Scaling of Sampling Weights For Two Level Models in Mplus 4.2

September 2, 2008

1 Introduction

In survey analysis the data usually is not obtained at random. Units are typically sampled with unequal probability. In two level analysis both the cluster units and the individual units can be sampled with unequal probabilities. Sampling weights are assigned at one or both levels to reflect these unequal selection probabilities. Suppose that the probability of selection for cluster j is p_j. The between level sampling weight is then $w_j = 1/p_j$. Suppose that the probability of selection for individual i in cluster j, given that cluster j is selected, is p_{ij}. The within level sampling weight is then $w_{ij} = 1/p_{ij}$. The original sampling weights are typically rescaled before they are used in the estimation. Let's denote by w_j^* and w_{ij}^* the rescaled weights. These weights are used directly in the estimation method. The pseudo ML estimates obtained by the Mplus estimator **MLR** maximize the weighted pseudo likelihood

$$\prod_j \left(\int \left(\prod_i f_{ij}(y_{ij}|x_{ij},\eta_j)^{w_{ij}^*} \right) f_j(y_j|x_j,\eta_j)\phi(\eta_j|x_j)d\eta_j \right)^{w_j^*} \tag{1}$$

where y_{ij} and y_j are the dependent variables on the within and the between level and their likelihood functions are f_{ij} and f_j respectively, x_{ij} and x_j are within level and between level covariates and η_j is the between level random effect with likelihood function ϕ. More information on the estimation method can be found in Asparouhov and Muthen (2006) and Asparouhov (2006).

In this note we describe how the scaled weights are obtained from the raw unscaled weights and the different methods and options available in Mplus 4.2. The within weight variable is specified by the **weight=** command while the between weight variable is specified by the **bweight=** command. There are three methods for scaling the within level weights and two methods for scaling the between level weights. The within level weights scaling is specified by the **wtscale=** command and the between level weight scaling is specified by the **bwtscale=** command. The three possible options for the within level scaling are **unscaled, cluster,** and **ecluster**. The two options for the between level scaling are **unscaled** and **sample**.

2 Scaling the within level weights

The scaling of the within level weights influences the parameter estimates and their standard errors as well as other output results. Following is a detailed description of the three different ways to scale the within level weights.

- **unscaled**

  The scaled weight is the same as the raw weight

  $$w_{ij}^* = w_{ij}.$$

  This scaling method can be used in special situations where the scaling is done prior to the Mplus analysis and the raw weights are unchanged and used during the estimation.

- **cluster** (default)

  The scaled weight is obtained by

  $$w_{ij}^* = w_{ij} \frac{n_j}{\sum_i w_{ij}}$$

  where n_j is the number of sample units in cluster j. With this scaling method the new within weights add up to the cluster sample size

$\sum_i w^*_{ij} = n_j$. This is the default scaling method. The method generally achieves the most accurate results in most situations, see Asparouhov (2006), scaling method A.

- **ecluster**

  The scaled weight is obtained by

  $$w^*_{ij} = w_{ij} \frac{n^*_j}{\sum_i w_{ij}}$$

  where n^*_j is the effective cluster sample size for cluster j

  $$n^*_j = \frac{(\sum_i w_{ij})^2}{\sum_i w^2_{ij}},$$

  see Potthoff et alt. (1992). With this scaling method the new within weights add up to the effective cluster sample size n^*_j. This method generally achieves somewhat more accurate estimates for variance and covariance parameters than the default **cluster** method, see Asparouhov (2006), scaling method B, and Stapleton (2004).

3 Scaling the between level weights

The scaling of the between level weights generally does not influence the parameter estimates and their standard errors. It will however influence the log-likelihood value and fit indices that are based on the log-likelihood value. Following is a detailed description of the two methods for scaling the between level weights.

- **unscaled**

  The scaled weight is the same as the raw weight

$$w_j^* = w_j$$

  Here as well this method is useful for special analysis where the scaling is done prior to Mplus analysis.

- **sample** (default)

  The scaled weight is obtained by

$$w_j^* = w_j \frac{n}{\sum_{i,j} w_{ij}^* w_j}$$

  where n is the total sample size. With this scaling the new composite weights which is the product of the within and the between weights $w_j^* w_{ij}^*$ sums up to the total sample size

$$\sum_{i,j} w_j^* w_{ij}^* = n.$$

  This is the default scaling method, and generally will produce meaningful values for the fit indices.

4 References

The following is a list of articles relevant to the topic of weight scaling for twolevel analysis.

Asparouhov, T. (2006). General Multilevel Modeling with Sampling Weights. Communications in Statistics: Theory and Methods, Volume 35, Number 3, pp. 439-460(22).

Asparouhov, T. and Muthen, B. (2006). Multilevel Modeling of Complex Survey Data. ASA Proceedings of Survey Research Section.

Pfeffermann, D.; Skinner, C.J.; Holmes, D.J.; Goldstein, H.; Rasbash, J. (1998) Weighting for unequal selection probabilities in multilevel models. Journal of the Royal Statistical Society, Series B, 60, 23-56.

Potthoff, R.F.; Woodbury, M.A.; Manton, K.G. (1992) "Equivalent Sample Size" and "Equivalent Degrees of Freedom" Refinements for Inference Using Survey Weights Under Superpopulation Models. Journal of the American Statistical Association, 87, 383-396.

Stapleton, L. (2002) The Incorporation of Sample Weights Into Multilevel Structural Equation Models. Structural Equation Modeling, 9(4), 475-502.