Modeling Non-response Adjustment Factors

Hee-Choon Shin, NORC, 55 East Monroe St., Chicago, IL 60603

Key Words: weighting, non-response adjustment, log-linear model.

Introduction

The most common method of adjustment for unit nonresponse is weighting, where respondents and nonrespondents are classified into adjustment cells based on some covariates known for all units in the sample, and a nonresponse adjustment factor is computed for all responding cases in each cell proportional to the inverse of the response rate in the cell. (Little, 1982; Little, 1986; Little, 1988)
The estimated response rates depend on the overall cell structure.

Traditional non-response adjustment

An arbitrary and *a*theoretical approach to non-response adjustment is to set a minimum cell size and minimum response rate (or adjustment factor) within each cell. Typically these two objectives are achieved by collapsing adjacent cells. The problem of this traditional approach is arbitrariness of all the decisions involved in determining the cell structures: decision on minimum cell size, decision on minimum response rate, and decision on the way how to collapse cells.

Propensity weighting

The theory of propensity scores, pioneered by Rosenbaum and Rubin(Rosenbaum & Rubin, 1983) and applied to survey nonresponse by Little (Little, 1986), is a theoretical approach to nonresponse adjustment. Define the response propensity for the i[th] unit

$$p(x_i) = \Pr(m_i = 0 \setminus x_i),$$

where m_i is the indicator for missingness (1=missing; 0 = response), and x_i is the set of variables for both respondents and nonrespondents. The theory is based on the following relationship:

$$\Pr[M \setminus p(X), Y] = \Pr[M \setminus p(X)].$$

Respondents are a random subsample within strata defined by $p(X)$, the propensity score. As succinctly described in Little and Rubin (2002), practical application involves the following steps:

1) Estimate $p(X)$ as $\hat{p}(X)$ by using logistic or probit regression,
2) Categorize the estimated $\hat{p}(X)$ into five or six values, and
3) Let the adjustment cells be equal to the categorized variable.

Even though the propensity weighting is a theoretical approach, we can not exclude the arbitrariness of the step 2, and the arbitrariness, to a less degree, in choosing a particular model in step 1. This paper does not challenge the well-established propensity weighting method itself, but attempt to exclude arbitrary decision-making activities in applying the method to actual data.

The current work is different from earlier other studies utilizing response probabilities in adjusting non-response bias (Alho, 1990; Sarndal, 1981). The current paper specifically deals with the way how to construct the adjustment cells.

Estimating non-response adjustment factors using multiplicative models

Consider classification variables A with categories $i = 1,\cdots,I$; another classification variable B with categories $j = 1,\cdots,J$; and a response indicator variable C with categories $k = 1,\cdots,K$. Here K is usually 2 but could be greater than 2. We will show our approach with 3-way cross-tabulation but extending our approach to higher-dimensional tables are straight-forward. Now consider the following model, a "saturated" multiplicative model with zero degrees of freedom:

$$\pi_{ijk} = \eta \tau_i^A \tau_j^B \tau_k^C \tau_{ij}^{AB} \tau_{ik}^{AC} \tau_{jk}^{BC} \tau_{ijk}^{ABC},$$

where π_{ijk} is the unconditional cell probability. Assuming $\pi_{ijk} > 0$ for all *i*, *j*, and *k*, and letting $v_{ijk} = \log \pi_{ijk}$, the τ parameters can be written as

$$\tau_i^A = \exp(\bar{v}_{i..} - \bar{v}_{...})$$
$$\tau_{ij}^{AB} = \exp(\bar{v}_{ij.} - \bar{v}_{i..} - \bar{v}_{.j.} + \bar{v}_{...})$$
$$\tau_{ijk}^{ABC} = \exp(v_{ijk} - \bar{v}_{ij.} - \bar{v}_{i.k} - \bar{v}_{.jk} + \bar{v}_{i..} + \bar{v}_{.j.} + \bar{v}_{..k} - \bar{v}_{...})$$

with similar formulae for τ_j^B, τ_k^C, τ_{jk}^{BC}, and τ_{ik}^{AC}. The dot subscript denotes summation with respect to the subscript it replaces and the bar denotes average. The parameter η is a scale factor ensuring $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\pi_{ijk} = 1$.

For details on estimation of the parameters can be found in Goodman's work (Goodman, 1970; Goodman, 1972).

Let k=1 be the indicator for response and k=2 otherwise. Non-response adjustment factor in the i[th] A and j[th] B, ρ_{ij}^{-1} is

$$\rho_{ij}^{-1} = \frac{\pi_{ij1} + \pi_{ij2}}{\pi_{ij1}}.$$

The traditional approach utilizes the saturated model to estimate the non-response adjustment factor, $\hat{\rho}_{ij}^{-1}$. Our approach is to find a parsimonious best model to calculate ρ_{ij}^{-1}.

Application

Consider a hypothetical data shown in Table 1. The traditional approach would examine the cell sizes and the cell specific response rates, and would collapse the categories on ad-hoc basis so that the predetermined minimum cell size and response rate may be maintained. When the propensity score method is utilized, we would fit a logistic or probit model to the data, and construct the adjustment cells by categorize the predicted propensity scores on ad-hoc basis.

Table 2 shows fitted cell probabilities for each model and goodness of fit statistics by applying the proposed multiplicative model. There are only two models: Model 8 and Model 9. As shown in Table 2, the model with all the effects including the three-way interaction fits the data perfectly (P-value=1). The traditional approach utilizes the saturated model (Model 9) to estimate the non-response adjustment factor, $\hat{\rho}_{ij}^{-1}$. The model (Model 8) with all two-way interaction or no-three way interaction model is the most parsimonious model (P-value = .2452). Considering parsimoniousness, we have to choose Model 8 to calculate the non-response adjustment factors.

Estimated adjustment factors and sizes of weighting effect for each model are shown in Table 3. The weighting effect is measured with

$$1 + \left(\frac{\hat{\sigma}_W}{\overline{W}}\right)^2,$$

where $\hat{\sigma}_W$ is the estimated standard error of the weight and $\overline{W}$ is the arithmetic average of the weights.

Estimated non-response adjustment factors in each cell for Model 1, $\hat{\rho}_{ij}^{-1}(M1)$ for example, is

$$\hat{\rho}_{ij}^{-1}(M1) = \frac{\hat{\pi}_{ij1} + \hat{\pi}_{ij2}}{\hat{\pi}_{ij1}},$$
$$= \frac{\eta\tau_i^A\tau_j^B\tau_1^C + \eta\tau_i^A\tau_j^B\tau_2^C}{\eta\tau_i^A\tau_j^B\tau_1^C},$$
$$= \frac{\tau_1^C + \tau_2^C}{\tau_1^C}.$$

The estimated adjustment factors for Model 1 $\hat{\rho}_{ij}^{-1}(M1)$ are a constant (1.6000) for all adjustment cells. Model 1 is equivalent to collapsing A and B into single-category variables respectively. Therefore Model 1 uses the marginal distribution of C to calculate the non-response adjustment factors. Since the weight adjustment was done by multiplying a constant, the weighting effect is 1 or no weighting effect. The weighting effect of Model 9 is 1.0661, the highest among possible models. The model (Model 8) with all two-way interaction or no-three way interaction model is the most parsimonious model (P-value = .2452) and the weighting effect of Model 8 is 1.0506, which is smaller than the one for Model 9. Estimated non-response adjustment factors in each cell for Model 8, $\hat{\rho}_{ij}^{-1}(M8)$ is

$$\hat{\rho}_{ij}^{-1}(M8) = \frac{\hat{\pi}_{ij1} + \hat{\pi}_{ij2}}{\hat{\pi}_{ij1}},$$
$$= \frac{\eta\tau_i^A\tau_j^B\tau_1^C\tau_{ij}^{AB}\tau_{i1}^{AC}\tau_{j1}^{BC} + \eta\tau_i^A\tau_j^B\tau_2^C\tau_{ij}^{AB}\tau_{i2}^{AC}\tau_{j2}^{BC}}{\eta\tau_i^A\tau_j^B\tau_1^C\tau_{ij}^{AB}\tau_{i1}^{AC}\tau_{j1}^{BC}}$$
$$= \frac{\tau_1^C\tau_{i1}^{AC}\tau_{j1}^{BC} + \tau_2^C\tau_{i2}^{AC}\tau_{j2}^{BC}}{\tau_1^C\tau_{i1}^{AC}\tau_{j1}^{BC}}.$$

The estimated cell-specific adjustment factors for Model 8, $\hat{\rho}_{ij}^{-1}(M8)$ depends on given levels of *i* and *j*.

Concluding Remarks

A new method of adjusting for non-response bias has been proposed. By applying the proposed method, we could develop non-response adjustment factors simply from the estimated cell frequencies under the most parsimonious model, given a set of variables which were observed for all sampled units.

References

Alho, J. M. (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika, 77*(3), 617-624.

Goodman, L. A. (1972). A general model for the analysis of surveys. *The American Journal of Sociology, 77*(6), 1035-1086.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association, 65*(329), 226-256.

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics, 6*(3), 287-296.

Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review / Revue Internationale De Statistique, 54*(2), 139-157.

Little, R. J. A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association, 77*(378), 237-250.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Sarndal, C. E. (1981). Frameworks for inference in survey sampling with applications to small area estimation and adjustment for nonresponse. *Bulletin of the International Statistical Institute*, 49, 494-513.

Table 1. Hypothetical Data

Variable			Frequency	Probability
A(*i*)	**B**(*j*)	**C**(*k*)	(n_{ijk})	(π_{ijk})
1	1	1	20	0.0294
1	1	2	10	0.0147
1	2	1	15	0.0221
1	2	2	15	0.0221
1	3	1	10	0.0147
1	3	2	20	0.0294
2	1	1	30	0.0441
2	1	2	10	0.0147
2	2	1	55	0.0809
2	2	2	15	0.0221
2	3	1	70	0.1029
2	3	2	30	0.0441
3	1	1	80	0.1176
3	1	2	30	0.0441
3	2	1	70	0.1029
3	2	2	30	0.0441
3	3	1	75	0.1103
3	3	2	95	0.1397

Table 2. Fitted Models and Goodness-of-fit Statistics

Variable			Model								
			9	8	7	6	5	4	3	2	1
A	B	C	{A,B,C, AB,AC, BC,ABC}	{A,B,C, AB,AC, BC}	{A,B,C, AC,BC}	{A,B,C, AB,BC}	{A,B,C, AB,AC}	{A,B,C, BC}	{A,B,C, AC}	{A,B,C, AB}	{A,B,C}
1	1	1	0.0294	0.0265	0.0202	0.0319	0.0221	0.0253	0.0175	0.0276	0.0219
1	1	2	0.0147	0.0176	0.0130	0.0123	0.0221	0.0097	0.0175	0.0165	0.0131
1	2	1	0.0221	0.0243	0.0218	0.0309	0.0221	0.0272	0.0195	0.0276	0.0243
1	2	2	0.0221	0.0198	0.0156	0.0132	0.0221	0.0117	0.0195	0.0165	0.0146
1	3	1	0.0147	0.0154	0.0241	0.0228	0.0221	0.0302	0.0292	0.0276	0.0365
1	3	2	0.0294	0.0287	0.0376	0.0213	0.0221	0.0282	0.0292	0.0165	0.0219
2	1	1	0.0441	0.0493	0.0697	0.0425	0.0434	0.0590	0.0603	0.0368	0.0511
2	1	2	0.0147	0.0095	0.0159	0.0163	0.0154	0.0227	0.0214	0.0221	0.0307
2	2	1	0.0809	0.0832	0.0751	0.0721	0.0760	0.0636	0.0670	0.0643	0.0568
2	2	2	0.0221	0.0197	0.0190	0.0309	0.0270	0.0272	0.0238	0.0386	0.0341
2	3	1	0.1029	0.0954	0.0831	0.0760	0.1085	0.0704	0.1006	0.0919	0.0852
2	3	2	0.0441	0.0517	0.0460	0.0711	0.0385	0.0659	0.0357	0.0551	0.0511
3	1	1	0.1176	0.1153	0.1012	0.1168	0.0958	0.1068	0.0876	0.1011	0.0925
3	1	2	0.0441	0.0464	0.0447	0.0449	0.0660	0.0411	0.0603	0.0607	0.0555
3	2	1	0.1029	0.0984	0.1090	0.1029	0.0871	0.1151	0.0973	0.0919	0.1027
3	2	2	0.0441	0.0487	0.0536	0.0441	0.0600	0.0493	0.0670	0.0551	0.0616
3	3	1	0.1103	0.1172	0.1207	0.1292	0.1480	0.1274	0.1460	0.1563	0.1541
3	3	2	0.1397	0.1328	0.1296	0.1208	0.1020	0.1192	0.1006	0.0938	0.0925
Degree of freedom			0	4	8	6	6	10	10	8	12
Goodness of fit Chi-Square			0	5.44	23.88	30.34	37.84	43.5	51.01	57.46	70.63
P-value			1	0.2452	0.0024	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Table 3. Nonresponse Adjustment factors and Weighting Effects

Variable		Model								
		9	8	7	6	5	4	3	2	1
A	B	{A,B,C, AB,AC, BC,ABC}	{A,B,C, AB,AC, BC}	{A,B,C, AC,BC}	{A,B,C, AB,BC}	{A,B,C, AB,AC}	{A,B,C, BC}	{A,B,C, AC}	{A,B,C, AB}	{A,B,C}
1	1	1.5000	1.6640	1.6410	1.3846	2.0000	1.3846	2.0000	1.6000	1.6000
1	2	2.0000	1.8166	1.7143	1.4286	2.0000	1.4286	2.0000	1.6000	1.6000
1	3	3.0000	2.8691	2.5591	1.9355	2.0000	1.9355	2.0000	1.6000	1.6000
2	1	1.3333	1.1925	1.2275	1.3846	1.3548	1.3846	1.3548	1.6000	1.6000
2	2	1.2727	1.2367	1.2535	1.4286	1.3548	1.4286	1.3548	1.6000	1.6000
2	3	1.4286	1.5419	1.5532	1.9355	1.3548	1.9355	1.3548	1.6000	1.6000
3	1	1.3750	1.4026	1.4416	1.3846	1.6889	1.3846	1.6889	1.6000	1.6000
3	2	1.4286	1.4951	1.4921	1.4286	1.6889	1.4286	1.6889	1.6000	1.6000
3	3	2.2667	2.1333	2.0741	1.9355	1.6889	1.9355	1.6889	1.6000	1.6000
Weighting Effect		1.0661	1.0506	1.0375	1.0254	1.0169	1.0254	1.0169	1.0000	1.0000