BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: CENSUS/SRD/RR-85/08

ESTIMATION OF THE CORRELATED COMPONENT OF RESPONSE VARIANCE FOR CATEGORICAL VARIABLES

by

S. Lynne Stokes
Center for Statistical Sciences
University of Texas
and
Mary Mulry-Liggan
Statistical Research Division

Recommended by: Howard Hogan

Report completed: May 21, 1985

Report issued: May 31, 1985

ESTIMATION OF THE CORRELATED COMPONENT OF RESPONSE VARIANCE FOR CATEGORICAL VARIABLES

S. Lynne Stokes and Mary H. Mulry-Liggan

S. Lynne Stokes is Research Fellow, Center for Statistical Sciences, University of Texas, Austin TX 78712, and Mary H. Mulry-Liggan is Mathematical Statistician, Statistical Research Division, U.S. Bureau of the Census, Washington, D.C. 20233. This research was supported by the U.S. Bureau of the Census.

ABSTRACT

The variance of the ANOVA estimator of the variance component for a balanced one-way random model which has a categorical response variable is derived, without making any distributional assumptions for the random effect. The tightest possible bounds for this variance are determined from the Markov-Krein Theorem by assuming knowledge of only the first two or three moments and possibly the support of the random effect. The bounds are shown to be useful for planning the survey design, including the sample size requirements, of an interpenetration experiment to estimate the correlated component of response variance for a categorical item in a sample survey. Examples illustrate that the resulting required sample sizes may deviate greatly from that which would be appropriate for estimating the correlated component for a continuous response variable assumed to meet the normality assumptions of the classical variance components model.

KEY WORDS: Variance component; Interpenetration; Interviewer variance; Survey non-sampling error; Markov-Krein Theorem.

1. INTRODUCTION

Errors introduced in the measuring, editing, or coding of responses in a sample survey affect the behavior of the estimators obtained from the sample and sometimes affect our ability to measure that behavior. Models designed to measure the impact of these errors indicate that the non-sampling errors may contribute substantially to the bias and/or variance of the estimators obtained from the sample. Furthermore, when these errors are positively correlated within the sample, as they might be when a single operator, such as an interviewer or coder, handles a number of cases, the usual estimators of the standard errors of means and totals are likely to be biased downward. If good estimates of this correlation, sometimes called the correlated component of response error, can be made, the estimates of the standard errors can be improved and problem items may be identified.

Most methods for estimating the correlated component require interpenetration of operators, a technique introduced by Mahalanobis (1946). In its most basic form, interpenetration requires the random sample of size n from a population of size N to be randomly divided into k subsamples of size $m = n/k$, and each subsample to be assigned to a single operator. Then the typical model describing y_{ijt}, the recorded value in the t^{th} survey replication for the j^{th} unit of the population, which is in operator i's assignment, is

$$y_{ijt} = \mu_j + e_{ijt}, \tag{1}$$

where $\mu_j = E(y_{ijt} | j)$. In some cases, μ_j can be thought of as the true value, or at least the true value together with any biases which cannot be separated from it. Then e_{ijt} is the error in that recorded value, and $E(e_{ijt}|\, j) = 0$. For $\bar{y}_t = \sum_i\sum_j y_{ijt}/km$, we have, ignoring the finite population correction factor,

$$V(\bar{y}_t) = [V(\mu_j + e_{ijt}) + (m-1)\,Cov(e_{ijt}, e_{ij't})] \,/\, km \qquad (2)$$

if $Cov\,(e_{ijt}, e_{i'j't}) = Cov(\mu_j, e_{ijt}) = 0$. (If the j^{th} unit isn't in operator i's assignment, let $y_{ijt} = 0$.) We will refer to the operator introducing the correlated errors as an interviewer, since that is a common source for such errors. When that is the case, the correlated component of response error is sometimes referred to as the interviewer variance.

In this paper we assume that data comes from only one survey trial, so the subscript t will be suppressed. We first examine a model proposed by Kish (1962) for the correlated errors in continuous data. He decomposed the error term in (1) as $e_{ij} = b_i + e'_{ij}$, where b_i, having mean 0 and variance σ_b^2, can be thought of as a random variable associated with the i^{th} interviewer and represents the average bias that he introduces into a measurement. e'_{ij} represents the composite of all other uncorrelated non-sampling errors (i.e., only one source of correlated errors is assumed), and $Cov(b_i, e'_{ij}) = 0$. Then (1) can be rewritten as

$$y_{ij} = \mu + b_i + \epsilon_{ij}, \qquad (3)$$

where $\mu = E\mu_j$ and $\epsilon_{ij} = (\mu_j - \mu) + e'_{ij}$, having mean 0 and variance σ_ϵ^2, contains both sampling and uncorrelated non-sampling errors. If we assume that $Cov(b_i, b_{i'}) = Cov(b_i, \mu_j) = 0$, then (3) is a random analysis of variance model where the correlated component of response error, CC = $Cov(e_{ij}, e_{ij'})$ is the variance component σ_b^2. (2) can be written as $V(\bar{y}) = V(y_{ij})[1+(m-1)\rho_b]/km$, where $\rho_b = V(b_i) / V(y_{ij}) = \sigma_b^2 / (\sigma_b^2 + \sigma_\epsilon^2)$.

Bailar and Biemer (1984) model the mechanism causing the correlated errors differently when the characteristic being observed is membership in a category. For each category, an interviewer can make two types of errors : ϕ_i is the probability that interviewer i records a unit reporting that it belongs to the category as not belonging to it and Θ_i is the opposite kind of error. Then (ϕ_i, Θ_i) is considered to be a random vector associated with the i^{th} interviewer. We can avoid consideration of individual characteristics of ϕ_i and Θ_i by defining a new random variable $p_i = E(y_{ij}| i) = Pr[y_{ij} = 1| i]$ to be the probability that interviewer i records a randomly chosen unit as belonging to the category. Then the categorical data equivalent of (3) is

$$y_{ij} = Ep_i + (p_i - Ep_i) + \epsilon_{ij}. \quad (4)$$

Since y_{ij} is limited to be either 0 or 1, ϵ_{ij} is restricted to $-p_i$ or $1 - p_i = q_i$, with probabilities q_i and p_i respectively, for each i. Similar statements to those following (3) can be made concerning the components of this model; i.e., $E(\epsilon_{ij}) = E[E(\epsilon_{ij}| i)] = E(-p_i q_i + q_i p_i) = 0$, and $Cov(\epsilon_{ij}, p_i) = E[Cov(\epsilon_{ij}| i)] + Cov[E(\epsilon_{ij}| i), p_i] = 0$. Note that

$$V(\epsilon_{ij}) = E[V(\epsilon_{ij}| i)] + V[E(\epsilon_{ij}| i)] = Ep_i q_i. \quad (5)$$

The correlated component for model (4), assuming that $Cov(p_i, p_{i'}) = 0$, is $Cov(y_{ij}, y_{ij'}) = \sigma_p^2$. The usual estimator for proportion, $\hat{P} = \sum_i\sum_j y_{ij}/km$, has variance

$$V(\hat{P}) = V(y_{ij})[1 + (m-1)\rho_p]/km, \tag{6}$$

where $\rho_p = V(p_i) / V(y_{ij}) = \sigma_p^2 / \mu_p(1 - \mu_p)$, $\mu_p = Ep_i$.

Model (4) is applicable to many problems outside of survey sampling. It is the model for estimating the variance component for a single random effect when the dependent variable is categorical. For example, let $y_{ij} = 1$ if the $(i,j)^{th}$ rat, which is a member of rat i's litter, survives to age A, with $y_{ij} = 0$ otherwise. If the k female rats are considered to be chosen from a large population, then σ_p^2 is a measure of the contribution of their variability to the variability of the estimated survival rate, $\hat{P}$. Discussion in the literature of properties of variance component estimators for models of this type, or even having any non-normal random effect, is rare.

2. THE ESTIMATOR AND ITS VARIANCE

In this and later sections, y_{ij} will denote the j^{th} unit in interviewer i's assignment rather than the j^{th} unit of the population. Then the usual ANOVA estimator of the variance component σ_b^2 or σ_p^2 is $\hat{CC} = (s_b^2 - s_w^2)/m$, where $s_b^2 = [m/(k-1)]\,\Sigma(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ and $s_w^2 = [k(m-1)]^{-1}\Sigma\Sigma(y_{ij} - \bar{y}_{i\cdot})^2$. $\hat{CC}$ is an unbiased estimator of the correlated component for both models (3) and (4).

Biemer and Stokes (1985) derived expressions for the variance of similar estimators of the correlated component, but only for the normal model. In other studies of interviewer variance, empirical variance estimation techniques, such as the jackknife (McLeod and Krotki 1978) and ultimate cluster methods (Bailey, Moore, and Bailar 1978) have been used for evaluating the precision of estimators of the correlated component. These methods are not useful for the planning of sample designs for interpenetration studies, however, since in that case, measures of precision of $\hat{CC}$ are needed before the data collection process begins. That application lead to the devēlopment of the methods discussed in this paper.

Under the assumption of normality of both e_{ij} and b_i in model (3), it is known (e.g., from Searle 1971, p. 474) that

$$V(\hat{CC}) = 2(k-1)^{-1}\,[\sigma_b^4 + 2\sigma_b^2\sigma_\epsilon^2/m + \sigma_\epsilon^4/m^2] + O(m^{-3}). \qquad (7)$$

If the assumption of normality of b_i is discarded, then one can show that

$$V(\hat{CC}) = k^{-1}[\mu_{b4} - ((k-3)/(k-1))\sigma_b^4] + 4\,\sigma_b^2\sigma_\epsilon^2\,/(k-1)m + 2\sigma_e^4\,/(k-1)m^2 + O(m^{-3}) \qquad (8)$$

where μ_{b4} is the 4^{th} central moment of the b_i's.

The response to most survey questions is categorical, rather than continuous, thus making (4) the appropriate model for the observations. The assumptions of normality of both the error term and p_i (since its support is the interval [0,1]) are then incorrect. Thus neither (7) nor (8) is appropriate for assessing the precision of $\hat{CC}$ for such characteristics. It can be shown, however, that for y_{ij} following model (4),

$$V(\hat{CC}) = k^{-1}[\mu_{p4} - ((k-3)/(k-1))\sigma_p^4] + 4\sigma_p^2 E(p_iq_i)/(k-1)m$$
$$+ 2E^2(p_iq_i)/(k-1)m^2 + 4A_p/km + 2B_p/km^2 + O(m^{-3}), \quad (9)$$

where $A_p = E(p_iq_i[(p_i - \mu_p)^2 - \sigma_p^2])$ and $B_p = V(p_iq_i) + E[p_iq_i(q_i-p_i)(p_i - \mu_p)]$. The first three terms of (9) are identical to those of (8), with μ_{p4} and σ_p^2 replacing μ_{b4} and σ_b^2, and (from (5)) $V(\epsilon_{ij}) = E(p_iq_i)$ replacing σ_ϵ^2. A_p and B_p are functions of the first four moments of p_i and can be either positive or negative. One might either over- or underestimate $V(\hat{CC})$ by incorrectly using (7) for categorical variables, as is sometimes done.

Table 1 shows, for several specified distributions of p_i, the ratio of $V(\hat{CC})$ for the categorical model (from (9)) to that of a normal model having comparable variance components ($\sigma_\epsilon^2 = E(p_iq_i)$ and $\sigma_b^2 = \sigma_p^2$ in (7)), where $m = 50$ and $k = 2$. This ratio is denoted by $R_{c,n}$. The symmetric distributions of p_i in the table (a, b, and c) may lead to values of $R_{c,n}$ which are either greater or less than 1, depending on whether the kurtosis (as measured by $\beta_{p2} = \mu_{p4}/\sigma_p^4$) is large or small. The positively skewed distributions, i.e., those having $\beta_{p1}=\mu_{p3}^2/\sigma_p^3 > 0$ (d, e, and f) all have $R_{c,n} > 1$. These distributions have large β_{p2} as well, since $\beta_2 > 1 + \beta_1$ for any random variable (see, e.g., Kendall and Stuart 1977, p. 95). The magnitude of β_2 affects the precision of $\hat{CC}$ in the continuous non-normal model in the same way, as can be seen from (8). A comparison of d and f in Table 1 shows that, in contrast to the continuous case, the mean of p_i affects the precision of $\hat{CC}$ as well.

For most applications in which estimates of variance components for categorical variables are needed, the investigator is not likely to have much information concerning the distribution of p_i, making (9) unusable either for variance estimation or experimental design. However when planning an interpenetration study for estimation of interviewer variance, one is likely, at least, to have information concerning the range of values for the first two moments of p_i for the categorical variables of interest. A range for μ_p will be known from the reported incidence of similar characteristics in previous studies, and a less precise range for ρ_p, and thus for $\sigma_p^2 = \mu_p(1-\mu_p)\rho_p$, can be given from knowledge gained from previous interpenetration studies, such as those reported by Kish (1962). For example, we know that for most demographic items, ρ (used here to mean either ρ_b or ρ_p) is near 0, for factual subject-matter items, ρ is less than about 0.04, and for highly controversial or attitudinal variables, ρ may be as large as 0.1. For applications other than to estimation of interviewer variance, either more or less information may be available about p_i.

One criterion which might be used for determining an adequate sample design for an interpenetration study is that of achieving a specified coefficient of variation ($CV(\hat{\theta}) = [Var(\hat{\theta})/\theta^2]^{1/2}$) for $\hat{CC}$. For example, in a telephone survey in which it is known that interviewers can complete m interviews during the allotted survey period, we might be interested in determining the number of interpenetrated interviewer assignments (k) needed to achieve a specified CV for σ_b^2, the correlated component from the continuous model. If the normality assumptions leading to (7) hold, the required k is that satisfying

$$(CV)^2 = 2\{1 + 2[(1-\rho_b)/\rho_b]/m + [(1-\rho_b)/\rho_b]^2/m^2\}/(k-1) + O(m^{-3}). \qquad (10)$$

Since all terms of (9) involve 3rd and 4th moments of p_i, required values of k cannot be obtained in this way for the categorical model unless further assumptions about p_i are made. The same is true for the continuous non-normal model because of the uncertainty in the magnitude of the first term of (8). The approach taken in this paper is to determine bounds, rather than exact values, for $V(\hat{CC})$ in the categorical case based on knowledge of the boundedness of p_i, its first two moments, and as little additional, speculative information as possible about its distribution. From these bounds, the survey or experimental designer can determine bounds for k or m.

3. VARIANCE BOUNDS IN THE CATEGORICAL MODEL

Bounds for $V(\hat{CC})$ given in (9) can be found by using corollaries of the Markov-Krein Theorem (DeVylder 1982,1983). These corollaries, which are stated in the Appendix, provide tight upper and lower bounds on the expected value of certain functions of a bounded random variable whose first several moments are known. By tight bounds is meant that the corollaries actually produce distributions which achieve each of those bounds, so that the bounds cannot be improved.

An upper and lower bound for the third central moment of p_i, μ_{p3}, can first be found by using Corollary 1 of the Appendix, since p_i is bounded on

[0,1], and since, as noted earlier, the survey planner will have some information about the first two moments of p_i. This interval, which can be obtained from (A.2), is

$$\sigma_p^2(\sigma_p^2 - \mu_p^2) / \mu_p \leq \mu_{p3} \leq \sigma_p^2[(1-\mu_p)^2 - \sigma_p^2] / (1-\mu_p). \quad (11)$$

In order that the conditions of Corollary 2 are satisfied, it must be possible to write (9) as $h(p_i)$, where $h^{(4)}(p_i) \geq 0$. This is true since $h(\cdot)$ is a polynomial of order 4, where the coefficient of p_i^4 is $[1 - 4/m + 6/m^2]/k \geq 0$ for all $\bar{m}$. Then the permissible μ_{p3} values from (11) can be used in Corollary 2 to produce tight upper and lower bounds for $V(\hat{CC})$; i.e., bounds which are actually achievable for some potential distribution of p_i. With the bounds obtained for $V(\hat{CC})$, the problem of determining the survey design parameters required to achieve a specified CV for $\hat{\sigma}_p^2$ can be addressed.

For telephone surveys, the number of interviewer assignments which may be interpenetrated simultaneously is constrained by the number of work-stations and interviewers available for a given shift or set of shifts, rather than by the cost of interpenetration. But for personal visit interviews, interpenetration experiments require increased amounts of travel by all involved interviewers and thus are constrained by the cost of that travel. For many surveys, such as the Current Population Survey (CPS) conducted by the U.S. Bureau of the Census, finding enumeration areas close enough together that interpenetration of large numbers of interviewer assignments is feasible, with the time and cost constraints required, is difficult. For that reason, interpenetration in personal visit interviewing is generally restricted to pairs of interviewer assignments, and a larger sample of interviewers is achieved by increasing the number, r, of interpenetrated pairs. (Biemer and Stokes (1985) found that this design is optimal for a similar estimator of the correlated component when the variable follows model (3) with normally distributed random effects and when the cost of interpenetration increases with $\sqrt{k}$.) Then the estimator of the correlated component is taken to be the average $\hat{CC}$ over all r pairs; i.e., $\hat{CC} = \Sigma\hat{CC}_i/r$, where $\hat{CC}_i$ denotes the estimate from the i^{th} pair, and, ignoring the finite population correction,

$$V(\hat{CC}) = \Sigma V(\hat{CC}_i)/r, \tag{12}$$

where $V(\hat{CC}_i)$ is given by (7) or (8) for the continuous model and (9) for the discrete model, with k = 2. (12), together with information from the corollaries about the range of $V(\hat{CC})$, may be used for determining bounds for r, the number of interpenetrated pairs required to achieve a given CV.

Bounds for k or r yielded by the use of the two corollaries as described will generally be very wide, especially for small values of ρ_p. Two approaches for improvement on these bounds will be discussed and examples of each given in Section 4. Each one is available if a certain type of information can be obtained (or reasonably guessed) about the distribution of p_i.

One approach is to make some assumption about the skewness of the distribution of p_i, and thus about μ_{p3}. For example, one may feel that the distribution of p_i is likely to be symmetric; i.e., the tendency and amount of overreporting of membership in a category is about the same as that of underreporting. Then one might set μ_{p3} to be near 0 and use Corollary 2 to obtain bounds for $V(\hat{CC})$. Or, one may believe that the larger interviewer errors tend to be in the direction of overreporting membership in a category. Then Corollary 2 may be used with μ_{p3} restricted to positive values.

The upper bound provided by the corollaries for the expected values are achieved by distributions which place positive probability at both endpoints of the assumed support of the random variable X and at only one point in the interior. For our application, that means the distribution of p_i leading to the upper bound on the required number of interviewers would have positive probability associated with both 0 and 1 for a single questionnaire item. The actual population of interviewers from which the sample can come is likely, for most survey field operations, to be somewhat homogeneous, since they must undergo screening and training before they can enter the available pool. Thus it seems unlikely that interviewer p_i's could achieve the extremes of both 0 and 1 on a single questionnaire item. Therefore, a second approach to shortening the intervals for k or r is to restrict the range [a,b] in Corollaries 1 and 2 to be narrower than [0,1]. For example, the investigator may believe, either from past experience or professional judgment, that a reported incidence of some characteristic less than 1/4 or higher than 3/4 would be highly unlikely. Then [a,b] could be set to [.25,.75] and the two corollaries applied.

4. APPLICATIONS TO SAMPLE DESIGNS

In this section, the method described in Section 3 is applied to the design of a sample survey whose purpose is to estimate the interviewer variance for some categorical variables. The number of interviewer assignments to interpenetrate is considered for both personal visit and telephone surveys.

Section 4.1 addresses the design of a new survey where little direct knowledge about the non-sampling errors for the survey items are available. The investigator is forced to use values of ρ_p which have been determined from other sources for similar questions and to make reasonable assumptions about the distributions of the p_i's. Section 4.2 considers the design of a telephone survey where there has been a pilot study. Estimates of the p_i's from an experimental telephone survey conducted by the U.S. Bureau of the Census serves as an example of a pilot survey. The data illustrates that a variety of distributions for the p_i's are plausible.

4.1 New Survey Design

Suppose an investigator is interested in designing an interpenetration experiment to estimate the interviewer variance σ_p^2 for a number of categorical questionnaire items. Two possible sample designs will be considered. One assumes that only interpenetration of pairs of interviewer assignments is feasible, so that the investigator must determine the minimum number of pairs (r) required to achieve a specified CV. The second design places no limit on the number of interviewer assignments which may be interpenetrated simultaneously, so that a minimum k, total number of interviewers, for a specified CV, is desired. The first type of design, as mentioned in Section 3, is likely to be required for personal visit interviews, while the second is more adaptable to telephone surveys. Interviewer sample sizes of m = 50 and/or m = 500 (which might correspond to an interviewer workload in a census) will be reported for each analysis.

Suppose that the only questionnaire items of interest to the investigator are ones for which the proportion of respondents recorded as falling in the categories are near 0.5, say between 0.3 and 0.7. The number of interpenetrated interviewers (k) or interviewer pairs (r) required to achieve a CV of 0.5 is investigated for three types of assumptions about the distribution of p_i. In Table 2, the bounds for r and k, when no assumptions about the 3^{rd} central moment or the range of p_i can be made, are shown . The last row of each subtable gives the value of r or k which would be required if the response to the questionnaire item were continuous and the normality assumptions needed for the validity of (10) were appropriate. The bounds are reported for likely values of ρ_p (or ρ_b for the last line), .01 to .10, and for μ_p near 0.5. (Only values of $\mu_p \le 0.5$ are necessary since the table is symmetric around $\mu_p = 0.5$.)

As discussed in Section 3, these bounds can generally be improved if certain information about the distribution of p_i is known. Suppose that the investigator in this example believes that the distribution of p_i is nearly symmetric. That belief can be translated into a restriction that μ_{p3} be close to 0. Two definitions of close were considered. For the first, whose results are reported in Table 3.a, we required that μ_{p3} be restricted to the 1/3 of its potential range (as given by (11)) that is closest to 0. So, for example, if (11) yields a possible range for μ_{p3} of (-c, +c), the restricted range is (-c/3, +c/3). Likewise, if (11) yields a possible range of (c_1,c_2), where $c_2 \geq c_1 \geq 0$, then the restricted range would be $(c_1, c_1 + (c_2 - c_1)/3)$. For Table 3.b, the stringent assumption that μ_{p3} be exactly 0 was required. While this requirement is weaker than the analogous one used for the continuous random ANOVA model (i.e., normality of the b_i's), it is stronger than can generally be justified. It is included here just to show how much could be gained by perfect knowledge of the third moment of p_i.

Restriction of the support of p_i from [0,1] to some subinterval [a,b] also results in improved bounds for the survey design parameters r and k. In Tables 3.c and 3.d, respectively, results are shown for restricted ranges [a,b] = [.1,.9] and [.25,.75], when no assumptions are made about μ_{p3}.

To be sure of achieving the stated CV for any possible distribution of p_i meeting the selected criteria, the investigator must choose r or k as the upper bound given in the table for the chosen pair of parameter values, μ_p and ρ_p. It is clear from Table 2 that there exist distributions for p_i which would make detection of small levels of interviewer correlation impossible due to the large number of interviewers which would be required to do so. Table 3 shows that a reduction in these upper bounds can usually be obtained by a restriction placed on either the support or the third central moment of p_i. Despite this, the maximum possible r and k are still so large for some parameter pairs (μ_p, ρ_p) that detection of small interviewer correlation would remain difficult, if not impossible, for many surveys. However, the application of both types of assumptions simultaneously is much more helpful in reducing the upper bounds of r or k, as is demonstrated by Table 4.

Several important observations which are generally true are illustrated by Tables 2 through 4:

(1) Detection of small values of σ_p^2 is potentially very difficult for categorical questionnaire items, much more difficult than for items having continuous normally distributed responses.

(2) The sample design parameter, r or k, which would be determined using the normal assumptions (i.e., equation (10)) is almost always near the lower, or most optimistic, end of its possible range under the categorical model assumption. This indicates that a serious underestimation of the number of interviewers needed in the interpenetration experiment can occur if (10) is used inappropriately for categorical variables.

(3) μ_p affects the upper bound of the variance of $\hat{\alpha}$ and thus of the required r or k. Estimation of interviewer variance becomes potentially more difficult (i.e., requires larger numbers of interviewers for the extreme p_i

distributions) as μ_p moves away from 0.5. This situation contrasts with the normal model case, where $V(\hat{CC})$ is not a function of the item mean.

(4) The lower bound of the interval for the required r or k is unaffected by restriction of the support of p_i, and is affected only slightly by the requirement of near symmetry. The only exception to this occurs when the support of p_i is restricted to such a point that the value of μ_{p3} producing the interval's lower bound is incompatible with the specified μ_p and ρ_p. For example, in the unrestricted case, those distributions having $\mu_{p3} = 0$ are the ones which allow the lower bounds for r and k. But there does not exist a distribution meeting the simultaneous restrictions $\mu_{p3} = 0$, $[a,b] = [.25, .75]$, $\mu_p = .3$, and $\rho_p = .10$. So the lower bound for r for any distribution for p_i meeting the latter three of these restrictions occurs when μ_{p3} takes its smallest possible value, which can be seen from (A.2) to be .0078, and is shown in Table 3.d to be 21.

(5) The upper bound of the interval for r or k is reduced most for distributions having μ_p near .5 by restriction of the support of p_i. For distributions having μ_p further from .5, the most reduction is gained by restricting μ_{p3}.

(6) An increase in interviewer assignment size m, over the range considered (50 to 500) is very helpful for detecting small σ_p^2 ($\rho_p = .01$), but makes little difference if detection of larger σ_p^2 is sufficient. This occurs

because the upper bound of (9) is dominated by its first term, which is not a function of m, when ρ_p is large.

(7) Changing the sample design from complete to pairwise interpenetration loses much efficiency when the distribution of p_i is of the type leading to values of r or k near their lower bound. For example, we see from Table 2.a that the minimum number of interviewers which would be required to achieve a CV of 0.5 for $\hat{\sigma}_p^2$ when ρ_p = .01 under the pairwise interpenetration design is 134 (2 x 67), while only 64 would be required for that p_i distribution if complete interpenetration were possible. By contrast, the largest possible required number of interviewers is similar for the two designs; for this example, the required number of interviewers is 500 (2 x 250) for pairwise and 428 for complete interpenetration.

4.2 Survey Design Following a Pilot Study

When data from a survey using similar procedures and asking similar questions is available, the investigator may have more information about the distribution of the p_i's to use in designing an interpenetration study. In this section, data for four questionnaire items from an experimental telephone survey illustrate how interviewer characteristics may differ from one questionnaire item to another. Two of these examples are then used to demonstrate how the methods of Section 3, combined with the descriptive information about the distribution of the p_i's can be employed by the investigator to determine the number of interviewers and the assignment size required. Since a telephone survey is under consideration, we assume that our object is to determine k, the number of interviewer assignments to be completely interpenetrated, needed to achieve the desired precision for an estimator of interviewer variance.

The data for these examples were obtained from an experimental random digit dial telephone survey of employment (called RDD-I) conducted by the U.S. Bureau of the Census in 1982 (Mulry-Liggan and Chapman 1982, Mulry-Liggan 1983). The questionnaire used was nearly identical to that used in the CPS. The survey was conducted in seven two-week periods (called replicates) and the data shown here were collected in the last three of these replicates. The interviewers' assignments were interpenetrated so that estimates of the correlated component could be obtained, as well as estimates of p_i (called $\hat{p}_i$) for individual interviewers. $\hat{p}_i$ is simply the observed proportion of interviewer i's assignment which belonged to the category being considered. Approximately 15 interviewers participated in the experiment, each of whom handled 39 or more cases for the items in these examples.

The first example considered the questionnaire item which recorded whether or not the respondent was willing to provide a complete address. The proportion of households whose addresses could be obtained was of interest to the designers of this survey. There was evidence from the survey data of positive interviewer variance for this item, and examination of the $\hat{p}_i$'s shows that their range was quite small, from a low of 0.62 to a high of 0.80. Larger numbers of interviewers had $\hat{p}_i$ values falling near the ends of that interval than near the middle, so that the histogram of the $\hat{p}_i$'s was U-shaped.

Contrasting with this were the results from another response characteristic measured, the proportion of respondents answering a direct salary question, before probing for salary category began. This item, too, showed evidence of positive interviewer variance, but the range of estimated p_i values was large, from a low of 0.23 to a high of 0.93. The observed distribution of these $\hat{p}_i$ values appeared more unimodal than that of the address question, with most of the values falling in an interval close to the average. These observations suggest that the interviewer population characteristics may really differ for these two items, even though they seem to be of a similar nature, both having to do with item response rate.

As a final example, we consider two possible response categories to the employment status question, "What were you doing most of last week, working or something else?" These two responses are coded on the questionnaire as "Unable to work" and "Other". The proportion of households having at least one member classified by an interviewer as "Unable to work" varied among the interviewers from a low of 0 to a high of only .13 and was very nearly symmetric. (The overall proportion so classified was .075.) By contrast, the proportion of households having a member classified as "Other" decreased as the survey progressed, but the histogram of the interviewer $\hat{p}_i$'s consistently retained the same characteristics. For the three replicates considered here, the overall proportions were 0.20, 0.11, and 0.05. For each replicate, however, two interviewers had discrepantly large $\hat{p}_i$ values (0.37 and 0.39 for the first replicate considered, 0.21 and 0.30 for the second, and 0.12 and 0.13 for the third), while the remaining $\hat{p}_i$ values covered only a small range.

The apparent change in the distribution of the p_i's during these three replicates appears to be the result of a learning process, although the interviewers had been working for eight weeks by the beginning of the first of these replicates. There were changes in the interviewer training and supervision as the experiment progressed, however, including the instigation of quality circles where the interviewers shared experiences. It is likely that this interaction eventually resulted in their behavior becoming more uniform. (Hopefully, it simultaneously resulted in more accurate classification, but that can't be demonstrated.) This is an illustration of what is meant by the claim that telephone interviewing from a centralized facility might provide better interviewer control than decentralized interviewing.

We now illustrate how these observations about the data from RDD-I could be used informally by the investigator for planning a design for a survey to estimate the interviewer variance for the address question and for the "Other" category of the labor force question. For the address question, the investigator might feel safe in restricting the range [a,b] to [.5,.9], an interval somewhat longer than that observed in the experimental survey. A further restriction of the 3rd central moment of p_i to be near 0 might be acceptable to the investigator as well on account of the symmetry of the histogram of $\hat{p}_i$'s.

Table 5 shows the bounds of the required k for several values of the assignment size m for CV = 0.5 when μ_p equals its estimate from RDD-I (which was 0.7) and when μ_{p3} is restricted to a small interval near 0. Examination of similar tables for $\mu_p = .65$ and .75 showed little deviation in either the bounds of k or the effect of varying m.

If the investigator wants to be certain of meeting the precision criteria for estimating very small values of σ_p^2 (e.g., $\rho_p = .01$), he or she must be prepared to interpenetrate more interviewer assignments than would be required for a continuous normally distributed response item. However, for larger values of σ_p^2, the entire interval for k may lie below that required for a normal variable. For example, from (10) with m = 25 and $\rho_p = .05$, we see that k = 32 interviewer assignments would have to be interpenetrated to achieve CV = 0.5. Table 5 illustrates that increasing the assignment size m reduces the number of interviewers required up to a point. That point is higher for the lower values of ρ_p.

To estimate k for the "Other" category of the labor force question, the investigator has to judge which of the situations in the three replicates is analogous to the new survey. If there is no planned interaction among interviewers, such as the quality circles for example, the investigator might decide that the range of p_i's cannot be more severely restricted than to [0, .4] or [0, .5]. If the survey is well-established and the interviewers experienced, he or she might feel comfortable further restricting [a,b], for example to [0, .2].

Since the histogram of $\hat{p}_i$'s always appears skewed to the right, restricting μ_{p3} to be non-negative is also a reasonable assumption. The resulting bounds for k when m = 50 and for μ_p between .05 and .20 are shown in Table 6. The ranges for k displayed there illustrate again the advantage of specifying as small an interval [a,b] as possible. By using the interval [0, .4] instead of [0, .5], there is, on the average, a 35% reduction in the upper bound for k. Table 6 also illustrates that the investigator can make the assumptions so restrictive that there does not exist a bounded random variable which satisfies them.

5. CONCLUSIONS

This paper has shown that the application of methods which may be appropriate for designing an experiment to estimate the correlated component of response error, or in general, any variance component, for continuous normal variables may be dangerously inappropriate for categorical variables. The examples of Section 4.1 have illustrated that usually there exist distributions of p_i for which far greater numbers of interviewers would be required to achieve the same CV that is attained more easily for the continuous normal models; in fact, the number of interviewers required for normal variables is often close to the lowest possible number which could be required for any distribution of p_i in the categorical case. Some of these extreme distributions of p_i are highly unlikely, however, at least for the application to interviewer variance estimation. If they can be ruled out by placing restrictions on the skewness and/or support of the p_i distribution, the upper bound for r or k can be reduced considerably, making the method more useful as a survey or experimental planning tool.

The investigator may be able to select the largest possible value required for r or k to ensure that the sample design has adequate precision for the interviewer variance estimates for important categorical items. On the other hand, if even the lower bound calls for an impossibly large number of interviewers, he or she is warned that for that item, at least, interviewer variance cannot be estimated sufficiently accurately with the available resources.

The investigator is left with a different problem if, after the best information about the distribution of p_i is utilized, the maximum number of interviewers available for inclusion in the interpenetration experiment falls between the lower and upper bound called for by this method. He or she then must realize that if the experiment proceeds, there is a risk of being unable to estimate the desired parameters adequately. If this path is chosen, however, the investigator at least receives the side benefit of having the opportunity to collect more information about the distribution of p_i, so that better prediction intervals may be possible for the next experiment. Alternatively, a modification of the sample design, such as increasing the interviewer assignment size m or the number of interviewers interpenetrated simultaneously, might be called for.

Another potential benefit of this method is that it might be used as an aid to variance estimation for CC. Without assuming some distribution for p_i, one is left with only empirical variance estimation methods, such as the jackknife, for evaluating the precision of $\hat{CC}$ after the data are collected. When large numbers of questionnaire items are being considered, these methods can become very expensive. The bounds for $V(\hat{CC})$ provided by the corollaries can serve as a cheap screening device for identifying items showing some evidence of positive interviewer variance. The investigator may then decide to produce empirical estimates of the variance for just those items. Besides the cost, another problem with the use of empirical variance estimation methods is that they are difficult to implement for estimators of the correlated component in some complex sample designs, whereas a "boundable" expression for $V(\hat{CC})$ (such as (9)) may be easily obtained for some such designs.

With an accumulation of information about interviewer behavior (and thus about the p_i distributions) for certain types of questionnaire items, or about whatever random effect is of interest in a variance components model, more accurate assessment of sample size requirements may be possible. If enough information is available, (9) may be directly usable; otherwise, we may at least learn more precisely what restrictions on the p_i distribution are acceptable for specific important items.

APPENDIX

This section is designed to state results and provide formulas which will enable the reader to easily derive upper and lower bounds for $V(\hat{CC})$ and for survey design parameters for their own survey using knowledge of specific questionnaire items.

1. Corollary 1 (Brockett and Cox 1984): Let X be a random variable having range [a,b] with $EX = \mu$ and $V(X) = \sigma^2$. Then for any function h for which $h^{(3)}(x) \geq 0$,

$$h(a)\pi + h(c)(1-\pi) \leq Eh(X) \leq h(d)\eta + h(b)(1-\eta), \qquad \text{(A.1)}$$

where

$$\pi = \frac{\sigma^2}{\sigma^2 + (a-\mu)^2}, \qquad \eta = \frac{(b-\mu)^2}{\sigma^2 + (b-\mu)^2}$$

$$c = \mu - \sigma^2/(a-\mu), \text{ and } \qquad d = \mu - \sigma^2/(b-\mu).$$

2. Corollary 1 can be used to show that the upper and lower bounds for the 3^{rd} central moment of X, μ_3, are

$$\sigma^2 [(a-\mu)^2 - \sigma^2] / (a-\mu) \leq \mu_3 \leq \sigma^2 [(b-\mu)^2 - \sigma^2] / (b-\mu). \qquad \text{(A.2)}$$

3. Corollary 2 (Brockett and Cox 1984): Let X be a random variable having

range [a,b] with $EX = \mu$, $V(X) = \sigma^2$ and $\mu_3 = E(X-\mu)^3$ known. Then for any function h for which $h^{(4)}(x) \geq 0$,

$$h(c_1)\pi + h(c_2)(1-\pi) \leq Eh(X) \leq h(a)n_1 + h(d)n_2 + h(b)(1-\eta_1-\eta_2),$$

where

$$d = \frac{\mu_3 - (a+b-2\mu)}{\sigma^2 - \mu(1-\mu)} + \mu, \qquad \eta_1 = \frac{\sigma^2 + (d-\mu)(b-\mu)}{(b-a)(d-a)},$$

$$\eta_2 = \frac{(b-\mu)(a-\mu) + \sigma^2}{(d-b)(d-\mu)}, \qquad c_1 = \mu + \frac{\mu_3 - \sqrt{\mu_3^2 + 4\sigma^2}}{2o^2}$$

$$c_2 = \mu + \frac{\mu_3 + \sqrt{\mu_3^2 + 4\sigma^2}}{2\sigma^2}, \qquad \pi = 1/2 + \frac{\mu_3}{\sqrt{\mu_3^2 + 4\sigma^2}}.$$

REFERENCES

BAILEY,L., MOORE, T., AND BAILAR, B.B. (1978), "An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample," Journal of the American Statistical Association,73, 16-23.

BAILAR, B.A. and BIEMER, P.P. (1984), "Some Methods for Evaluating Nonsampling Error in Household Censuses and Surveys," W.G. Cochran's Impact on Statistics, edited by P.S.R.S. Rao, John Wiley & Sons.

BIEMER, P.P., and STOKES, S.L. (1985),"Optimal Design of Interviewer Variance Experiments in Complex Surveys," Journal of the American Statistical Association,80,158-166.

BROCKETT, P.L., and COX, S.H. (1984),"Insurance Calculations Using Incomplete Information," Working Paper 83/84-2-22, The University of Texas at Austin, Department of Finance.

DeVYLDER, F. (1982), "Best Upper Bounds for integrals with respect to measures allowed to vary under conical and integral constraints," Insurance: Mathematics and Economics, 1(2), pp. 109-130.

DeVYLDER, F. (1983), "Maximization, under Equality Constraints, of a Functional of a Probability Distribution," Insurance: Mathematics and Economics, Vol. 2, 1-16.

KENDALL, M. and STUART, A. (1977), The Advanced Theory of Statistics, Volume 1, Macmillan Publishing Co., NY.

KISH, L. (1962), "Studies of Interviewer Variance for Attitudinal Variables," Journal of the American Statistical Association, 57, 92-115.

McLEOD, A., and KROTKI, K.P. (1979), "An Empirical Investigation of an Improved Method of Measuring Correlated Response Variance," Survey Methodology, 5, No. 2, 59-78.

MAHALANOBIS, P.C. (1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," Journal of the Royal Statistical Society, 109, 325-370.

MULRY-LIGGAN, M.H. and CHAPMAN, D.W. (1982), "The Sampling Design of the Bureau of the Census Random Digit Dialing Experiment," Proceedings of the Survey Research Section of the American Statistical Association.

MULRY-LIGGAN, M.H. (1983), "A Comparison of Random Digit Dialing Survey and the Current Population Survey," Proceedings of the Survey Research Section of the American Statistical Association.

SEARLE, S.R. (1971), Linear Models, John Wiley & Sons, Inc., NY.

Table 1 Distributions for p and Resulting Precision of $\hat{CC}$

	f(p)	μ_p	ρ_p	β_{p1}	β_{p2}	$R_{c,n}$
a.	$\begin{cases} .9 & \text{if } .45 \le p \le .55 \\ .25 & \text{if } .55 < p \le .75 \text{ or } .25 \le p < .45 \\ 0 & \text{otherwise} \end{cases}$	.500	.013	0	8.91	1.21
b.	Beta(.5,.5,.418,.582)*	.500	.013	0	1.50	0.93
c.	Beta(37,37,0,1)	.500	.013	0	2.92	0.99
d.	Beta(1,10,0,.5)	.045	.040	2.31	5.78	1.64
e.	Beta(1,20,0,.5)	.024	.022	2.99	7.07	1.84
f.	Beta(1,10,.25,.75)	.295	.008	2.31	5.78	1.08

* Beta(r,s,a,b) denotes a random variable having density function
$f(p) = (p - a)^{r-1}(b - p)^{s-1} / B(r,s)(b - a)^{r+s+1}$ for $a \le p \le b$.

Table 2 Bounds for Required r and k to Achieve CV = 0.5
[a,b] = [0,1], μ_{p3} Unrestricted

	Number of Interviewers							
	r (Interpenetration of pairs)				k (Complete interpenetration)			
μ_p \ ρ_p	.01	.025	.05	.10	.01	.025	.05	.10
				a. m = 50				
.5	(67,250)	(22,94)	(12,47)	(8,24)	(64,428)	(19,162)	(9,78)	(6,37)
.4	(67,348)	(22,132)	(12,65)	(8,32)	(64,624)	(19,238)	(9,114)	(6,53)
.3	(67,514)	(22,198)	(12,98)	(8,49)	(64,957)	(19,371)	(10,181)	(6,86)
Normal	72	26	16	12	73	27	17	13
				b. m = 500				
.5	(8,204)	(6,83)	(5,43)	(5,23)	(6,397)	(5,157)	(4,77)	(4,37)
.4	(8,302)	(6,121)	(5,61)	(5,31)	(6,593)	(5,233)	(4,113)	(4,53)
.3	(8,469)	(6,188)	(5,94)	(5,47)	(6,926)	(5,366)	(4,179)	(4,86)
Normal	12	10	9	9	13	11	10	10

Table 3 Bounds for Required r and k to Achieve CV = 0.5, m = 50

	Number of Interviewers							
	r (Interpenetration of pairs)				k (Complete interpenetration)			
μ_p \ ρ_p	.01	.025	.05	.10	.01	.025	.05	.10
	a. μ_{p3} restricted to middle third of its range, [a,b]=[0,1]							
.5	(67,250)	(22,94)	(12,47)	(8,24)	(64,428)	(19,162)	(9,78)	(6,37)
.4	(67,277)	(22,104)	(12,52)	(8,27)	(64,483)	(19,183)	(9,88)	(6,42)
.3	(67,312)	(22,118)	(12,58)	(8,29)	(64,553)	(19,210)	(9,101)	(6,47)
	b. μ_{p3} = 0, [a,b] = [0,1]							
.5	(67,250)	(22,94)	(12,47)	(8,24)	(64,428)	(19,162)	(9,78)	(6,37)
.4	(67,250)	(22,94)	(12,47)	(8,24)	(64,428)	(19,161)	(9,77)	(6,36)
.3	(67,249)	(22,92)	(12,45)	(8,23)	(64,426)	(22,159)	(10,75)	(6,34)
	c. [a,b] = [.1,.9], μ_{p3} unrestricted							
.5	(67,183)	(22,67)	(12,33)	(8,18)	(64,296)	(19,109)	(9,51)	(6,24)
.4	(67,262)	(22,97)	(12,48)	(8,24)	(64,453)	(19,169)	(9,80)	(6,36)
.3	(67,397)	(22,151)	(12,75)	(8,37)	(64,723)	(19,277)	(10,134)	(6,63)
	d. [a,b] = [.25,.75], μ_{p3} unrestricted							
.5	(67,111)	(22,38)	(12,19)	(8,10)	(64,152)	(19,51)	(9,23)	(6,10)
.4	(67,162)	(22,57)	(12,28)	(8,14)	(64,253)	(19,89)	(9,40)	(6,17)
.3	(67,255)	(23,94)	(17,46)	(21,23)	(64,438)	(22,163)	(19,77)	(30,35)

Table 4 Bounds for Required r and k to Achieve CV = 0.5
m = 50, [a,b] = [.25,.75]

	Number of Interviewers							
	r (Interpenetration of pairs)				k (Complete interpenetration)			
μ_p \ ρ_p	.01	.025	.05	.10	.01	.025	.05	.10
	a. μ_{p3} restricted to middle third of its range							
.5	(67,111)	(22,38)	(12,19)	(8,10)	(64,152)	(19,51)	(9,23)	(6,10)
.4	(67,118)	(22,40)	(12,19)	(8,11)	(64,166)	(19,55)	(9,23)	(6,11)
.3	(67,134)	(23,51)	(12,29)	(21,22)	(64,196)	(22,76)	(19,43)	(30,32)
	b. μ_{p3} = 0							
.5	(67,111)	(22,38)	(12,19)	(8,10)	(64,152)	(19,51)	(9,23)	(6,10)
.4	(67,104)	(22,34)	(12,16)	*	(64,138)	(19,44)	(9,17)	*
.3	(67,71)	*	*	*	(64,71)	*	*	*

* There is not a bounded random variable satisfying the specified conditions.

Table 5. Bounds for Required k to Achieve CV = 0.5, μ_p = 0.7, [a,b] = [.5, .9], μ_{p3} restricted to middle third of its range

	ρ_p			
m	.01	.025	.05	.10
25	(189,257)	(45,70)	(18,28)	(9,12)
50	(64,134)	(19,44)	(9,19)	(6,9)
75	(37,106)	(13,37)	(7,17)	(5,8)
100	(25,95)	(10,34)	(6,15)	(5,7)

Table 6. Bounds for Required k to Achieve CV = 0.5, $m = 50$, $\mu_{p3} > 0$

	ρ_p			
μ_p	.01	.025	.05	.10
	a. [a,b] = [0,.2]			
0.15	(66,78)	*	*	*
0.10	(68,127)	(21,38)	(10,15)	(6,7)
0.05	(75,326)	(21,118)	(11,53)	(10,23)
	b. [a,b] = [0,.4]			
0.20	(65,171)	(20,57)	(10,24)	(6,11)
0.15	(65,298)	(20,107)	(10,49)	(6,21)
0.10	(68,824)	(21,317)	(10,153)	(6,73)
	c. [a,b] = [0,.5]			
0.30	(64,165)	(19,56)	(10,25)	(6,11)
0.20	(65,298)	(20,107)	(10,49)	(6,21)
0.10	(68,824)	(21,317)	(10,153)	(6,73)

* There is not a bounded random variable satisfying the specified conditions.