Journal of Official Statistics
Vol. 3, No. 4. 1987, pp. 389–401

On the Design of Interpenetration Experiments for Categorical Data Items

S. Lynne Stokes and Mary H. Mulry[1]

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 the range of possible values 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 nonsampling error; Markov-Krein Theorem.

1. Introduction

Errors introduced in the measuring, editing, or coding of responses in a sample survey contribute substantially to the bias and/or variance of the estimators obtained from the sample. 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 increase in the variance of estimators of means and totals can be particularly severe. To see why this is true, let y_{ij} denote the recorded response of the jth unit in the ith operator's assignment, where $i = 1,\ldots,k$ and $j = 1,\ldots,m$. Then

$$\mathrm{Var}(\bar{y}) = (\sigma^2/km)[1 + (m-1)\varrho],$$

where $\bar{y} = \sum_i \sum_j y_{ij}/km$, $\sigma^2 = \mathrm{Var}(y_{ij})$ and $\varrho\sigma^2 = \mathrm{Cov}(y_{ij}, y_{ij'})$. So if the operator's assignment is large, even a small value of ϱ can substantially increase $\mathrm{Var}(\bar{y})$. This increase is known as the correlated component of response variance.

[1] S. Lynne Stokes is Assistant Professor, Department of Management Science and Information Systems and Center for Statistical Sciences, University of Texas, Austin TX 78712, U.S.A. Mary H. Mulry 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.

Because of its potentially large effect, the U.S. Bureau of the Census has included estimation of the correlated component of response variance due to interviewers, also called interviewer variance, as part of its evaluation programs for the last four decennial censuses. As a result of the first of those studies in 1950, the operation of the 1960 and subsequent censuses was changed to predominantly self- enumeration so that the impact of the interviewer errors would be reduced.

Now there is interest within the Census Bureau and other survey research organizations in estimating interviewer variance for key items in sample surveys as well. In addition to providing a way to adjust the variance of estimates, it may help identify items subject to large interviewer errors, which could then be mitigated by questionnaire redesign or increased interviewer training.

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 interviewer. Thus each interviewer's sample is randomly distributed over the entire population, requiring increased travel for each interviewer.

In personal visit sample surveys, in contrast to censuses, the sample units may be sparsely distributed. Even when the population is divided into subpopulations, each covering less area, the added expense in time and money of the increased interviewer travel may be enormous. Even locating groups of interviewers whose usual enumeration areas are in close enough proximity that interpenetration is feasible may be difficult. For that reason, careful planning of the sample design for interpenetration experiments is important. If sufficiently precise estimation of interviewer variance would require a larger number of interpenetrated interviewer assignments or a larger number of sample units than can be afforded, then this should be known before, rather than after the data are collected.

To solve the sample size problem completely, the researcher must know the distribution of the estimator of interviewer variance. Samplers rarely have such complete information about their estimators when planning a survey, however. They generally must make sample size decisions from knowledge of the behavior of the mean and variance of their estimator only.

Several estimators of interviewer variance have been suggested (Kish (1962), Fellegi (1974), Biemer and Stokes (1985), and most of them are actually ANOVA or MINQUE (Hartley, Rao, and La Motte (1978)) estimators of variance components from random or mixed ANOVA models. The mean and variance of these estimators are known, at least for balanced designs, when all variance components and the response variable itself are normally distributed. Most survey data are categorical, however, and when analysis is performed, each category is treated as a dichotomous variable. Little has been written about the distribution of the estimators of variance components in that case.

Williams (1982) and Anderson and Aitkin (1985) have suggested estimators of a parameter measuring interviewer variability (but which is not interviewer variance) for dichotomous variables based on maximum likelihood. The advantage of their methods is that other factors besides the interviewer that may affect the responses, such as geographical area, may be included in the model. However, neither of them supplies expressions for the variance of their estimators which could be used for planning sample designs.

In other studies of interviewer variance, empirical variance estimation techniques, such as the jackknife (McLeod and Krotki (1979)) and ultimate cluster methods (Bailey, Moore, and Bailar (1978)) have been used for evaluating the precision of estimators of interviewer variance on categorical data items. These methods are not useful for the planning of sample designs for interpenetration studies, however. For interpenetration studies, of course, measures of the precision of the estimator are needed before the data collection process begins.

The problem of determining the number of interviewers needed to participate in an interpenetration experiment whose goal was to estimate interviewer variance for dichotomous items led to the development of the methods discussed in this paper. Here, the approach taken is to at least determine bounds for the variance of the usual estimator of interviewer variance for dichotomous items from balanced interpenetration experiments, without making unsubstantiated distributional assumptions about the errors interviewers make. These bounds can then be used to determine a lower bound for the number of interviewers needed to achieve a specified precision for the estimate of interviewer variance. As more is known about the way in which interviewers make errors, the researcher may be willing to make more stringent distributional assumptions, thus narrowing the interval of possible values for the estimator's variance.

A striking observation from the analysis which follows is the discrepancy in the sample sizes which may be required between the categorical and continuous (normal) data cases. Since researchers sometimes use the theory of the normal model for designing interpenetration experiments for surveys whose key items are categorical, this discrepancy could lead to estimators with insufficient precision.

In Section 2, two commonly used estimators of interviewer variance are presented, along with their variances, for the continuous and categorical data models. In Section 3, methods for obtaining bounds on the variances of these estimates for dichotomous items are suggested. The methods are illustrated in Section 4 by numerical example. A discussion follows in Section 5.

2. The Estimator and Its Variance

Assume that k interviewer assignments are interpenetrated and that each interviewer has m assigned units. Let y_{ij} denote the response of the jth unit in interviewer is assignment. When the characteristic being observed is membership in a category, so that the response y_{ij} is limited to 0 or 1, Bailar and Biemer (1984) model the mechanism causing the correlated errors in the following way. For each category, an interviewer can make two types of errors: η_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 probability. Then (η_i, θ_i) is considered to be a random vector associated with the ith interviewer. But consideration of individual characteristics of η_i and θ_i can be avoided by defining a new random variable $p_i = E(y_{ij}|i) = Pr[y_{ij} = 1|i]$ to be the conditional probability that interviewer i records a randomly chosen unit as belonging to the category. Then we may write

$$y_{ij} = \mu + (p_i - \mu) + \varepsilon_{ij}, \qquad (2.1)$$

where $\mu = Ep_i = E[E(y_{ij}|i)] = E(y_{ij})$, which implies that $E(\varepsilon_{ij}) = 0$. Since y_{ij} is limited to either 0 or 1, ε_{ij} is restricted to $-p_i$ or $1 - p_i = q_i$, with conditional probabilities q_i and p_i respectively, for each i. Thus $V(\varepsilon_{ij}) = E[V(\varepsilon_{ij}|i)] + V[E(\varepsilon_{ij}|i)] = Ep_iq_i = \mu(1-\mu) - \phi_1$, where $\phi_1 = V(p_i)$. Similarly, one can show that $\mathrm{Cov}(\varepsilon_{ij},p_i) = \mathrm{Cov}(\varepsilon_{ij},\varepsilon_{ij}') = 0$. When written in this form, it is obvious why estimators of variance components are appropriate, even in the categorical case, for estimating the interviewer variance ϕ_1 and the intra-interviewer correlation $\varrho = \mathrm{Cov}(y_{ij}, y_{ij'})/\mathrm{Var}(y_{ij'}) = \phi_1/\mu(1-\mu)$.

The ANOVA estimator from the design just described, in which k interviewers randomly split mk units, with y_{ij} denoting the jth unit in interviewer is assignment, $i = 1,\ldots,k$; $j = 1,\ldots, m$ is

$$\hat{\phi}_1 = (s_b^2 - s_w^2)/m, \tag{2.2}$$

where

$$s_b^2 = [m/(k-1)] \sum_i (\bar{y}_i - \bar{y})^2, \tag{2.3}$$

$$\text{and } s_w^2 = [k(m-1)]^{-1} \sum_i \sum_j (y_{ij} - \bar{y}_i)^2. \tag{2.4}$$

In the dichotomous case, (2.3) and (2.4) can be rewritten as

$$s_b^2 = [m/(k-1)] \sum_i (\bar{p}_i - \bar{p})^2$$

$$\text{and } s_w^2 = [k(m-1)]^{-1} \sum_i \bar{p}_i(1-\bar{p}_i),$$

where $\bar{p}_i = \sum_j y_{ij}/m$, $\bar{p} = \sum_i \bar{p}_i/\mathrm{k}$.

This estimator can be used only when it is feasible for all interviewers to have units chosen from the total population. For example, it was used by Kish (1962) in a survey of employees of a factory. It might also be used in some telephone surveys since assignment of a random subsample of the total population to each interviewer does not generate the additional costs of increased travel.

For personal visit interviews, however, interpenetration experiments require increased amounts of travel by all 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 of interpenetrated pairs r, rather than by increasing k. Then the estimator of interviewer variance is taken to be

$$\tilde{\phi}_1 = \sum_i \hat{\phi}_{1i}/r, \tag{2.5}$$

where $\hat{\phi}_{1i}$ is the estimate of interviewer variance (computed from (2.2) with $k = 2$) for the ith pair.

The classical one-way random ANOVA model has the same form as (2.1), but y_{ij} is generally assumed to be a continuous random variable and $V(\varepsilon_{ij}|i)$ is assumed to be homogeneous, say equal to ϕ_0 over all groups. If, in addition, both p_i and ε_{ij} were assumed to be normal, the variance of the ANOVA estimate of ϕ_1, $\hat{\phi}_1$ as given by (2.2), is known (e.g., from Searle (1971, p. 474) to be

$$V(\hat{\phi}_1) = 2(k-1)^{-1} [\phi_1{}^2 + 2\phi_1\phi_0/m + \phi_0^2/m^2] + 0(m^{-3}). \tag{2.6}$$

If the assumption of normality of p_i is discarded, we have from Appendix 1

$$V(\hat{\phi}_1) = k^{-1}[\mu_4 - ((k-3)/(k-1))\phi_1^2] + 4\,\phi_1\phi_0/(k-1)m + 2\phi_0^2/(k-1)m^2 + 0(m^{-3}), \tag{2.7}$$

where μ_4 is the 4th central moment of the p_i s. It is clear that when the kurtosis of p_i, $\beta_2 = \mu_4/\sigma_1{}^2 > 3$, the expression in (2.7) is greater than that in (2.6). When $\beta_2 < 3$, it is smaller, and when $\beta_2 = 3$, as in the normal distribution, (2.6) and (2.7) are identical. Ideally, then, the kurtosis of the distribution of the p_i s should be considered when planning sample designs for estimating interviewer variance for continuous variables.

When y_{ij} is dichotomous, we have seen that the distribution assumed for the variance component p_i determines the distribution of ε_{ij}, so that their distributions cannot be specified separately. Furthermore, the conditional variance of the error term, $V(\varepsilon_{ij}|i) = p_iq_i$, is not homogeneous across interviewers unless $\phi_1 = 0$. Therefore, (2.6) and (2.7) are not appropriate expressions for $V(\hat{\phi}_1)$ and so cannot be used for survey design as they can in the continuous case. It can be shown, however, following the method described in Appendix 1, that for y_{ij} following model (2.1),

$$V(\hat{\phi}_1) = k^{-1}[\mu_4 - ((k-3)/(k-1))\phi_1^2] + 4\phi_1 E(p_iq_i)/(k-1)m + 2E^2(p_iq_i)/(k-1)m^2 + 4A_p/km + 2B_p/km^2 + 0(m^{-3}), \quad (2.8)$$

where $A_p = E\{p_iq_i[(p_i - \mu)^2 - \phi_1]\}$ and $B_p = V(p_iq_i) + E\ [p_iq_i(q_i - p_i)(p_i - \mu)]$. The first three terms of (2.8) are identical to those of (2.7), with $E(p_iq_i)$ replacing ϕ_0. A_p and B_p are functions of the first four moments of p_i and can be either positive or negative. So the discrepancy between $V(\hat{\phi}_1)$ for the normal and the dichotomous variable models (i.e., between (2.6) and (2.8)) may arise not only from p_is non-normality, but also from A_p and/or B_p being non-zero.

From (2.5) we have

$$V(\tilde{\phi}_1) = \sum_i V(\hat{\phi}_{1i})/r, \quad (2.9)$$

where $V(\hat{\phi}_{1i})$ can be obtained from (2.6), (2.7), and (2.8) with $k = 2$.

Since no other methods have been available, researchers have used (2.6) to plan sample designs even when the variables of interest are categorical. Table 1 shows that this procedure can lead to either over- or under-estimation of $V(\hat{\phi}_1)$. It gives, for several specified distributions of p_i, the ratio of $V(\hat{\phi}_1)$

Table 1. Distributions for p and resulting precision of $\hat{\phi}_1$

	$f(p)$	μ	ϱ	β_1	β_2	$R_{c,n}$
a.	.9 if $.45 \leq p \leq .55$ .25 if $.55 < p \leq .75$ or $.25 \leq p < .45$ 0 otherwise	.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 $a \leq p \leq b$.

for the dichotomous data model (from (2.8)) to that of a normal model (from (2.6)) having comparable variance components ($\phi_0 = E(p_i q_i)$), where $m = 50$ and $k = 2$. This ratio is denoted by $R_{c,n}$. The distributions represented in the table were chosen to have a variety of shapes and to have intra-interviewer correlations covering a range of typical values for ϱ. Distribution (c) is very close to normal and has $R_{c,n}$ close to 1. From (a) and (b) we see that symmetric distributions can have $R_{c,n}$ values either greater or smaller than 1, depending on whether the kurtosis of p_i is large or small. The positively skewed distributions, i.e., those having $\beta_1 = \mu_3^2/\phi_1^{3/2} > 0$ (d, e, and f) all have $R_{c,n} > 1$. These distributions have large β_2 as well, however, since $\beta_2 > 1 + \beta_1$ for any random variable (see, e.g., Kendall and Stuart (1977, p. 95)). In general, it appears from the table as though the kurtosis of p_i determines whether the precision of $\hat{\phi}_1$ is higher or lower for dichotomous variables than for the case where all variance components are normally distributed. We noted earlier that when y_{ij} is continuous and $V(y_{ij}|i)$ homogeneous, a similar statement holds.

The criterion which is often used for determining an adequate sample size for an unbiased estimator is that of achieving for it a specified coefficient of variation ($CV(\hat{\phi}_1) = [\mathrm{Var}(\hat{\phi}_1)/\phi_1^2]^{1/2}$). For example, in a 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 $\hat{\phi}_1$. If y_{ij} were a continuous variable and all the assumptions required for the validity of (2.6) held, then the required k is that satisfying

$$(CV)^2 = 2\{1 + 2[(1-\varrho)/\varrho]/m + [(1-\varrho)/\varrho]^2/m^2\}/(k-1) + 0(m^{-3}), \quad (2.10)$$

where $\varrho = \phi_1/(\phi_1 + \phi_0)$. The values of k determined from (2.10) to achieve $CV = 0.5$ for some specified values of ϱ and m are shown in Table 2.

The number of pairs of interpenetrated interviewer assignments, r, needed to achieve a specified CV for $\tilde{\phi}_1$ can also be determined from (2.9) and (2.10) by setting $k = 2$. The results from this computation for $CV = 0.5$ are also reported in Table 2.

If the sample designer plans to use (2.8) to determine the needed number of interviewers to participate in an interpenetration experiment to estimate ϕ_1 for categorical items, he or she must first make assumptions about the first four moments of p_i. Even an assumption of normality alone for p_i is not sufficient, since in the dichotomous case, $CV(\hat{\phi}_1)$ depends on $Ep_i = \mu$, which it does not for the continuous model. Furthermore, an assumption of normality for p_i is not natural, since p_i denotes a probability and, as such, is restricted to the interval [0,1].

Table 2. Required r and k to achieve CV = 0.5 for normal model

	Number of interviewers							
	r (Interpenetration of pairs)				k (Complete interpenetration)			
ϱ	.01	.025	.05	.10	.01	.025	.05	.10
$m = 50$	72	26	17	12	73	27	17	13
$m = 500$	12	10	9	9	13	11	10	10

For most applications in which estimates of variance components for categorical variables are needed, the sample designer is not likely to have much information about the distribution of p_i. On the other hand, he or she is likely, at least, to have information about the range of values for the first two moments of p_i for a given item. A range for μ will be known from the reported incidence of similar characteristics in previous studies, and a less precise range for ϱ, and thus for $\phi_1 = \mu(1-\mu)\ \varrho$, can be given from knowledge gained from previous interpenetration studies, such as those reported by Kish (1962) or Groves and Magilavy (1987). For example, we know that for most demographic items, ϱ is less than 0.01, 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.

In Section 3, a method for obtaining bounds, rather than exact values, for $V(\hat{\phi}_1)$ and thus for $V(\tilde{\phi}_1)$, is suggested. The goal of this method is to obtain these bounds while requiring knowledge only of the boundedness of p_i, its first two moments, and as little speculative information as possible about its distribution. From these bounds, the method can determine bounds for k or r.

3. Variance Bounds in the Categorical Model

Bounds for $V(\hat{\phi}_1)$ or $V(\tilde{\phi}_1)$ can be found by using corollaries of the Markov-Krein Theorem (DeVylder (1982, 1983)). These corollaries, which are stated in Appendix 2, provide tight upper and lower bounds on the expected value of certain functions of a bounded random variable whose first several moments are known. The corollaries actually produce distributions which achieve each of those bounds, implying that the width of the interval cannot be reduced without excluding values for a possible distribution. The random variables having the most extreme moments are discrete ones with two or three point masses.

An upper and lower bound for the third central moment of p_i, μ_3, is found first by applying Corollary 1 of the Appendix to $h(p_i) = (p_i - \mu)^3$. The corollary's assumptions are satisfied if the survey planner has some information about the first two moments of p_i. This interval, which can be obtained from (A.2), is

$$\phi_1(\phi_1 - \mu^2)/\mu \leqslant \mu_3 \leqslant \phi_1[(1-\mu)^2 - \phi_1]/(1-\mu). \tag{3.1}$$

To use the same approach for obtaining an upper and lower bound for $V(\hat{\phi}_1)$ given that the first 3 moments of p_i are known, one must be able to express $V(\hat{\phi}_1)$ from (2.8) as $Eh(p_i)$, where $h^{(4)}(p_i) \geqslant 0$. Such a function exists and is given in (A.3) of Appendix 2. Then Corollary 2 yields bounds for $V(\hat{\phi}_1)$ which require knowledge of μ_3, which may in turn be obtained from (3.1). With the bounds thus obtained for $V(\hat{\phi}_1)$, the problem of determining the survey design parameters required to achieve a specified CV for $\hat{\phi}_1$ or $\tilde{\phi}_1$ can be addressed.

Using this approach, however, yields bounds for k or r which are, in most cases, too wide to be useful. This is disturbing, especially since the Markov-Krein corollaries provide bounds which are tight; i.e., there actually exist distributions for p_i which lead to the endpoints of the intervals and thus to the most extreme sample sizes. But those extreme distributions are unlikely candidates for the behavior of the population of interviewers. For example, the upper bound for $V(\hat{\phi}_1)$ provided by the corollaries is achieved by distributions which place positive probability at each endpoint of p_is support; i.e., at 0 and 1, and at only one point in the interior of the interval. But the 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 is unlikely that interviewer p_is could achieve the extremes of both 0 and 1 on the same questionnaire item.

By reducing the class of distributions from which p_i may be chosen, the range of possible values of $V(\hat{\phi}_1)$ or $V(\tilde{\phi}_1)$ and thus of k or r is reduced. The class is reduced by imposing realistic constraints on the distributions. Two types of constraints are considered in this paper.

The first places restrictions on the skewness of the distribution of p_i, and thus on μ_3. The researcher may impose this type of restriction if, for example, he or she feels that the tendency and amount of overreporting by interviewers of membership in a category is about the same as that of underreporting. Then μ_3 would be restricted to be near 0 and Corollary 2 could be applied to obtain the needed bounds. Conversely, if the larger interviewer errors tend to be in the direction of overreporting membership in a category, then μ_3 might be restricted to positive values.

A second type of constraint considered here is that of restricting 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 does not occur. Then [a,b] could be set to [.25,.75] and the two corollaries applied. Both of these restrictions result in narrowed bounds for $V(\hat{\phi}_1)$ and thus for k or m. Examples of the use of each type of constraint are given in Section 4.

4. Examples

In this section, the method described in Section 3 is illustrated by numerical examples. Suppose an investigator is interested in designing an interpenetration experiment to estimate the interviewer variance ϕ_1 for a number of categorical questionnaire items. We assume that no direct information about interviewer behavior for the survey items is available, so that he or she is forced simply to make reasonable assumptions about the distribution of the p_is.

The two possible sample designs discussed earlier will be considered here. 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. This means 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. An interviewer sample size of $m = 50$, which may be appropriate for the interviewer workload in a telephone survey conducted from a centralized facility, will be assumed.

Suppose that the questionnaire items of interest to the investigator are ones for which the proportion of respondents recorded as belonging to 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 restrictions of the two types discussed in Section 3, as well as a combination of the two. These restrictions are discussed in the following paragraphs.

We first assume that the distribution of p_i is approximately symmetric. This belief can be translated into a restriction that μ_3 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_3 be restricted to the 1/3 of its potential range (as given by

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/ϱ	.01	.025	.05	.10	.01	.025	.05	.10
	a. μ_3 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. $\mu_3 = 0$, $[a,b]$ = [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_3 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_3 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)

(3.1)) that is closest to 0. For the second, whose results are reported in Table 3.(b), the stringent assumption that μ_3 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 random component), it is stronger than can generally be justified.

Second, we assume that the range of p_i is restricted. 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 (d), respectively, results are shown for restricted ranges [a,b] = [0.1,0.9] and [0.25,0.75], when no assumptions are made about μ_3.

Third, we assume both near symmetry and restricted range for p_i. The results of the simultaneous assumption of both restrictions

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/ϱ	.01	.025	.05	.10	.01	.025	.05	.10
	a. μ_3 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. $\mu_3 = 0$							
.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.

is shown in Table 4.

Several important observations can be made from Tables 3 and 4.

1. Detection of small values of ϕ_1 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 (shown in Table 2) is always nearer the lower, or more optimistic, end of its possible range than to its upper bound, under the categorical model assumption. This indicates that a serious underestimation of the number of interviewers needed in the interpenetration experiment can occur if (2.6) is used inappropriately for categorical variables.

3. μ affects the upper bound of the variance of $\hat{\phi}_1$ and thus of the required r or k. Estimation of the interviewer variance becomes potentially more difficult (i.e., requires larger numbers of interviewers for the extreme p_i distributions) as μ moves away from 0.5. This contrasts with the normal model case, where $V(\hat{\phi}_1)$ is not a function of the item mean.

4. The lower bound of the interval for the required r or k is unaffected by the restriction of the support of p_i, and is affected only slightly by the requirement of near symmetry. The only exception occurs when the support of p_i is restricted to such a point that the value of μ_3 producing the interval's lower bound is incompatible with the specified μ and ϱ.

5. The upper bound of the interval for r or k is reduced most for distributions having μ near 0.5 by restriction of the support of p_i. For distributions having μ further from 0.5, the most reduction is gained by restricting μ_3.

6. Changing the sample design from complete to pairwise interpenetration loses much efficiency when the distribution of p_i leads to values of r or k near their lower bound.

5. Conclusions

This paper has shown that the application of methods which may be appropriate for designing an experiment to estimate interviewer variance, or in general, any variance component, for continuous normal variables may be dangerously inappropriate for categorical variables. The examples of Section 4 have illustrated that usually there exist distributions of p_i that require far greater numbers of interviewers 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 sometimes 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 planning tool. The method described in Section 3 and the formulae available in the Appendix make the task of deriving bounds for k or r for new sets of restrictions a straightforward one for the survey designer.

The investigator may be able to select the largest possible value required for r or k yielded by the method of Section 3, once reasonable restrictions are assumed for p_i. If this is possible, he or she will be assured 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 resources available.

The investigator is left with a different problem if the maximum number of interviewers available for 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 not being able to estimate the desired parameters adequately.

If data from a previous interpenetration experiment involving similar questionnaire items is available, it might be used to aid the investigator in his or her choice of restrictions for the distribution of p_i. An understanding of interviewer behavior, and thus knowledge of the distribution of p_i, may be improved by observation of the individual p_is. An investigation of a formal method for utilizing this information is currently underway.

Appendix 1

To derive (2.7) and (2.8), we first note from (2.2) that

$$V(\hat{\phi}_1) = [V(s_w^2) + V(s_b^2) - 2\mathrm{Cov}(s_w^2, s_b^2)]/m^2. \quad \text{(A.1)}$$

This expression can be evaluated by writing the first term, for example, as $E[V(s_w^2|p_i, i = 1,\ldots,k)] + V[E(s_w^2|p_i, \mathrm{i} = 1,\ldots,k)]$ and the third term as $E[\mathrm{Cov}(s_w^2, s_b^2|p_i, i = 1,\ldots,k)] + \mathrm{Cov}[E(s_w^2|p_i, i = 1,\ldots,k), E(s_b^2|p_i, i = 1,\ldots,k)]$. In the continuous case, the conditional expected values, variances, and covariances above are derived under the assumption that $y_{ij}|i, j = 1,\ldots,m$, are independent $N(p_i, \phi_0)$ random variables. In the dichotomous case, $y_{ij}|i, j = 1,\ldots,m$ are independent Bernoulli(p_i) random variables. In both cases, then, the conditional distributions of the observations are independent, but not identically distributed. After taking due note of this fact, each of the terms in (A.1) can be derived by repeated use of properties of these random variables, yielding (2.7) in the continuous case and (2.8) in the dichotomous case.

Appendix 2

This section is designed to state results and provide formulas which will enable the reader to easily derive upper and lower bounds for $V(\hat{\phi}_1)$ 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 $h^{(3)}(x) \geqslant 0$,

$$h(a)\pi + h(c)(1-\pi) \leqslant Eh(X) \leqslant h(d)\,\eta + h(b)(1-\eta),$$

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. By letting $h(x) = (x-\mu)^3$, Corollary 1 can be used to show that the upper and lower bounds for the 3rd central moment of X, μ_3, are

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

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 $h^{(4)}(x) \geqslant 0$,

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

where

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

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

$$\eta_2 = \frac{(b-\mu)(a-\mu) + \sigma^2}{(d-b)(d-\mu)},$$

$$c_1 = \mu + \frac{\mu_3 - \sqrt{\mu_3^2 + 4\sigma^2}}{2\sigma^2},$$

$$c_2 = \mu + \frac{\mu_3 + \sqrt{\mu_3^2 + 4\sigma^2}}{2\sigma^2},$$

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

4. Corollary 2 can be used to find the upper and lower bounds for $V(\hat{\phi}_1)$ when the first 3 moments of X, μ, σ^2, and μ_3, are known, by letting

$$h(x) = \frac{1}{k}\left(1 - \frac{4}{m} + \frac{6}{m^2}\right)x^4$$

$$+ \frac{1}{k}\{-\mu(4\mu_3 + 6\mu\sigma^2 + \mu^3) - \frac{k-3}{k-1}\sigma^4$$

$$+ \frac{2}{m}[2(1+2\mu)$$

$$-(5+2\mu)/m][\mu_3 + 3\mu\sigma^2 + \mu^3]\}$$

$$+ \frac{2}{mk(k-1)}[2\nu\sigma^2 + \nu^2/m]$$

$$+ \frac{2}{mk}\{\mu[(2+\mu)\mu_2' + \mu^2]$$

$$+ \frac{1}{m}[(2+3\mu)\mu_2' - \mu^2]\}, \qquad \text{(A.3)}$$

where $\nu = \mu(1-\mu) - \sigma^2$ and $\mu'_2 = \mu^2 + \sigma^2$.

6. References

Anderson, D.A. and Aitkin, M. (1985): Variance Component Models with Binary Response: Interviewer Variability. Journal of the Royal Statistical Society, B, 47, pp. 203–210.

Bailey, L., Moore, T., and Bailar, B.A. (1978): An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample. Journal of the American Statistical Association, 73, pp. 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, Inc., New York.

Biemer, P.P. and Stokes, S.L. (1985): Optimal Design of Interviewer Variance Experiments in Complex Surveys. Journal of the American Statistical Association, 80, pp. 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, 2, pp. 1–16.

Fellegi, I. (1974): An Improved Method of Estimating the Correlated Response Variance. Journal of the American Statistical Association, 69, pp. 496–501.

Groves, R.M. and Magilavy, L.J. (1987): Measuring and Explaining Interviewer Variance in Telephone Surveys. Public Opinion Quarterly, Vol. 51.

Hartley, H.O., Rao, J.N.K, and La Motte, L.R. (1978): A Simple Synthesis – Based Method of Variance Component Estimation, Biometrics, 34, pp. 233–242.

Kendall, M. and Stuart, A. (1977): The Advanced Theory of Statistics, Vol. 1. Macmillan Publishing Co., New York.

Kish, L. (1962): Studies of Interviewer Variance for Attitudinal Variables. Journal of the American Statistical Association, 57, pp. 92–115.

McLeod, A. (1978), and Krotki, K.P. (1979): An Empirical Investigation of an Improved Method of Measuring Correlated Response Variance. Survey Methodology, 5, No. 2, pp. 59–78.

Mahalanobis, P.C. (1946): Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, 109, pp. 325–370.

Mulry-Liggan, M.H. and Chapman, D.W. (1982): The Sampling Design of the Bureau of the Census Random Digit Dialing Experiment. American Statistical Association, Proceedings of the Survey Research Section, pp. 133–138.

Mulry-Liggan, M.H. (1983): A Comparison of Random Digit Dialing Survey and the Current Population Survey. American Statistical Association, Proceedings of the Survey Research Section, pp. 214–219.

Searle S.R. (1971): Linear Models. John Wiley & Sons, Inc., New York.

Williams, D.A. (1982): Extra-binomial Variation in Logistic Linear Models. Applied Statistics, 31, pp. 144–148.

Received July 1986
Revised August 1987