# MODELS FOR SMOOTHING POST-STRATIFICATION WEIGHTS

**Laura C. Lazzeroni, U.C.L.A. and Roderick J.A. Little, University of Michigan**
**Department of Biostatistics, U.C.L.A. School of Public Health, Los Angeles CA 90024**

## 1. Introduction

We consider the estimation of a finite population mean of a continuous outcome variable **Y** based on a sample survey, in the presence of an ordinal post-stratifier **W** with known population distribution. Suppose **W** has H levels. For post-stratum $h$, let $N_h$ be the population size; $n_h$, the sample size; $Y_{hi}$, the value of **Y** for the $i^{th}$ individual in the population; and $y_{hi}$, the value of **Y** for the $i^{th}$ sample observation. Let:

$$N = \sum_h N_h \,;\; n = \sum_h n_h \,;$$

$$\bar{Y} = \sum_{h,i} Y_{hi} / N;\; \bar{Y}_h = \sum_i Y_{hi} / N_h;\;\; \bar{y}_h = \sum_i y_{hi} / n_h$$

denote respectively the total population size, the total sample size, the population mean, the within-stratum population mean, and the within-stratum sample mean.

It is assumed that given the value of **W**, the probability of inclusion in the sample does not depend upon the value of **Y**. Two standard approaches are to estimate $\bar{Y}$ by (a) the sample mean $\bar{y} = \sum_h p_h \bar{y}_h$, where $p_h = n_h / n$; and (b) the post-stratified mean $\bar{y}_{ps} = \sum_h P_h \bar{y}_h$, where $P_h = N_h / N$. The sample mean is an appropriate estimate of $\bar{Y}$ when **Y** and **W** are unrelated. Even if the variables are related, $\bar{y}$ is unbiased so long as the probability of inclusion in the sample does not depend upon **W**. However, $\bar{y}$ does not use information from the known distribution of **W** so that it may not be the best estimate. Due to sampling variability or systematic bias in the sampling procedures, the proportion $p_h$ falling within stratum $h$ of a given sample deviates from its respective population proportion $P_h$. Since $\bar{Y} = \sum_h P_h \bar{Y}_h$ and $\bar{y} = \sum_h p_h \bar{y}_h$, $\bar{y}$ is biased for $\bar{Y}$ conditional upon the sample configuration of **W**.

Given a set of sample values for **W**, $\bar{y}_{ps}$ is conditionally unbiased and may have much a smaller mean squared error than $\bar{y}$ (Holt and Smith 1979; Little 1993). Although $\bar{y}_{ps}$ incorporates information from **W**, $\bar{y}_h$ is used to estimate the stratum mean, whether a stratum contains a few or many observations. When a stratum contains few observations, the estimator of the stratum mean might be improved by borrowing strength from information from neighboring strata. Also, in our setting, the post-stratified mean does not reflect the ordinal nature of the post-stratifier. That is, it has the same form as for an unordered categorical post-stratifier.

Both $\bar{y}$ and $\bar{y}_{ps}$ can be written as

$$\sum_{h,i} w_h y_{hi} / \sum_{h,i} w_h$$

where $w_h$ is a weight attached to each observation in stratum $h$. Conditional on the observed values of **W**, the variance of these estimates is

$$\sum_{h,i} w_h^2 \left(1 - \frac{n_h}{N_h}\right) S_h^2 / \left(\sum_{h,i} w_h\right)^2$$

where $S_h^2$ is the within-stratum variance of $Y_{hi}$. For $\bar{y}_{ps}$, $w_h = P_h / p_h$, and for $\bar{y}$, $w_h = 1$ for all $h$. When $p_h$ is much smaller than $P_h$, $\bar{y}_{ps}$ relies greatly upon each observation in that stratum, inflating the variance of the final estimate. In fact, $p_h$ can equal zero, in which case adjustments are needed for $\bar{y}_{ps}$ to be defined.

Modifications of $\bar{y}_{ps}$ that reduce its variance have been proposed, which can be written as weighted averages of the observations where the original post-stratification weights have been smoothed to reduce variability. One method is to truncate the post-stratification weights larger than some maximum allowable value. Simultaneously, smaller weights are adjusted upwards. The truncation point may be fixed in an *ad-hoc* way, or based on the data (Potter 1990). A second approach is to pool or collapse strata. Strategies for choosing how and when to collapse strata have been suggested by Kalton and Maligalig

(1991), Little (1993), and Tremblay (1986). If sampling or nonresponse depends upon $\mathbf{W}$, modeling of these rates has been suggested (e.g. Kalton and Maligalig 1991). Observation weights can be based on the estimated rates, which will usually be smoother than the observed rates.

## 2. Proposed Methods

The ideal compromise between $\bar{y}$ and $\bar{y}_{ps}$ would use the ordinal structure of $\mathbf{W}$ to aid in the prediction of the stratum means. It would control variance by not weighting any individual observation too highly. When the sample means are well-observed and $\mathbf{Y}$ is strongly related to $\mathbf{W}$, the estimate should look like $\bar{y}_{ps}$. When $\mathbf{Y}$ and $\mathbf{W}$ are not strongly related, the estimate should look like $\bar{y}$. Also, since surveys contain large numbers of variables, the ideal method would have general applicability without requiring a lot of hands-on modeling for each outcome. However, if arbitrary choices are needed for the sake of generality, the resulting estimates should be insensitive to these choices.

We consider methods based on models for the outcome, which can be viewed within either a superpopulation or Bayesian framework. The general form of these models is

$$Y_{hi}|\mu_h \sim_{ind.} N(\mu_h, \sigma^2) \qquad (1)$$

and

$$\mu \sim N_H(\mathbf{X}\beta, \mathbf{D}), \qquad (2)$$

where $\mu = (\mu_1, \ldots, \mu_H)^T$, $\mathbf{X}$ is a known $H \times Q$ design matrix, $\beta$ is a $Q \times 1$ vector of unknown parameters and $\mathbf{D}$ is an $H \times H$ covariance matrix. For individuals not included in the sample, $Y_{hi}$ can be estimated by $\hat{\mu}_h$, its expected value given the data. The estimated finite population mean is

$$\sum_h [n_h \bar{y}_h + (N_h - n_h)\hat{\mu}_h]/N = \bar{y} + \sum (P_h - p_h)\hat{\mu}_h$$

The estimates $\hat{\mu}_h$ of the stratum means shrink the sample means $\bar{y}_h$ towards the $h$th element of $\mathbf{X}\hat{\beta}$, with a degree of shrinkage that tends to zero as the within-stratum sample size $n_h$ increases. Hence the estimator behaves like the poststratified mean in large samples, but smoothes the within-stratum means when the sample size is small.

Previous work has proposed inference under the exchangeable random effects (XRE) model obtained by setting $\mathbf{X} = \mathbf{I}$, the identity matrix, and $\mathbf{D} = \mathbf{I}\sigma^2_\mu$, where $\sigma^2_\mu$ is the between-stratum variance, in (2) (Holt and Smith 1979, Scott and Smith 1969, Little 1983, 1991, Ghosh and Meeden 1986). If $\sigma^2_\mu$ is set equal to zero then $\mu_h \equiv \mu$, $\hat{\mu}_h = \bar{y}$ and the final estimate is also $\bar{y}$. If $\sigma^2_\mu$ is set equal to infinity, a fixed effects ANOVA model is obtained, $\hat{\mu}_h = \bar{y}_h$ and the final estimate is $\bar{y}_{ps}$. If $\sigma^2_\mu$ is estimated from the data (an empirical Bayes approach), the resulting estimate of $\bar{Y}$ shrinks the means $\bar{y}_h$ towards $\bar{y}$. While these properties are appealing, simulations in Little (1991) indicate that model-based confidence intervals are sensitive to departures from the assumption of exchangeability in the post-stratum means, as noted by Morris (1983) in the general context of empirical Bayes estimation. The exchangeability assumption is questionable when W is ordinal, since a systematic relationship between Y and W might be expected. We develop extensions of the XRE model for the ordinal setting.

Two ways are suggested for incorporating the ordinal nature of $\mathbf{W}$ in the model (2). The more standard approach is to include functions of $\mathbf{W}$ in the construction of $\mathbf{X}$ so that the mean structure of $\mu$ depends upon $\mathbf{W}$. In particular, the regression (REG) model for the stratum means sets

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ \ldots & \ldots \\ 1 & H \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \sigma^2_\mu \times \mathbf{I}.$$

Another approach is to model $\mathbf{D}$ to incorporate greater positive correlation between $\mu_h$ and $\mu_{h'}$ when $h$ and $h'$ are close in value. For the AR1 model of the stratum means,

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}, \mathbf{D} = \sigma_\mu^2 \times \begin{bmatrix} 1 & \rho & \cdots & \rho^{H-1} \\ \rho & 1 & \cdots & \rho^{H-2} \\ \rho^2 & \rho & \cdots & \rho^{H-3} \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{H-1} & \rho^{H-2} & \cdots & 1 \end{bmatrix}$$

The REG and AR1 models each contain four parameters, compared with three for the XRE model. Of course these models could be elaborated, or a combined REG / AR1 model (with five parameters) could be fitted with some loss of parsimony; here we focus on the properties of the REG and AR1 models treated separately.

Assume that exactly $G$ strata contain data. The sufficient statistics are $\bar{\mathbf{y}}$ and $S$ where $\bar{\mathbf{y}}$ is the $G \times 1$ vector consisting of the ordered observed stratum means and $S = \sum_{h,i} (y_{hi} - \bar{y}_h)^2$. Let $\mathbf{Z}$ be the $G \times H$ submatrix of the $H \times H$ identity matrix with row $h$ deleted if $n_h = 0$. Let $\mathbf{R}$ be the $G \times G$ diagonal matrix with the $g^{th}$ diagonal element equal to $\sigma^2 / n_h$ if $h$ is the $g^{th}$ stratum for which $n_h > 0$. Define $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}} + \mathbf{R}$. Conditional on the observed $\mathbf{W}$,

$$\bar{\mathbf{y}} \sim N_G(\mathbf{Z}\mathbf{X}\beta, \mathbf{V})$$

and independently

$$S / \sigma^2 \sim \chi^2_{n-G}$$

This model for the observed sample means is a special case of a random effects model. Following Laird and Ware (1982), the resulting maximum likelihood (ML) estimates are

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{Z}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}\bar{\mathbf{y}}$$

and

$$\hat{\mu} = \mathbf{X}\hat{\beta} + \mathbf{D}\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}(\bar{\mathbf{y}} - \mathbf{Z}\mathbf{X}\hat{\beta})$$

when $\sigma^2$ and $\sigma^2_\mu$ are known. When the variance parameters are unknown, maximum likelihood estimates are used instead so that $\mathbf{D}$ becomes $\hat{\mathbf{D}}$ and $\mathbf{V}$ becomes $\hat{\mathbf{V}}$. Let $\bar{y}_r$ and $\bar{y}_a$ denote the estimate of the finite population mean based on the REG and AR1 model respectively. For these models, $\hat{\mu}_h$ is a weighted average of the sample observations. The estimates $\bar{y}_r$ and $\bar{y}_a$ can also be written as weighted averages of the observations for some set of smoothed weights, $w_h$.

**3. An Example**

The Los Angeles Epidemiologic Catchment Area survey of mental health was based on an equal probability sample of households in two areas, East Los Angeles and West Los Angeles (Eaton and Kessler 1985). Population distributions within these catchment areas were taken from the 1980 U. S. Census. Data for the age-related sampling proportions and population distribution are given in Little (1993).

Eight demographic groups defined by Ethnicity (H = Hispanic, N = Not Hispanic), Gender (F or M) and Catchment Area (E or W) were each analyzed separately. Sample sizes varied from 112 to 738. The outcome variable $\mathbf{Y}$ was a score measuring depression based on a set of questions from the survey. Although not continuous, $\mathbf{Y}$ takes 71 distinct values from 0 to 51. The sample was post-stratified on age, $\mathbf{W}$, with each year representing one stratum. Respondents varied from 18 to 96 while individuals in the population were recorded at age 102. No demographic group had individuals in all strata. Data from 3 groups, HMW, NME and NMW, are presented here. The left panels of Figure 1 show $\mathbf{Y}$ plotted against $\mathbf{W}$ showing an apparent downward trend with age but with lots of variability especially at younger ages. Plots of $P_h$ and $p_h$ plotted against age (not included here) suggest that some age ranges may be systematically undersampled.

The AR1 and REG models were both applied to the data, using a modified version of the scoring algorithm described by Jennrich and Schluchter (1986) to maximize the likelihood. Penalty functions were used to insure that $\hat{\sigma}^2 > 0$, $\hat{\sigma}^2_\mu > 0$ and $0 < \hat{\rho} < 1$.

Under the constraints, $\hat{\sigma}^2_\mu$ was essentially zero in three cases for the AR1 model and six cases for the REG model. For the AR1 model, this means that $\hat{\rho}$ is meaningless and it is not reported. The estimated slope is always negative under the REG model. Because the AR1 model describes the stratum means rather than the observations, $\hat{\rho}$ can be quite large, greater than 0.9 in two cases. A simple intercept model for the outcomes was also fit to the data. This is the simplest super-population model that leads to the estimate $\bar{y}$. For the AR1 and the REG models, Table 1 contains the difference of the log-likelihood from this simplest model as well as maximum likelihood estimates of the parameters.

The right panels of Figure 1 show $\hat{\mu}_h$, the estimated ~~or predicted means~~, plotted against age, from post-stratification, and from the AR1 and REG models. For the sample mean, the predicted means are identically equal to $\bar{y}$. Straight lines correspond to $\hat{\sigma}_\mu = 0$ since the predicted mean is then just the estimated fixed effect. The NME group has the least smooth predicted values and the largest ratio of $\hat{\sigma}^2_\mu$ to $\hat{\sigma}^2$. This group also has the smallest meaningful estimates of $\rho$ for the AR1 model. Plots of the weights (not shown here) indicate that for both models, the variability of the weights is small compared to truncation-based methods that have been proposed. The AR1 model displays weights that vary with the observed sampling rates whereas the weights of the REG model are usually linear.

Table 2 shows $\bar{y}_a$, $\bar{y}_r$, $\bar{y}$ and $\bar{y}_{ps}$ for three of the groups. In general, all four estimates are similar for this data set. Also listed is $Q = \sum_{h,i} w_h^2 / (\sum_{h,i} w_h)^2$, which is based on expression 1 and is meant to reflect the finite population sampling variance of the estimates. This shows that both model-based estimates are considerably less variable than the post-stratified mean, which is not surprising given the small variability of the weights.

## 5. Conclusion

The REG and AR1 models for the post-stratum means appear reasonable generalizations of the XRE model for smoothing post-stratum means based on an ordinal post-stratifier. More detailed assessments of these methods, on populations simulated under a variety of conditions, are needed to answer the question of which method is preferable for routine survey use. Smoothing based on a model like REG and AR1 seems more principled than arbitrary procedures that truncate the weights. The models in effect provide different weights for each survey outcome, and hence involve more computation than methods that provide a single smoothed weight for all outcomes. But computation is less of an issue in the era of high-speed computers, and any procedure directed at reducing variance should tailor the weights, depending on the degree of association of the post-stratifier with the outcome.


## Acknowledgments

This research was supported by NIMH grant MH 37188.


## References


Eaton, W.W. and Kessler, L.G. (1985, eds.) *Epidemiologic Field Methods in Psychiatry: the NIMH Epidemiologic Catchment Area Program*, New York: Academic Press.

Ghosh, M. and Meeden, G. (1986). "Empirical Bayes Estimation of Means from Stratified Samples. *Journal of the American Statistical Association*", 81, 1058-1062.

Holt, D. and Smith, T.M.F. (1979), "Post Stratification", *Journal of the Royal Statistical Society, Series A*, 142, 33-46.

Kalton, G. and Maligalig, D.S. (1991), "A Comparison of Methods of Weighting Adjustment for Nonresponse," *Proceedings of the 1991 Annual Research Conference*, pp. 409-428, U.S. Bureau of the Census, Arlington VA.

Little, R.J.A. (1983). "Estimating a finite population mean from unequal probability samples", *Journal of the American Statistical Association*, 78, 596-604.

Little, R.J.A. (1991). "Inference with survey weights", *Journal of Official Statistics*, 7, 405-424.

Little, R.J.A. (1993). "Post-stratification: a modeler's perspective", *Journal of the American Statistical Association*, 88, 1001-1012.

Morris, C.N. (1983). "Parametric Empirical Bayes Inference: Theory and Applications", *Journal of the American Statistical Association*, 78, 47-55.

Potter, F. (1990), "A Study of Procedures to Identify and Trim Extreme Sampling Weights," *Proceedings of the Survey Research Methods Section, American Statistical Association, 1990*, 225-230.

Scott, A. and Smith, T.M.F. (1969), "Estimation in Multistage Sampling," *Journal of the American Statistical Association*, 64, 830-840.

Tremblay, V. (1986), "Practical Criteria for Definition of Weighting Classes," *Survey Methodology*, 12, 85-97.

**TABLE 1. Maximum Likelihood Estimation for AR1 and REG Models**

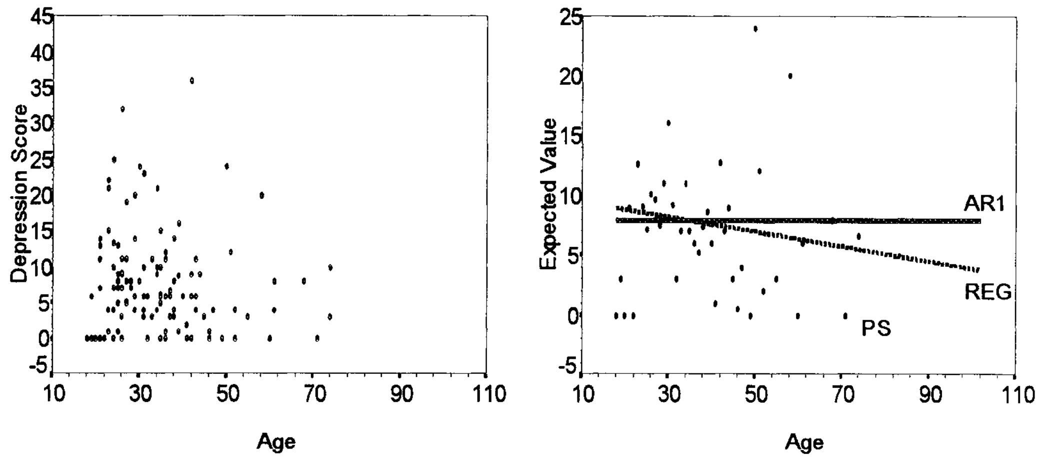
| Group | Model | Difference in Log-Likelihood | MLE of Intercept | MLE of $\sigma^2$ | MLE of $\sigma_\mu^2$ | MLE of $\rho$ | MLE of Slope |
|---|---|---|---|---|---|---|---|
| HMW | AR1 | 0.00 | 7.90 | 51.15 | 0.00 | NA | |
| *N=11448, n=112* | REG | 0.63 | 10.05 | 49.59 | 0.00 | | -0.06 |
| NME | AR1 | 1.47 | 6.66 | 37.18 | 6.46 | 0.30 | |
| *N=11287, n=124* | REG | 1.98 | 8.66 | 37.37 | 5.64 | | -0.04 |
| NMW | AR1 | 2.01 | 5.84 | 47.66 | 0.98 | 0.92 | |
| *N=53593, n=711* | REG | 2.79 | 7.81 | 48.01 | 0.07 | | -0.04 |

**TABLE 2. Estimates of Overall Means from Four Methods**

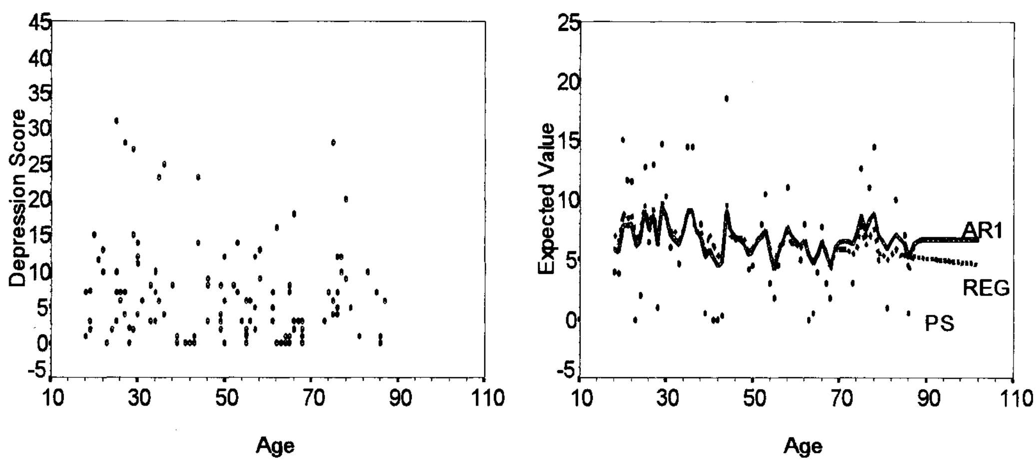
| | HMW | NME | NMW |
|---|---|---|---|
| $\bar{y}_a$ | 7.90 (0.0089) | 6.67 (0.0083) | 6.14 (0.0014) |
| $\bar{y}_r$ | 7.85 (0.0090) | 6.74 (0.0083) | 6.17 (0.0014) |
| $\bar{y}$ | 7.90 (0.0089) | 6.61 (0.0081) | 6.25 (0.0014) |
| $\bar{y}_{ps}$ | 6.83 (0.0141) | 6.62 (0.0124) | 6.04 (0.0017) |

**Figure 1.**

**Hispanic Males in West Los Angeles**




**Non-Hispanic Males in East Los Angeles**




**Non-Hispanic Males in West Los Angeles**