Characteristics of Nonrespondents in the Schools and Staffing Surveys' School Sample**

Pao-Sheng Shen, Randall J. Parmer and Andre I. Tan, U.S. Bureau of the Census
Pao-Sheng Shen, U.S. Bureau of the Census, Washington, D.C. 20233

1. Introduction

In the 1991 Schools and Staffing Survey (SASS), there are some nonrespondents. One strategy for adjusting for nonresponse is to estimate the variables of interest with a poststratification estimator. Each respondent observation is weighted by the inverse of the respondent proportions of the observations in its cell, which is defined on the auxiliary variables such as grade level, enrollment and urbanicity. In doing this, one is implicitly modeling the nonresponse mechanism by assuming that the probability of nonresponse may vary among cells but not within cells. Hence, it is important to choose suitable adjustment cells such that the response probabilities of individuals within cells are as homogeneous as possible. This approach is discussed in detail by Schaible (1979).

The first objective of our research is to identify the auxiliary variables correlated with nonresponse and make recommendations for nonresponse adjustment cells. The second objective is to identify subpopulation with low response rate where field resources can be concentrated to improve the overall response rate. The data used are from a sample of 8995 public schools and 2741 list frame private schools.

Section 2 of the article presents a brief description of the 1991 SASS. In Section 3, we discuss the methodology. To identify the auxiliary variables correlated with nonresponse, adjusted Chi-square tests are used for testing the correlation between the auxiliary variables and response status. For estimation of response rates in subpopulations, due to the small subpopulation sizes, procedures depending on the distribution created by the sampling plan are unstable or not available. The logit-estimates, which are simply the application of the "pseudo" maximum likelihood estimate (pseu-MLE) from Roberts, Rao and Kumer (1987), were used to estimate the response rates for subpopulations of public schools. For private schools, subpopulation sample sizes are too sparse to support the existence of a unique pseu-MLE. Hence, empirical Bayasian-logit estimates which are based on the "pseudo" maximum posterior estimate (pseu-MPE) defined in Section 3.3 were used as alternatives. Section 4 contains a summary of our results and conclusions.

2. The 1991 Schools and Staffing Survey (SASS)

2.1 Frame Construction

The 1991 Schools and Staffing Surveys consists of a school, a teacher, and for public schools a Local Education Agency or school district survey. Public schools were identified on the Common Core of Data or CCD. This CCD was matched to the previous SASS public school sampling frame. Non-matches from the previous frame were included with the CCD to make up the public school sampling frame for 1991. Public schools were stratified by state, grade level, and Indian/non-Indian.

The private schools were selected from a list frame, constructed by matching multiple lists obtained from private school organizations, State Departments of Education, and a private vendor. This frame is thought to include 80-90% of private schools. To increase the coverage of the survey, an area frame was constructed by selecting 120 PSUs, consisting of counties or groups of counties. Within these sample counties, lists of schools were obtained from local sources, such as yellow pages, churches and fire marshals. These lists were unduplicated with the list frame. The remaining schools, not matching to the list frame, make up the area frame.

2.2 Design

Public schools were stratified by state, grade level, and Indian/non-Indian. Probabilities of selection were computed, proportional to the square root of the number of teachers in the school conditioned on the 1988 selection. The probabilities were adjusted to obtain the desired proportion of overlapping schools from 1988. Approximately 9900 public sample schools were selected systematically within each stratum.

Private schools were stratified by affiliation, grade level, and census region for the list frame, and by PSU and grade level for the area frame. Probabilities of selection were computed and adjusted similarly to the public schools. Approximately 3300 private schools were selected, systematically within each stratum.

2.3 Data Collection

School questionnaires were mailed to schools. They were asked to fill them out and mail them back to the Census Bureau. After four weeks, if the school hadn't responded, we sent out a second questionnaire. If after three more weeks the school hadn't responded, we called them and attempted to complete the interview by telephone. Schools still not responding by telephone were classified as noninterviews.

2.4 Estimation

Schools' probabilities of selection were adjusted for school merges and other situations that would affect the probability of selection. The inverse of the probability of selection became the basic weight. This basic weight was adjusted to account for noninterviews using noninterview adjustment cells. A ratio adjustment was also applied which adjusted the characteristics of the sample schools to the characteristics of the whole sample frame.

3. Methodology

3.1 Testing

The response status (yes or no) is considered to be the response variable. The continuous auxiliary variables are divided into 2 -5 groups. The standard Chi-Squared tests for independence (denoted as X_I^2 when auxiliary variables are not used for stratification) or tests for homogeneity (denoted as X_H^2 when auxiliary variables are used for stratification) are not appropriate due to the complex sample design of SASS. As a result, some adjustments that take into account the design are necessary in order to make valid inferences from survey data. Rao and Scott (1984) derived a first-order correction denoted by δ to the standard Chi-Squared test which requires the knowledge of only the cell design effects (deffs) and the deffs for marginals provided the model admits a direct solution to likelihood equations under multinomial sampling. These results are applicable to the test results in our study.

However, because of a shortage of information on cell deffs, only X^2_I and some of X_H^2 were adjusted. The reason is that the empirical study by Holt, Schott and Ewings (1980) indicated that the distortion of nominal significance level is substantially smaller with X_I^2 than with X_H^2. The deffs for adjusting tests were obtained based on the estimated variance of all the individual cells using 48 pseudo-replicates originated by the U.S. Bureau of the Census (Simmons and Baird (1968)).

** This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

3.2 Subpopulation Estimation

Some variables of interest that we identified as correlated with nonresponse, were chosen to construct subpopulations. By the levels of the variables chosen, the populations of public schools and list frame private schools were divided into 20 and 48 subpopulations respectively. In certain subpopulations, sample sizes are too small to have accurate estimates by using the standard methods based on the selection probabilities. One strategy is to borrow information across subpopulations by using an unsaturated logit regression model. Due to difficulties in obtaining appropriate likelihood functions for our design, "pseudo" maximum likelihood estimates (pseu - MLE) (Roberts, Rao and Kumar (1987)) can be used to replace maximum likelihood estimates (MLE) of regression coefficients. This strategy was implemented on the estimation of response rates for subpopulations of public schools and the estimates based on the regression model are referred to as logit-estimates. However, with only 2741 samples for list frame private schools, the observed response or nonresponse frequencies are zero for some subpopulations. These conditions may make pseu-MLE not unique (Albert and Anderson (1984)). To solve the existence problem, an empirical Bayes approach was proposed and Bayesian-logit estimates were used as alternatives.

The approach is described in Section 3.3. The goodness-of-fit of the model was based on a likelihood ratio test corrected by an upper bound on δ proposed by Rao and Scott (1987). The upper bound can be obtained using information on cell deffs and marginal deffs. The test is conservative and applicable to the model not admitting a direct solution to the likelihood equation.

3.3 An Empirical Bayesian Approach for Subpopulation Estimation of Binary Data from Complex Sample Surveys

Without loss of generality, suppose that the population is partitioned into I×J subpopulations according to factor A_i ($i=1,\ldots,I$) by factor B_j ($j=1,\ldots,J$). Let $P=(P_{11},\ldots,P_{IJ})'$ where P_{ij} denotes the proportions that schools in the ijth subpopulation are respondents. Let $\hat{N}_{ij}$ denote the 1991 SASS survey estimate of the ijth subpopulation total $\hat{N}_{ij1}$ the corresponding estimate of response frequencies. With large $\hat{N}_{ij}$ and reasonably large frequencies, $\hat{N}_{ij1}$, the ratio estimate

$$\hat{P}_{ij} = \frac{\hat{N}_{ij1}}{\hat{N}_{ij}}$$

is often used to estimate P_{ij}. When the data are too few $\hat{P}_{ij}$ can be very unstable. In this situation, it seemed much more appropriate to borrow information across subpopulations by using an unsaturated logistic model. A logistic regression model for the response rate P_{ij} of the ijth subpopulation is given by $P_{ij}=f_{ij}(\beta)$, where

$$\log\left[\frac{f_{ij}(\beta)}{1-f_{ij}(\beta)}\right] = X'_{ij}\beta \qquad (3.1)$$

In (3.1) X_{ij} is an S-vector of known constants derived from the factor levels and β is an S-vector of unknown parameters. The pseu-MLE of β can be obtained from solving the following "pseudo" likelihood equations through iterative calculations:

$$X'D(\hat{\omega})f(\hat{\beta}) = X'D(\hat{\omega})\hat{P} \qquad (3.2)$$

where $X = (X_{11},\ldots,X_{IJ})$ is an S×IJ matrix of rank S, $D(\hat{\omega}) = diag(\hat{\omega}_{11},\ldots,\hat{\omega}_{IJ})$,

$$\hat{\omega}_{ij} = \frac{\hat{N}_{ij}}{\sum_{ij}\hat{N}_{ij}} = \frac{\hat{N}_{ij}}{\hat{N}}$$ is the estimated subpopulation relative size ω_{ij},

$\hat{P} = (\hat{P}_{11},\ldots,\hat{P}_{IJ})'$ and $f(\hat{\beta})=(f_{11}(\hat{\beta}),\ldots,f_{IJ}(\hat{\beta}))'$.

Under the assumption that $n^{1/2}(\hat{P} - f(\beta))$ converge in distribution to N(0, V), the estimated asymptotic covariance matrix of $f(\hat{\beta})$ is (Robert, Rao and Kumar (1987))

$$\hat{V}_f = [D(\hat{\omega})]^{-1}\hat{\Delta}X\hat{V}_\beta X'\hat{\Delta}[D(\hat{\omega})]^{-1} \qquad (3.3)$$

where

$\hat{\Delta} = diag(\hat{\omega}_{11}\hat{f}_{11}(1-\hat{f}_{11}),\ldots,\hat{\omega}_{IJ}\hat{f}_{IJ}(1-\hat{f}_{IJ}))$ and

$\hat{V}_\beta = n^{-1}(X'\hat{\Delta}X)^{-1}[X'D(\hat{\omega})\hat{V}D(\hat{\omega})X](X'\hat{\Delta}X)^{-1}$

where $\hat{V}$ denote the survey estimate of the covariance matrix V.

However, when any of $\hat{N}_{ij1}$ or $\hat{N}_{ij0}$ $(=\hat{N}_{ij}-\hat{N}_{ij1})$ is zero, a unique pseu-MLE $\hat{\beta}$ may not exist for the regression model considered. A sufficient condition for the existence of a unique $\hat{\beta}$ is $0 < \hat{N}_{ij1} < \hat{N}_{ij}$ for all i, j (Albert and Anderson (1984)). The empirical Bayesian approach developed next solves the existence problem and has an intuitively appealing interpretation. First, we model the distribution of P_{ij} as a Beta distribution with parameters a_i and b_i. That is, we will assume the hierarchical prior $P_{ij} \overset{iid}{\sim} Beta(a_i,b_i)$ for $i=1,\ldots,I$ and $j=1,\ldots,J$, so that the P_{ij} have density function

where $B(a_i,b_i)$ is the complete beta function.

$$h(P_{ij}) = [B(a_i,b_i)]^{-1} P_{ij}^{a_i-1} (1-P_{ij})^{b_i-1}$$

Note that the hierarchial prior model is equivalent to grouping like subpopulations (with the same level of factor A) into strata (different levels of factor A) and modeling the subpopulations within a stratum to have a common distribution. For list frame private schools, based on the data 1988 and 1991 SASS, the variation of response rate within each association is smaller than the variation of response rate among the associations. Also, there were reasonably large sample sizes in each association. Hence association was used as factor A and the combination of the other variables was used as factor B to construct the hierarchical model.

Next, we estimate a_i and b_i for $i=1,...,I$ from the marginal distribution of data by integrating the following "pseudo" likelihood equation with respect to P_{ij}.

$$\prod_{ij} \left[(P_{ij})^{n\left(\frac{\hat{N}_{ijl}}{\hat{N}}\right) + a_i - 1} (1-P_{ij})^{n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right) + b_i - 1} [B(a_i,b_i)]^{-1} \right] \tag{3.4}$$

The result of integration of (3.4) is

$$\prod_{ij} \left[B\left(n\left(\frac{\hat{N}_{ijl}}{\hat{N}}\right) + a_i,\ n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right) + b_i \right) [B(a_i,b_i)]^{-1} \right] \tag{3.5}$$

The expression in equation (3.5) is maximized under the constraint $a_i>0$, $b_i>0$ using numerical method to obtain the pseu-MLE of a_i and b_i denoted by $\hat{a}_i$ and $\hat{b}_i$ for $i=1,...I$. The value of β obtained by solving the following equation will be called the pseu-MPE of β and denoted by $\tilde{\beta}$ and the estimator $f(\tilde{\beta})$ will be referred as empirical Bayesian-logit estimator.

$$X'D(\tilde{\omega})f(\tilde{\beta}) = X'D(\tilde{\omega})\tilde{P} \tag{3.6}$$

where $\tilde{P} = (\tilde{P}_{11},...,\tilde{P}_{IJ})'$, $\tilde{P}_{ij} = \dfrac{n\left(\frac{\hat{N}_{ijl}}{\hat{N}}\right) + \hat{a}_i}{n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right) + \hat{a}_i + \hat{b}_i}$

$D(\tilde{\omega}) = diag(\tilde{\omega}_{IJ},...,\tilde{\omega}_{IJ})$,

$f(\tilde{\beta}) = (f_{11}(\tilde{\beta}),...f_{IJ}(\tilde{\beta}))'$ $\quad \tilde{\omega}_{ij} = \dfrac{\left[n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right) + \hat{a}_i + \hat{b}_i \right]}{n + J\sum_i (\hat{a}_i + \hat{b}_i)}$

The pseu-MPE always exists since $\hat{a}_i > 0$ *and* $\hat{b}_i > 0$ for all i.

REMARK.

First, note that $\tilde{P}_{ij}$ can be written as

$$\left[\pi_{ij} \hat{P}_{ij} + (1-\pi_{ij}) \left(\frac{\hat{a}_i}{\hat{a}_i + \hat{b}_i} \right) \right] I_{(\hat{N}_{ij}>0)} + \left(\frac{\hat{a}_i}{\hat{a}_i + \hat{b}_i} \right) I_{(\hat{N}_{ij}=0)}$$

where

$$\pi_{ij} = \frac{\hat{\tau}_i^2}{\hat{\tau}_i^2 + \hat{\sigma}_{ij}^2} \qquad \hat{\tau}_i^2 = \frac{\hat{a}_i \hat{b}_i}{(\hat{a}_i + \hat{b}_i)^2 (\hat{a}_i + \hat{b}_i + 1)}$$

$$and \quad \hat{\sigma}_{ij}^2 = \frac{\hat{P}_{ij}(1-\hat{P}_{ij})}{n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right)}$$

Note that $\hat{\tau}_i^2$ is the estimated variance of the i^{th} stratum based on the superpopulation model (prior distribution imposed on P) and $\hat{\sigma}_{ij}^2$ is an intuitively estimated sampling variance for the ij^{th} subpopulation when the subpopulation sample size is not zero. The smaller the sampling variance relative to stratum variance, the more weight $\hat{P}_{ij}$ gets. Just as intuitively reasonable, for large relative sampling variance, which can be defined as infinity when sample size is zero, little weight should be given to $\hat{P}_{ij}$, and there should be a borrowing of strength from the other observations in the same stratum. Secondly, under the model (3.1), it follows that β has a prior $\pi(\beta)$ in the form

$$\pi(\beta) \propto \prod_{ij} [f_{ij}(\beta)]^{a_i} [1-f_{ij}(\beta)]^{b_i}$$

Solving the Equation (3.6) is equivalent to maximizing the following "pseudo" posterior likelihood function with respect to β

$$\prod_{ij} [f_{ij}(\beta)]^{n\left(\frac{\hat{N}_{ijl}}{\hat{N}}\right) + \hat{a}_i} [1 - f_{ij}(\beta)]^{n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right) + \hat{b}_i}$$

The conditional asymptotic covariance of $f(\tilde{\beta})$ can be derived as follows:

LEMMA. Let β_0 denote the conditional expected value of $\tilde{\beta}$ when $P = P_0$.

Suppose that

(A) The conditional distribution of $n^{1/2}(\tilde{P} - f(\beta_0))$, as n tends to infinity, is normal with mean 0 and variance V_0 and,

(B) For all i,j, we have $\tilde{\omega}_{ij}-\omega^*_{ij}=o_p(1)$ where ω^*_{ij} 's are some design-dependent constants.

then the conditional asymptotic variance of $f(\tilde{\beta})$, denoted by $V_{\tilde{f}}$, is $V_{\tilde{f}} = [D(w^*)]^{-1}\Delta_0 \, X \, V_{\tilde{\beta}} \, X' \, \Delta_0 \, [D(w^*)]^{-1}$ where

$$V_{\tilde{\beta}} = n^{-1}(X'\Delta_0 X)^{-1}[X'D(\omega^*)V_0 D(\omega^*)X](X'\Delta_0 X)^{-1} \quad (3.7)$$

$D(\omega^*) = \mathrm{diag}(\omega^*_{11},\dots,\omega^*_{IJ})$ and
$\Delta_0 = \mathrm{diag}\{\omega^*_{11}f_{11}(\beta_0)[1-f_{11}(\beta_0)],\dots,\omega^*_{IJ}f_{IJ}(\beta_0)[1-f_{IJ}(\beta_0)]\}$
Proof:

Let $U(\beta,D(\tilde{\omega})) = \sum_{ij} U_{ij}(\beta,D(\tilde{\omega})) = \sum_{ij} X'_{ij}\tilde{\omega}_{ij}(\hat{P}_{ij} - f_{ij}(\beta))$.

By Equation (3.6), $U(\tilde{\beta},D(\tilde{\omega})) = 0$

Under the assumption (B),

$$U(\beta,D(\tilde{\omega})) = U(\beta,D(\omega^*)) + o_p(1) \textit{ for all } \beta \quad \text{as } n \to \infty$$

Now, treat $U(\beta, D(\omega^*))$ as a function of β only and denoted by $L(\beta)$. Regularity conditions are satisfied by $L(\beta)$ and as n is large $n^{1/2}(\tilde{\beta} - \beta_0)$ can, using a Taylor expansion, be approximated by

$$\left\{\left[\frac{\partial}{\partial\beta}L(\beta)\right]_{\beta=\beta_0}\right\}^{-1}L(\beta_0)=\left\{\sum_{ij}\left[-\frac{\partial}{\partial\beta}U_{ij}(\beta,D(\omega^*))\right]_{\beta=\beta_0}\right\}^{-1}\left[n^{1/2}\sum_{ij}U_{ij}(\beta_0,D(\omega^*))\right]$$
$$= (X'\Delta_0 X)^{-1}X'D(\omega^*)\,[n^{1/2}(\hat{P} - f(\beta_0))]$$

Under the assumption (A), it follows that $n^{1/2}(\tilde{\beta} - \beta_0)$, as n tends to infinity, converges in distribution to $N(0,V_{\tilde{\beta}})$.

Similarly, noting that

$$n^{1/2}(f(\tilde{\beta}) - f(\beta_0)) = \left[\frac{\partial f(\beta)}{\partial\beta}\right]_{\beta=\beta_0}[n^{1/2}(\beta - \beta_0)] = [D(\omega)]^{-1}\Delta_0 \, X \, [n^{1/2}(\tilde{\beta} - \beta_0)]$$

it follows that $n^{1/2}(f(\tilde{\beta}) - f(\beta_0))$, as n tends to infinity, converges in distribution to $N(0,V_{\tilde{f}})$.

Let $\hat{V}$ denote the survey estimate of the covariance matrix V_0 (given the prior parameters $\hat{a}_i$ and $\hat{b}_i$ for $i=1,\dots,I$). Then (3.7) can be estimated by

$$\hat{V}_{\tilde{\beta}} = (X'\tilde{\Delta}X)^{-1}[X'D(\tilde{\omega})\hat{V}D(\tilde{\omega})X](X'\tilde{\Delta}X)^{-1}$$

where $\tilde{\Delta} = \mathrm{diag}\;\{\tilde{\omega}_{11}f_{11}(\tilde{\beta})[1-f_{11}(\tilde{\beta})],\dots,\tilde{\omega}_{IJ}f_{IJ}(\tilde{\beta})[1-f_{IJ}(\tilde{\beta})]\}$

Similarly, the asymptotic covariance of $f(\tilde{\beta})$ can be estimated by

$$\hat{V}_f = [D(\tilde{\omega})]^{-1}\tilde{\Delta}X\hat{V}_{\tilde{\beta}}X'\tilde{\Delta}[D(\tilde{\omega})]^{-1} \quad (3.8)$$

In our study, the computer programs were written in SAS® to perform the required maximization of the logarithm of equation (3.5) to obtain the estimated prior parameters $\hat{a}_i$ and $\hat{b}_i$ for all i. Then SAS/CATMOD was used to obtain the $\hat{\beta}$ and $\tilde{\beta}$ for public schools and private schools respectively. Due to small sample sizes for certain subpopulations, a pseudo-replication scheme is not applicable to the estimation of V and V_b . One way around this is to aggregate, temporarily, some of the subpopulations of small sample sizes to the same group. In other words, define disjoint groups of subpopulations and implement a pseudo-replication scheme to estimate the covariance of groups. Assign the estimated group standard deviation to all subpopulations belonging to the same group. In our study, this strategy was used to obtain subpopulation design effects, $\hat{V}$ and $\tilde{V}$ and then $\hat{V}_f$ and $\tilde{V}_f$ were calculated.

4. Results and Conclusions

4.1 Testing

Table 1 and Table 2 illustrate the estimated deffs $\hat{\delta}$ and results of X^2_I, X^2_H, $X^2_I/\hat{\delta}$ and $X^2_H/\hat{\delta}$ for some auxiliary variables selected. From Table 1 and 2, we note that the deffs for public schools are much higher than those for list frame private schools. One explanation is that our design is not planned to reduce the variance of the estimation of response rate. However, it happened that for private schools, both grade level and association, which are strongly correlated with response status, were used for stratification, while for public schools, grade level, which is used for stratification, is weakly correlated with response status. Even though design effects for public schools are very high, it turned out that the size of the modified tests based on $X^2_I/\hat{\delta}$ was significant at $\alpha = .001$ for urbanicity and at $\alpha = .01$ for enrollment and the modified test based on $X^2_H/\hat{\delta}$ was significant for state at $\alpha = .001$. For private schools, the size of the modified tests based on $X^2_H/\hat{\delta}$ was significant at $\alpha = .001$ for grade level and association. The size of the modified tests based on $X^2_I/\hat{\delta}$ was significant at $\alpha = .001$ for affiliation and urbanicity and at about $\alpha = .10$ for enrollment.

Table 1
Test Results for Public Schools

Variables	df	δ	X_p^2	$*X_p^2/\delta$	X_l^2	$**X_l^2/\delta$
Grade Level	1	12.1	12.3	1.0	-	-
State	30	14.9	2115.6	141.9	-	-
Urbanicity	1	16.9	-	-	808.7	47.9
Enrollment	4	15.6	-	-	215.6	13.8

Table 2
Test Results for Private Schools

Variables	df	δ	X_p^2	X_p^2/δ	X_l^2	X_l^2/δ
Grade Level	1	2.2	343.0	155.9	-	-
Urbanicity	1	1.7	-	-	137.4	80.8
Enrollment	1	2.0	-	-	5.1	2.6
Affiliation	1	1.8	-	-	177.7	98.7
Association	4	2.5	904.6	361.8	-	-

*For testing homogeneity:

$$\delta = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (1-\hat{P}_j)\, \hat{d}_{j(i)}\, (1-n_{i+}/n)$$

where n is sample size, n_i+ is row margin, $\hat{P}_j = \sum_i (n_{i+}/n)\hat{P}_{j(i)}$, $\hat{P}_{j(i)}$ is the estimated cell proportions within the i^{th} row population and $\hat{d}_{j(i)} = n_{i+}\, \hat{V}(\hat{P}_{j(i)})/\hat{P}_j(1-\hat{P}_j)$ is the estimated deffs of $\hat{P}_{j(i)}$.

where $\hat{V}(\hat{P}_{j(i)})$ denotes the estimated variance of $\hat{P}_{j(i)}$.

**For testing independence:

$$\bar{\delta} = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (1-\hat{P}_{i+}\hat{P}_{+j})\hat{d}_{ij} \sum_{i=1}^{I} (1-\hat{P}_{i+})\hat{d}_i(v) - \sum_{j=1}^{J} (1-\hat{P}_{+j})\hat{d}_j(c)$$

where $\hat{P}_{i+}$ and $\hat{P}_{+j}$ are the estimated row and column marginals proportion, $\hat{d}_i(r)$ *and* $\hat{d}_j(c)$ are the estimated deffs of $\hat{P}_{i+}$ and $\hat{P}_{+j}$ respectively, and $\hat{d}_{ij}$ is the estimated deff of $\hat{P}_{ij}$, which is the estimated proportion for the ij^{th} cell.

4.2. Subpopulation Estimation

For public schools, the population was divided into 20 subpopulations by grade level, urbanicity and enrollment. Based on the unadjusted chi-square of each term, some interaction terms appear to be nonsignificant and are excluded from the full model. The following reduced model was chosen to explain the variation in the response rate.

$$V_{ijk} = \log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} \qquad (4.1)$$

where π_{ijk} denotes the response rate of the ijk^{th} subpopulation.
α_i denotes the effect of the i^{th} enrollment, $i=1, ..., 5$.
β_j denotes the effect of the j^{th} urbanicity, $j=1, 2$.
γ_k denotes the effect of the k^{th} grade level, $k=1$,
$(\alpha\beta)_{ij}$ denotes the interaction of the i^{th} enrollment by the j^{th} urbanicity.
$(\alpha\gamma)_{ik}$ denotes the interaction of the i^{th} enrollment by the k^{th} grade level.

Similarly, for list frame private schools, the population was divided in 48 subpopulations by association, grade level, urbanicity, and enrollment. The following model was chosen to explain the variation in the response rate.

$$V'_{ijkl} = \log\left(\frac{\pi'_{ijkl}}{1-\pi'_{ijkl}}\right) = \mu' + \alpha'_i + \beta'_j + \gamma'_k + \lambda'_l + (\alpha'\beta')_{ij} + (\alpha'\gamma')_{ik} + (\alpha'\lambda')_{il} + (\beta'\lambda')_{jl} + (\alpha'\gamma'\lambda')_{ikl} + (\beta'\gamma'\lambda')_{jkl} \qquad (4.2)$$

where π'_{ijkl} denotes the response rate of the $ijkl^{th}$ subpopulation
α'_i denotes the effect of the i^{th} association, $i=1, ..., 6$.
β'_j denotes the effect of the j^{th} grade level $j=1,2$.
γ'_k denotes the effect of the k^{th} urbanicity, $k=1,2$.
λ'_l denotes the effect of the l^{th} enrollment, $l=1,2$.
$(\alpha'\beta')_{ij}$ denotes the interaction of the i^{th} association by the j^{th} grade level,and so on.

For testing the goodness-of-fit of the model, the adjusted likelihood ratio test proposed by Rao and Scott (1987) were used. The adjustment is based on the upper bound on δ which requires the information of cell deffs (subpopulation deffs). The deffs of the subpopulation were estimated using 48 pseudo-replicates. The estimated deff for the i^{th} subpopulation is equal to

$$\hat{Var}(\hat{P}_i) \;/\; (n\hat{W}_i)^{-1}\, \hat{P}_i\, (1-\hat{P}_i)$$

where $\hat{P}_i$ is the estimated response rate for the i^{th} subpopulation
$\hat{Var}(\hat{P}_i)$ is the estimated variance of $\hat{P}_i$ using 48 pseudo-replicates
$\hat{W}_i$ is the estimated relative size for the i^{th} subpopulation
n is the total sample size.

For public schools, the upper bound on δ is estimated by the average deffs available (= 6.4) and multiplied by R_1/R_1-m_1, where R_1 is the number of subpopulations (= 20) and m_1 (= 15) is the number of parameters to be estimated for model (4.1). Hence the upper bound was estimated by (6.4)(20/5) = 25.7. The result for the adjusted likelihood ratio = (2.4)/25.7 = 0.09, which is not significant at the 5% level when compared to $X_5^2(0.05)=11.1$ Note that due to the high deffs for public schools, the test is very conservative.

Similarly, for list frame private schools, the upper bound on δ is estimated by the average deffs available (= 2.1) and multiplied by $R_2/R_2 - m_2$, where R_2 is the number of subpopulation (= 48) and m_2 (= 31) is the number of parameters to be estimated for model (4.2). Hence the upper bound was estimated by (2.1)(48/17) = 5.9. The result for the adjusted likelihood ratio = 40.1/5.9 = 6.8, which is not significant at the 5% level when compared to $X_{17}^2(0.05)=27.6$.

Based on model (4.1) and (4.2), the estimated response rate for subpopulations of public schools and private list frame schools are presented in Table 3 and Table 4 respectively. The corresponding estimated asymptotic standard deviations are also listed.

4.3. Conclusion

The empirical Baysian strategy used here for estimating response rate of subpopulations is a two-staged approach (one stage to estimate the prior, one stage to estimate the parameters given the estimated prior). The prior used is data-dependent. Although this strategy is not classical Baysian, it is in the spirit of an empirical-Bayesian procedure. This approach has the advantage of allowing information from all subpopulations to be used to provide estimates of response rate within each subpopulation. The disadvantage is that the computations are difficult. Under the hierarchical prior assumption, the estimated subpopulations' response rates were shrinked toward the marginal (association) response rate. The estimated asymptotic standard deviations did not include the uncertainty in the pseu-MLE of prior parameters. A possible remedy for this problem was suggested by Carlin and Gelfand (1991).

In summary, the variation of response rate for public schools is much smaller than that for private schools. For public schools, the nonresponse adjustment cells currently used by the U.S. Bureau of the Census are state by grade level by enrollment by urbanicity. Based on our results of testing, it seems to be a good choice. When further collapsing is necessary, cells can be collapsed with grade level first, enrollment second and urbanicity third. For private list frame, the nonresponse adjustment cells currently used by the U.S. Bureau of the Census are association by grade level by urbanicity. Based on the results of testing, it indicated that enrollment may also be a good candidate for creating nonresponse adjustment cells. If further collapsing is necessary, the cells can be collapsed with enrollment first, grade level second, urbanicity third and association fourth.

5. References

[1] Albert, A., and Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistics Regression Models, Biometrika, 71, 1-10.

[2] Carlin, P. Bradley and Gelfand, E. Alan (1991). A Sample Reuse Method for Accurate Parametric Empirical Bayes Confidence Interval. Journal of Royal Statistical Society. Sec. B 53, No. 1, 189-200

[3] Holt, D., Scott, A. J., and Ewings, P. D. (1980). Chi-squared Test With Survey Data, Journal of Royal Statistical Society, Ser. A., 143, 302-320.

[4] Rao, J. N. K. and Scott, A. J. (1984). On Chi-squared Tests for Multi-way Contingency Tables With Cell Proportions Estimated from Survey Data. Ann. Statist., 12, 46-60.

[5] Rao, J. N. K. and Scott, A. J. (1987). On Simple Adjustments Chi-squared Tests With Sample Survey Data. Ann. Statist., 15, 385-397.

[6] Roberts, G., Rao, J. N. K. and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. Biometrika, 74, 1-12.

[7] Schaible, Wesley L. (1979), "Estimation of Finite Population Totals from Incomplete Sample Data: Prediction Approach," in Symposium on Incomplete Data, Preliminary Proceedings, Washington, D.C., U.S. Department of Health, Education, and Welfare, 170-187.

[8] Simmons, W.R. and Baird, J. (1968). Use of Pseudo-replication in the NCHS Health Examination Survey, Proceedings of the Social Statistics Section of the American Statistical Association.

Table 3
Estimated Response Rate and Asymptotic Standard Deviation for Subpopulation of Public Schools

		Urbanicity	
Enrollment	Grade Level	1	2
1	1	* 0.967 (0.0017)	* 0.976 (0.0017)
	2	0.952 (0.0059)	0.964 (0.0048)
2	1	0.926 (0.0040)	0.980 (0.0014)
	2	* 0.807 (0.0083)	* 0.945 (0.0083)
3	1	0.938 (0.0028)	0.974 (0.0016)
	2	* 0.900 (0.0093)	* 0.958 (0.0093)
4	1	0.939 (0.0026)	0.974 (0.0020)
	2	* 0.947 (0.0081)	* 0.977 (0.0081)
5	1	0.923 (0.0024)	0.970 (0.0020)
	2	0.900 (0.0035)	0.960 (0.0067)

The asymptotic standard deviation of the cells marked by '*' have been estimated based on the aggregation of subpopulations.

Table 4
Estimated Response Rate and Asymptotic Standard Deviation for Subpopulations of List Frame Private Schools

			Association					
Enrollment	Grade Level	Urbanicity	1	2	3	4	5	6
1	1	1	0.943 (0.0074)	* 0.934 (0.011)	* 0.982 (0.079)	* 0.732 (0.084)	* 0.873 (0.018)	* 0.861 (0.009)
1	1	2	0.989 (0.012)	* 0.982 (0.011)	* 0.985 (0.079)	* 0.789 (0.084)	* 0.824 (0.018)	* 0.769 (0.009)
1	2	1	* 0.882 (0.009)	* 0.875 (0.011)	* 0.931 (0.079)	* 0.787 (0.084)	* 0.731 (0.025)	* 0.753 (0.010)
1	2	2	* 0.989 (0.009)	* 0.991 (0.011)	* 0.989 (0.079)	* 0.773 (0.084)	* 0.875 (0.025)	* 0.861 (0.010)
2	1	1	0.883 (0.043)	* 0.970 (0.007)	* 0.932 (0.081)	* 0.721 (0.013)	* 0.854 (0.020)	* 0.805 (0.014)
2	1	2	0.972 (0.006)	* 0.984 (0.007)	* 0.974 (0.081)	* 0.972 (0.013)	* 0.962 (0.020)	* 0.954 (0.14)
2	2	1	* 0.891 (0.010)	* 0.981 (0.007)	* 0.885 (0.081)	* 0.603 (0.013)	* 0.853 (0.023)	* 0.839 (0.019)
2	2	2	* 0.902 (0.010)	* 0.982 (0.007)	* 0.876 (0.081)	* 0.872 (0.013)	* 0.951 (0.023)	* 0.883 (0.019)

The asymptotic standard deviation of the cells marked by '*' have been estimated based on the aggregation of subpopulations.