{opt}, Optimal Sample Size, n_{opt} , and QD Effects

Domain Proportion R_o	Ratio of QD Variance to Total Variance δ	Optimal No. of QD's k_{opt}	Optimal Sample Size n_{opt}	QD Effect $k = k_{opt}$ $\hat{P}$	QD Effect $k = k_{opt}$ $\hat{P}$*	QD Effect k=2 $\hat{P}$	QD Effect k=10 $\hat{P}$
.001	.0075	6	47,000	.812	.801	.871	.828
	.012	7	46,500	.718	.711	.819	.725
.005	.0075	12	44,000	.446	.447	.678	.448
	.012	15	42,500	.341	.343	.627	.350
.01	.0075	16	42,000	.298	.299	.608	.311
	.012	20	40,000	.219	.221	.573	.243

5. References

Bradburn, N.M., Rips, L.J. and Shevell, S.K. (1987) "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys", Science, 236, 157-161.

Brown, N.R., Rips, L.J. and Shevell, S.K. (1985) "The Subjective Dates of National Events in Very-Long-Term Memory", Cognitive Psychology, 17, 139-177.

Hansen, M.H., Hurwitz, W.N. and Bershad, M.A. (1961) "Measurement Errors in Censuses and Surveys", Bulletin of the International Statistical Institute, 38, 351-374.

Lord, F. and Novick R.N. (1968) Statistical Theories of Mental Test Scores, Addison-Wesley.

Marquis, K.H., Marquis, M.S., and Polich, J.M. (1986) "Response Bias and Reliability in Sensitive Topic Surveys", Journal of the American Statistical Association, 81, 381-389.

Monsees, M.L. and Massey, J.T. (1979) "Adapting Procedures for Collecting Demographic Data in a Personal Interview to a Telephone Interview", Proceedings of the Social Statistics Section, American Statistical Association, 130-135.

Nathan, Gad and Sirken, Monroe G. (1986) "Response Error Effects of Survey Questionnaire Design", Proceedings of the Section on Survey Research Methods, American Statistical Association, 493-498.

National Center for Health Statistics (1986) "National Laboratory for Collaborative Research in Cognition and Survey Measurement". A brochure.

Tversky, A. and Kahneman, D. (1974) "Judgments Under Uncertainty: Heuristics and Biases", Science, 185, 1124-1131.

This research was supported by NSF grant (SES-8612320).