WEIGHTING ADJUSTMENTS FOR UNIT NONRESPONSE WITH MULTIPLE OUTCOME VARIABLES

Sonya L.Vartivarian, Department of Statistics, University of Michigan. E-mail: slvartiv@umich.edu
Roderick J. Little, Department of Biostatistics, University of Michigan. E-mail: rlittle@umich.edu

KEY WORDS: Sampling weights; Survey Inference; Unit Nonresponse Adjustment

1. Introduction

Weighting is a common form of unit nonresponse adjustment in sample surveys where entire questionnaires are missing due to noncontact or refusal to participate. Weights are inversely proportional to the probability of selection and response. A common approach computes the response weight as the inverse of the response rate within adjustment cells based on covariate information. When the number of cells thus created is too large, a coarsening method such as response propensity stratification can be applied to reduce the number of adjustment cells. Simulations in Vartivarian and Little (2002) indicate improved efficiency and robustness of weighting adjustments based on the joint classification of the sample by two key potential stratifiers: the response propensity and the predictive mean, both defined in Section 2. Predictive mean stratification has the disadvantage that it leads to a different set of weights for each key outcome. However, potential gains in efficiency and robustness make it desirable to use a joint classification. Here, we consider the efficiency and robustness of weights that jointly classify on the response propensity and predictive mean, but that base the predictive mean dimension on a single canonical outcome variable.

2. Coarsening the Set of Covariates

Let $D = (X,Z)$ be all fully-observed survey variables X and design variables Z, Y be the set of outcome variables and R be a response indicator. In principle, adjustment cells might be based on a joint classification of the variables D. We assume that given the classification D, nonresponse is missing at random (MAR; Rubin, 1976; Little and Rubin, 2002), that is

$$R \amalg Y \mid D \qquad (1)$$

where $\amalg$ denotes independence. Since classification on D may be unrealistic, we seek a coarsening A of D, such that

$$R \amalg Y \mid A\,. \qquad (2)$$

Little (1986) defines the *response propensity* as

$$p(D) = pr(R = 1 \mid D)$$

and supposes that $p(D) > 0$ for all observed values of D. Then Little (1986) uses the theory of Rosenbaum and Rubin (1983) to show that (1), i.e. ignorable nonresponse, implies that

$$Y \amalg R \mid p(D). \qquad (3)$$

The response propensity can be modeled via a logistic regression fit to the sampled cases, for example, and a grouped version of the response propensity can be the basis for adjustment cells, where grouping can be based on the quintiles of the distribution of the estimated response propensity. If A is a coarsening of D based on response propensity stratification, then (2) holds approximately so adjustment based on A controls nonresponse bias.

Modeling the distribution of the outcome Y given D is the second strategy for reducing the number of adjustment cells suggested in Little (1986). Since (1) implies that the distribution of the population values Y for respondents and nonrespondents are homogeneous given D, Little (1986) notes that pooling over values of D such that the distribution is constant results in subpopulations where the outcome Y and response R are still independent. Thus, Little (1986) specifies a model for the distribution of Y given D such that the distribution of the outcome Y differs only in the location parameter for different values of D. Then forming adjustment cells such that the location parameter is constant within the adjustment cells satisfies (2), implying

$$Y \amalg R \mid \hat{y}(D)\,. \qquad (4)$$

Since the location parameter is usually unknown, an estimate is obtained as the predicted mean $\hat{y}(D)$ from the regression of the outcome Y given D fitted to responding cases, for example. A grouped version $\hat{y}_G(D)$ of the predictive mean can be the basis for forming adjustment cells. One possible choice is to base the groups on the quintiles of the distribution of $\hat{y}(D)$. If A is a coarsening of D based on the predictive mean stratification, then (2) holds approximately so adjustment based on A controls nonresponse bias.

3. Joint Classification of the Response Propensity and the Predictive Mean

Simulations in Vartivarian and Little (2002) examine the response propensity and the predictive mean as potential stratifiers. Response propensity stratification is attractive since it requires less modeling and has zero large sample bias for domain and cross-class means, where cross-classes are classes that cut across adjustment cells. However, it does not control variance and can be very inefficient. Predictive mean stratification has the advantage of controlling both bias and variance of the overall mean. Yet, it produces a different stratification and thus a different set of weights for each outcome, and it does not in general control the bias of cross-class means. Following Little and Rubin (2000), Vartivarian and Little (2002) propose to cross classify on the response propensity scores $\hat{p}(D)$ and the best linear predictor $\hat{y}(D)$ to form adjustment cells. The motivation is to capture the bias-reduction property of response propensity stratification and gains in efficiency for predictive mean stratification. The joint classification also has potential gains in robustness because of the "double robustness" property, where unbiasedness is gained if one of the models is correct, and efficiency is gained if both are correct.

Simulations in Vartivarian and Little (2002) suggest that an improvement in efficiency is gained in situations where the response propensity stratification alone is inefficient, with some loss in efficiency when the response propensity is efficient. Further, the simulations demonstrate robustness of the joint classification to misspecification of the model for the response propensity or the predictive mean. However, the simulations in Vartivarian and Little (2002) focus on the simple situation of a single outcome *Y*, where predictive mean stratification yields a one-dimensional classification variable. In surveys with multiple key outcomes, the method of crossclassification proposed in Vartivarian and Little (2002) would lead to a different set of weights for each outcome, as illustrated in Table 1, which is practically cumbersome and leads to complications for multivariate analysis.

Table 1. Number of Sets of Weights Needed for Adjustment Cell Stratification

Method of Adjustment Cell Stratification	Number of Weight Sets	Number of Cross-classifiers
Response Propensity Score	1	1
Predictive Mean based on 1st Canonical Covariate	1	1
Response Propensity Score and Predictive Mean based on 1st Canonical Covariate	1	2
Predictive Mean of Each Outcome Y	k	1
Response Propensity Score And Predictive Mean of Each Outcome Y	k	2

We desire a compromise predictive mean that limits the number of sets of weights, but potentially offers gains in efficiency and double robustness when crossclassified with the response propensity.

4. Joint Classification of the Response Propensity and the Canonical Covariate

One approach to limiting the sets of weights to a smaller number, whilst retaining the efficiency of estimation for the means of individual outcomes is to base weights on a predictive mean stratification using the first canonical covariate determined by the set of outcomes and covariates. Let $Y=(Y_1,...,Y_k)$ denote the set of k continuous key outcomes and $D=(X_1,...,X_p)$ denote the set of p covariates. We assume that $k \le p$ in our application of the theoretical canonical correlation results, though the results also hold when the smaller set is the covariate set. Let $s_{jj}^{(yy)}$ be the sample variance of outcome Y_j, $j=1,\dots,k$, and $s_{ij}^{(yy)}$ be the sample covariance of Y_i and Y_j, for $i \ne j$ and $i,j=1,\dots,k$, resulting in the sample variance-covariance matrix $\hat{\Sigma}_{YY}$ of *Y*:

$$\hat{\Sigma}_{YY}=\begin{bmatrix} s_{11}^{(yy)} & s_{12}^{(yy)} & & \cdots & s_{1k}^{(yy)} \\ \vdots & s_{22}^{(yy)} & & & \vdots \\ & & \ddots & & \\ \vdots & & & \ddots & \vdots \\ s_{k1}^{(yy)} & \cdots & & \cdots & s_{kk}^{(yy)} \end{bmatrix}.$$

Similarly, let $s_{jj}^{(xx)}$ be the sample variance of covariate X_j, $s_{ij}^{(xx)}$ be the sample covariance of X_i and X_j, for $i \ne j$ and $i,j=1,\dots,p$,where all variances and covariances are based on complete cases. The resulting complete case sample variance-covariance matrix of *D* is denoted by $\hat{\Sigma}_{DD}$,

$$\hat{\Sigma}_{DD}=\begin{bmatrix} s_{11}^{(xx)} & s_{12}^{(xx)} & & \cdots & s_{1p}^{(xx)} \\ \vdots & s_{22}^{(xx)} & & & \vdots \\ & & \ddots & & \\ \vdots & & & \ddots & \vdots \\ s_{p1}^{(xx)} & \cdots & & \cdots & s_{pp}^{(xx)} \end{bmatrix}.$$

Finally, let the vector $\mathbf{V}$,

$$\mathbf{V}_{((k+p)\times 1)} = \left[\begin{array}{c}\mathbf{Y}\\ \hdashline \mathbf{D}\end{array}\right] = \left[\begin{array}{c} Y_1 \\ Y_2 \\ \vdots \\ Y_k \\ \hdashline X_1 \\ X_2 \\ \vdots \\ X_p \end{array}\right]$$

have variance-covariance matrix

$$\hat{\Sigma} = \left[\begin{array}{c:c} \hat{\Sigma}_{YY} & \hat{\Sigma}_{YD} \\ \hdashline \hat{\Sigma}_{DY} & \hat{\Sigma}_{DD} \end{array}\right],$$

where

$$\hat{\Sigma}_{YD} = \begin{bmatrix} s_{11}^{(yx)} & s_{12}^{(yx)} & & \cdots & s_{1p}^{(yx)} \\ s_{21}^{(yx)} & s_{22}^{(yx)} & & & s_{2p}^{(yx)} \\ & & \ddots & & \\ \vdots & & & \ddots & \vdots \\ s_{k1}^{(yx)} & s_{k2}^{(yx)} & & \cdots & s_{kp}^{(yx)} \end{bmatrix}.$$

Then, we reduce the dimension of the outcome set Y by choosing a linear combination of the outcomes, say $U = a'Y$, where $a' = (a_1,\ldots,a_k)$ is a coefficient vector, to replace the entire set Y. We would like this set of outcomes to be maximally correlated with the set of covariates D. Specifically, we would like U to be maximally correlated with a linear combination of D, say $T = b'D$, where $b' = (b_1,\ldots,b_p)$ is a coefficient vector. The relevant result taken from Johnson and Wichern (1992, p.472) is as follows and can be found in standard multivariate texts such as Kshirsagar (1972):

Result 10.2. Let $\hat{\rho}_1^{*2} \geq \hat{\rho}_2^{*2} \geq \cdots \geq \hat{\rho}_k^{*2}$ be the k ordered eigenvalues of $\hat{\Sigma}_{YY}^{-1/2}\hat{\Sigma}_{YD}\hat{\Sigma}_{DD}^{-1}\hat{\Sigma}_{DY}\hat{\Sigma}_{YY}^{-1/2}$ with corresponding eigenvectors $\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2,\ldots,\hat{\mathbf{e}}_k$, where the $\hat{\Sigma}_{YY}$, $\hat{\Sigma}_{DD}$ and $\hat{\Sigma}_{YD}$ are as defined previously, and $k \leq p$. Let $\hat{\mathbf{f}}_1,\hat{\mathbf{f}}_2,\ldots,\hat{\mathbf{f}}_p$ be the eigenvectors of $\hat{\Sigma}_{DD}^{-1/2}\hat{\Sigma}_{DY}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YD}\hat{\Sigma}_{DD}^{-1/2}$, where the first k $\hat{\mathbf{f}}$'s may be obtained from $\hat{\mathbf{f}}_j = (1/\hat{\rho}_j^*)\Sigma_{DD}^{-1/2}\Sigma_{DY}\Sigma_{YY}^{-1/2}\hat{\mathbf{e}}_j$, for $j = 1,\ldots,k$. The jth sample canonical variate pair is:

$$\hat{U}_j = \hat{\mathbf{e}}_j'\hat{\Sigma}_{YY}^{-1/2}Y;\ \hat{T}_j = \hat{\mathbf{f}}_j\hat{\Sigma}_{DD}^{-1/2}D \qquad (5)$$

The first sample canonical variate pair $(\hat{U}_1,\hat{T}_1)$ have the maximum sample canonical correlation

$$r(\hat{U}_1,\hat{T}_1) = \hat{\rho}_1^*. \qquad (6)$$

The proof can be found in the same referenced text.

We replace our p covariates by the linear combination that has the maximum correlation $\hat{\rho}_1^{*2} = r^2(\hat{U}_1,\hat{T}_1)$ with the first canonical variate $\hat{U}_1$ of the outcomes. Note that using a standardized set of outcomes and covariates may be desirable because of increased interpretability of the coefficients and of descriptive summary measures. For example, if the measurement scale is vastly different for each of the variables, using the standardized variables then allows the coefficients to more accurately reflect the contribution of each variable to its canonical variate. Also, when the variables are standardized, one can then examine the proportion of total sample standardized variance within the outcome set explained by its first canonical variate:

$$R^2_{Y|\hat{U}_1} = \frac{1}{k}\sum_{j=1}^{k} r^2_{\hat{U}_1,Y_j}, \qquad (7)$$

where the $r^2_{\hat{U}_1,Y_j}$ is the coefficient of determination between $\hat{U}_1$ and Y_j, for $j = 1,\ldots,k$. This quantity may be used as one indication of how well the efficiency of the estimate of the mean of Y_j is preserved when using $\hat{U}_1$ as a compromise predictive mean crossclassifier. Note that the full set of outcomes Y may be used in subsequent analyses as the data reduction of the outcomes we consider is only with respect to the formation of adjustment cells and obtaining the corresponding weights.

5. Simulation Study

We consider a case with three outcome variables and four covariates. The population is structured such that one outcome is unbiased, while the remaining two outcomes are biased, but have different relationships with the response mechanism. We assume a MAR response mechanism.

5.1 Simulation Superpopulation Structure

A superpopulation model for six covariates, $D = (X_1, X_2, X_3, X_4, X_5, X_6)$, and three outcomes, $Y = (Y_1, Y_2, Y_3)$, is assumed to be a multivariate normal distribution with zero mean vector and the following covariance matrix Σ:

$$\Sigma = \left(\begin{array}{c:c} \Sigma_{YY} & \Sigma_{YD} \\ \hdashline \Sigma_{DY} & \Sigma_{DD} \end{array} \right)$$

$$= \left(\begin{array}{ccc:cccccc} 1 & 0 & 0 & 0.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.64 & 0.64 & 0 & 0 \\ \hdashline 0.9 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.64 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.64 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0. & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right).$$

It should be noted that covariate X_1 is highly correlated with outcome Y_1, covariate X_2 is highly correlated with outcome Y_2, and covariates X_3 and X_4 have a high multiple correlation with outcome Y_3. Covariates X_5 and X_6 are uncorrelated with all of the outcomes.

The following probit model was assumed for response:

$$P(R=1 \mid D=\{x_1\ x_2\ x_3\ x_4\ x_5\ x_6\}) \\ = \Phi\{0.2+\beta_1 *(x_1+x_3)+\beta_2 *(x_5+x_6)+\varepsilon\},$$

ε is a standard normal error term, and the coefficients $\beta=\{\beta_1\ \beta_2\}$ are varied according to Table 2 forming three different response mechanisms.

Table 2. Simulation Response Probability Coefficients.

	β_1	β_2
1.	0.5	0.5
2.	0.2	0.8
3.	0.8	0.2

Therefore, we have a total of three superpopulations, each determined by one value of $\beta=\{\beta_1\ \beta_2\}$. The response rate is approximately fifty-five percent in each scenario. One hundred replicate simple random samples, each of size n = 2200, were drawn for each scenario.

5.2 Modeling the Predictive Mean

Predictive mean models for Y_i, i = 1, 2, 3, include all covariates and were fit to the respondent sample. The notation y1F denotes a multiple regression of Y_1 on $D=(X_1,X_2,X_3,X_4,X_5,X_6)$ fit to the respondent data, whereas y2F and y3F are based on outcomes Y_2 and Y_3, respectively. Five adjustment cells were formed according to the quintiles of the distribution of the predicted values for each model.

5.3 Modeling the Compromise Predictive Mean

A canonical correlate analysis was performed, fit to the respondent data. All covariates and outcomes were included. The first canonical covariate of the outcome set was then used as the compromise outcome variable. A regression of the compromise outcome variable on all covariates was fit to the respondent data. The quintiles of the distribution of predicted values from this regression formed the five adjustment classes. The first canonical correlate is referred to as ycc1F since all covariates and outcomes were included in the canonical sets.

5.4 Modeling the Response Propensity

Two probit response propensity models were fit to the sample data: a full model denoted by pF including all covariates that the response propensity depends on, and a mismodeled response propensity denoted by pM where covariates X_5 and X_6 are omitted. Specifically, pF denotes a probit regression of the response indicator R on $(X_1,X_2,X_3,X_4,X_5,X_6)$, whereas pM denotes a probit regression of R on (X_1,X_2,X_3,X_4). The models are summarized in Table 3. Joint classifications are denoted by the predictive mean followed by the response propensity model. For example, a joint classification of y3F and pM is represented as "y3FpM", where a joint classification of ycc1F and pM is denoted by "ycc1FpM" .

Table 3. Models for Classifiers.

Model	Mean Classification	Covariates Included in Model
1.	pF	$(X_1,X_2,X_3,X_4,X_5,X_6)$
2.	pM	(X_1,X_2,X_3,X_4)
3.	y1F	$(X_1,X_2,X_3,X_4,X_5,X_6)$
4.	y2F	$(X_1,X_2,X_3,X_4,X_5,X_6)$
5.	y3F	$(X_1,X_2,X_3,X_4,X_5,X_6)$
6.	ycc1F	(Y_1,Y_2,Y_3) $(X_1,X_2,X_3,X_4,X_5,X_6)$

5.5 Results

The root mean square error (RMSE) and absolute bias (AB) relative to the superpopulation mean were examined for each outcome variable. In this simulation study, averaging over all three populations, we can see that a joint classification of the compromise predictive mean and the response propensity does not entail a great loss with respect to the RMSE. In fact, the compromise stratification performs similarly to the correct model for the response propensity, but the potential benefit of protection from model misspecification and possible gains from double robustness is an advantage in using the crossclassification. See Figures 1 and 2, where the mean before deletion of cases due to nonresponse and the respondent mean are also included, denoted by meanbd and meanr, respectively.

Outcome Y_1 represents a case where the predictive mean adds efficiency to the response propensity as seen in y1FpF. This efficiency is compromised by using the canonical correlate instead of the predictive mean in the crossclassification, but this loss may be offset by potential protection for model misspecification. Also, the canonical crossclassification does not show a great loss when compared to the response propensity classification alone: average 10000(RMSE) = 348.59 and 343.59 for ycc1FpF and pF, respectively.

For outcome Y_2, since the response probability does not involve the covariate associated with outcome Y_2, there is no bias. Using the canonical crossclassification offers a slight loss in precision over the response propensity in this case. The predictive mean classification is efficient as expected with a high correlation between Y_2 and X_2.

Outcome Y_3 has considerable bias that is corrected by the canonical crossclassification as well as the classification by the response propensity or the predictive mean. However, the response propensity is inefficient in this case. The canonical crossclassification performs well here compared to the response propensity classification, with gains in efficiency. Much of the efficiency gained by the using the predictive mean as a crossclassifier is retained when the predictive mean dimension is replaced by the compromise canonical correlate.

6. Summary

This research is promising for survey practitioners in that a relatively fast and easy compromise predictive mean that leads to only one set of weights (with two classifiers) may be applied when multiple key outcomes are present. The efficiency is comparable to that based on the predictive mean stratification alone, but potential protection from model misspecification is allowed with the second compromise canonical classification used with the response propensity.

6. Further Work

Future work should examine the efficiency of the canonical variate classification when the set of outcomes and covariates are not multivariate normal. It may also be useful to consider the canonical covariate based on the residual space, after partialing out the response propensity since the predictive mean dimension is viewed as an additional classification that may improve the efficiency of the response propensity.

ACKNOWLEDGEMENTS

This research was supported by grant DM59803720 from the National Science Foundation.

REFERENCES

1. Botman SL, Moore TF, Moriarty CL, Parsons VL. Design and Estimation of the National Health Interview Survey, 1995-2004. *National Center for Health Statistics, Vital Health Statistics* 2000; 2, 130.
2. Cochran WG. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics*, 1968; 24: 295-313.
3. Goskel H, Judkins D, and Mosher, D. Nonresponse Adjustments for a Telephone Follow-up to a National In-Person Survey. *Journal of Official Statistics*, 1992; 4: 417-431.
4. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc.: New Jersey, 1992.
5. Kshirsagar, AM. *Multivariate Analysis*, New York: Marcel Dekker, Inc., 1972.
6. Little RJ. Survey nonresponse adjustments. *International Statistical Review*, 1986; 54: 139-157.
7. Little RJ and Rubin DB. Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes: Concepts and Analytical Approaches. *Ann. Rev. Pub. Health*, 2000; 21:121-45.
8. Little RJ and Rubin DB. *Statistical Analysis with Missing Data,* 2nd Edition, John Wiley and Sons, Inc: New York, 2002.
9. Little RJ and Vartivarian S. On weighting the rates in non-response weights. *Special Issue, Statistics in Medicine* on papers from the 2001 CDC/ATSDR Symposium on Statistical Methods; *Statist. Med.,* 2003; 22:1589-1599.
10. Robins JM, Rotnitzky A and van der Laan M. Comment on “On Profile Likelihood” by Murphy S and van der Vaart A. W. *Journal of the American Statistical Association*; 2000; 95: 477-482.
11. Rosenbaum PR and Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*; 1983; 70: 41-55.
12. Rubin DB. Inference and missing data. *Biometrika*, 1976; 63, 3: 581-92.
13. Vartivarian S and Little RJ. On the formation of weighting adjustment cells for unit nonresponse. *Proceedings of the Survey Research Methods Section, American Statistical Association, 2002*.
14. Zeng D. *Adjusting for Dependent Censoring using High-Dimensional Auxiliary Covariates*. University of Michigan Press: Ann Arbor, MI, 2000.

Figure 1: Overall Performance of Various Weight Classifications

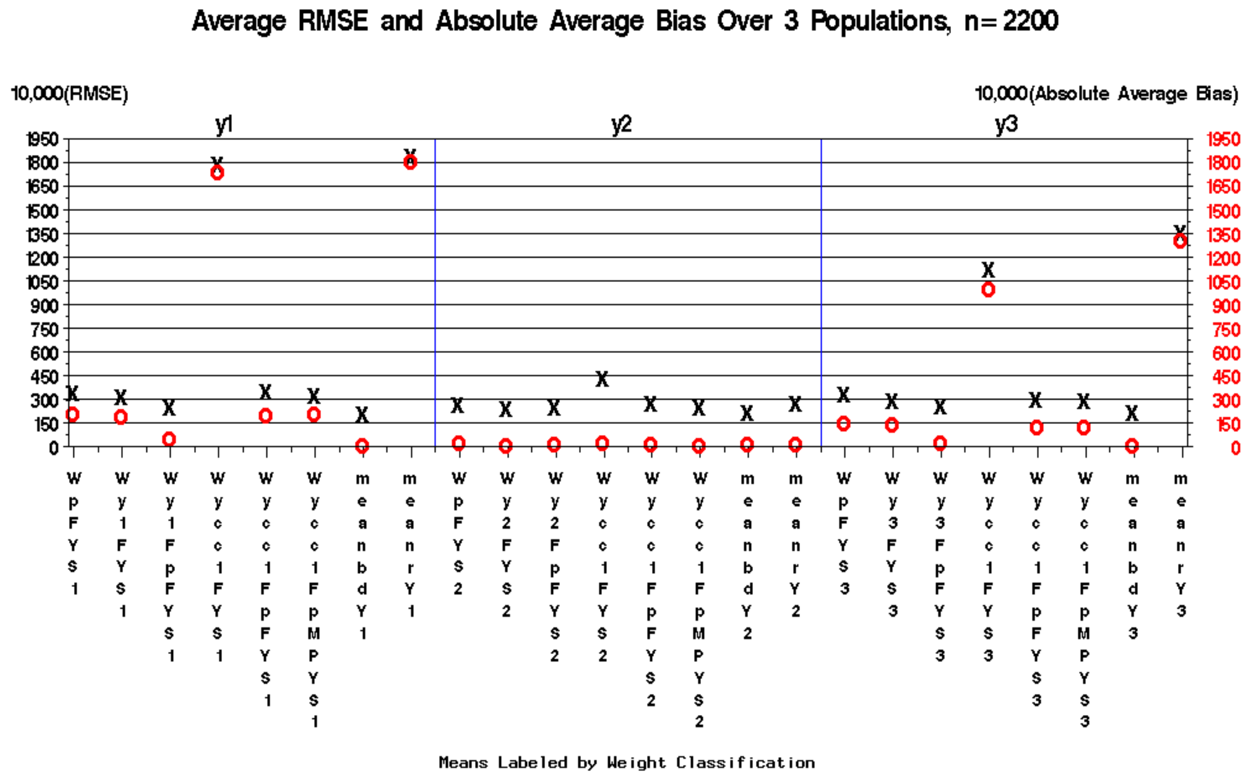

Figure 2: Performance of Various Weight Classifications with RMSE < 450