

# Methods Used for Estimating Mixed-Effects Models in EdSurvey

*Developed by Trang Nguyen and Claire Kelley\*†*

*July 18, 2018*

The `EdSurvey` package gives users functions to analyze education survey data efficiently. This vignette describes the methods used to estimate mixed-effects models with plausible values and survey weights in `EdSurvey`.

## Note:

New users looking for an introduction and overview of `EdSurvey` should see *Using EdSurvey to Analyze NCES Data: An Illustration of Analyzing NAEP Primer*.<sup>1</sup>

## Introduction

The `EdSurvey` package fits a **Weighted Mixed** model, also known as a multilevel, mixed, or hierarchical linear model (HLM) using the `WeMix` package. The weights could be inverse selection probabilities, such as those developed for an education survey where schools are sampled probabilistically and then students inside these schools are sampled probabilistically. Although mixed-effects models are already available in R, `EdSurvey` is unique in implementing methods for mixed models using weights at multiple levels and allowing plausible values in the dependent variable. For a full treatment of the mathematics behind survey weighting mixed-effects models, see the `WeMix` package's vignette.

## Motivation

The principle advantage of mixed-effects models is the ability to appropriately model data with observations that are nested into hierarchies. Educational data are often nested or clustered; we may observe students in classrooms that are themselves nested in schools. To model an outcome with predictive variables measured at different levels (e.g., to model student-level outcomes using classroom or school-level characteristics), researchers often used mixed-effects models. Traditionally in a mixed model, fixed effects are used for variables whose effect does not differ across clusters, whereas random effects (which may be slopes or intercepts) are used to model variables that may have different effects in different clusters. Mixed-effects models are particularly common in psychology and behavioral science (McNeish, Stapleton, & Silverman, 2017).

Although common in education research, mixed-effects models certainly are not the only way to handle clustered data. Conventional approaches, such as linear regression models, can still account for nested data effects by including dummy variables for each group to mimic random effects, and researchers may use cluster robust standard errors or the jackknife variance estimates to account for covariance within groups (McNeish et al., 2017). However, although the jackknife covers variance estimation at the highest group level, mixed-effects models may be preferable to regression models because they account for pseudoreplication when estimating covariances and account for differences in cluster size when calculating standard errors. Moreover, mixed-effects models more easily allow more complex group-level effects including slopes that differ by group.

---

\*This publication was prepared for NCES under Contract No. ED-IES-12-D-0002 with American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

†The authors would like to thank Paul Bailey and Mike Cohen for reviewing this document.

<sup>1</sup>Available online at <http://www.air.org/sites/default/files/EdSurvey.pdf>

For examples of how NCES recommends using weighted mixed-effects models, as well as their summary of the mathematical background and description of hierarchical linear model's insufficiency in this case see appendix D in the NCES working paper on analysis of TIMSS data at <https://nces.ed.gov/pubs2001/200105.pdf>.

## Fitting Mixed-Effects Models

To provide a brief recap, the mixed model in its simplest single-level form can be expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (1)$$

where  $\mathbf{y}$  is the vector of outcomes,  $\mathbf{X}$  is a matrix of covariates associated with regressors that are assumed to be fixed,  $\boldsymbol{\beta}$  is the vector of fixed-effect regression coefficients,  $\mathbf{Z}$  is a matrix of covariates associated with regressors that are assumed to be random, and  $\mathbf{u}$  is the vector of random effects. The meaning of  $\mathbf{u}$  being random is simply that a level is shared within a group and across groups  $\mathbf{u} \sim MVN(\vec{\mathbf{0}}, \boldsymbol{\Sigma})$ , where  $MVN(\cdot, \cdot)$  is the multivariate-normal distribution,  $\vec{\mathbf{0}}$  is the mean vector of all zeros, and  $\boldsymbol{\Sigma}$  is the covariance matrix of the  $MVN$ .

To generalize to the multilevel case, when there are more than two levels, eq. 1 can be rewritten as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{l=2}^L \mathbf{Z}^{(l)}\mathbf{u}^{(l)} + \boldsymbol{\epsilon} \quad (2)$$

where a superscript ( $l$ ) is added to  $\mathbf{Z}$  and  $\mathbf{u}$  to indicate that they are at the  $l$ th level. Note: In the summation,  $l$  starts at  $l = 2$  because random effects cannot be at the lowest level of observation ( $l = 1$ ).

This model is then estimated using maximum likelihood estimation using adaptive quadrature following the methodology of Rabe-Hesketh & Skrondal (2006). Details can be found in the `WeMix` package's vignette.

## Adjusting Weights Before Fitting Mixed-Effects Models

As discussed in Rabe-Hesketh & Skrondal (2006), using unscaled weights in mixed models can lead to biased estimates, especially when level-1 weights are very different from 1 and cluster size (e.g., the number of students within a school) is small. The following are the two most common scaling methods for level-1 weights, named Method A and B for consistency with the literature.

Method A:

$$w_{ij}^* = w_{ij} \left( \frac{n_j}{\sum_i w_{ij}} \right) \quad (3)$$

where  $w_{ij}$  are the full sample weights,  $i$  indexes the individuals,  $j$  indexes the groups, and  $n_j$  represents the number of observations in group  $j$ .

Method B:

$$w_{ij}^* = w_{ij} \left( \frac{\sum_i w_{ij}}{\sum_i w_{ij}^2} \right) \quad (4)$$

The default scaling method for PISA in the `EdSurvey` package is Method A for level one weight and no scaling for higher level weights, following the example in Rabe-Hesketh et al. (2006).

TIMSS Method:

$$w_{ij}^* = \frac{\text{total student weight}}{\text{adjusted school-level weight}} \quad (5)$$

For TIMSS, TIMSS Advanced, and PIRLS, adjusted student weights are calculated by dividing the total student weight (TOTWGT) by the adjusted school-level weight (SCHWGT) as explained in the *TIMSS 2015*

*codebook*.<sup>2</sup> No further scaling is applied to these weights, and second-level weights are used without scaling as well.

To implement other weighting schemes, users can choose to scale weights outside the function using their preferred scaling method and turn off the weight transformation in `mixed.sdf` by setting `weightTransformation=FALSE`. Please refer to Appendix A for an example workflow.

For surveys such as NAEP, there have not been any widely used scaling methods in the literature, so the `EdSurvey` package does not do any weight transformation. For surveys with at least some recommendations, such as TALIS (Carsten, 2009), you can see the suggested methods in Table 1.

For other surveys, users should create their own weights following the weighting scheme recommended previously.

## Mixed-Effects Models With Plausible Values

In the case of plausible values, consider a model where the dependent variable  $\mathbf{y}$  has  $K$  plausible values. The coefficients  $\beta$  of the plausible values are simply the average of the coefficient calculated for each plausible value. that is:

$$\beta = \frac{\sum_{k=1}^K \beta_k}{K} \quad (6)$$

The variance of the estimated parameters is calculated as the sum of the imputation variance and the sampling variance.

The imputation variance is defined as follows:

$$Var_{Imputation} = \frac{K+1}{(K-1)K} \sum_{k=1}^K (\beta_k - \beta)^2 \quad (7)$$

And the sampling variance is defined as follows:

$$Var_{Sampling} = \frac{\sum_{k=1}^K var(\beta_k)}{K} \quad (8)$$

where  $var(\beta)$  is the variance of each plausible value model calculated using a sandwich estimator. The total variance is then the sum of the imputation and sampling variance. The intraclass correlation also is then calculated as the average of the intraclass correlation across all models.

---

<sup>2</sup>More information can be found at <https://timssandpirls.bc.edu/publications/timss/2015-methods/chapter-15.html>.

**Table 1. Recommended Weighting Scheme for EdSurvey Data**

Survey	Implemented in EdSurvey	Recommended Scaling	Citation
PISA	Yes	Method A for level 2; no scaling for level 2	Koretz, McCaffery, & Sullivan (2001); TIMSS (2015)
TIMSS	Yes	TIMSS Method	Koretz et al. (2001); TIMSS (2015)
TIMSS Advanced	Yes	TIMSS Method	Koretz et al. (2001); TIMSS (2015)
PIRLS	Yes	TIMSS Method	Koretz et al. (2001); TIMSS (2015)
TALIS	No	First for teacher-level weights scale with $\frac{\text{Final Teacher Weight}}{\text{Final School Weight}}$ and then apply Method A for student level weights	Carstens (2009)
ECLS-K	No	No suggestions for multilevel models	
ELS	No	No suggestions for multilevel models	
CIVED	No	User guide does not provide suggestions for multilevel models, but suggest that sample design and weighting are the same as TIMSS.	Schulz & Sibberns (2004)
ICCS	No	Student weights should be calculated as a product of class and student weight factors; school weights should be calculated as a product of school weight factors and level-one weight factors.	Köhler, Weber, Brese, Schulz, & Carstens (2016)
ICILS	No	Level one weights should be calculated as within-school unit weights, so for teacher-level analysis, this would be the product of teacher-level weighting factor or for student-level analysis, it would be the product of student level weighting factors. For level two, weights are the selection probability of the school adjusted for nonresponse. So for a student level analysis this would be $WGTFAC1 \times WGTADJ1S$ ; for teacher-level analysis, it would be $WGTFAC1 \times WGTADJ1T$ .	Jung & Carstens (2015)
PIAAC	No	No suggestions for multilevel models	

## Appendix A: Example Workflow With Scaling Weights to User Specification

This example illustrates how the user might implement the student-level weighting as described in the section on adjusting weights section when using a survey (NAEP in this example) that does not have a weighting scheme previously implemented.

```
library(EdSurvey)

# Use NAEP Primer data
sdf <- readNAEP(system.file("extdata/data", "M36NT2PM.dat", package = "NAEPprimer"))

# Subset data to a sample of interest (this step is optional)
sdf <- subset(sdf, scrpsu < 500)

# Extract variables of interest to a light.edsurvey.data.frame
lsdf <- getData(sdf, c("composite", "dsex", "b017451", "scrpsu", "origwt", "smsrswt"),
                addAttributes=TRUE)

# Transform weights using your method (Note that this method is not recommended for NAEP)
lsdf$pwt1 <- lsdf$origwt/lsdf$smsrswt
lsdf$pwt2 <- lsdf$smsrswt

# fast is an argument from WeMix::mix that allows the function to run faster using c++
m1 <- mixed.sdf(composite ~ dsex + b017451 + (1|scrpsu), data=lsdf,
                weightVar = c('pwt1', 'pwt2'), fast=TRUE, verbose=1)

# print out result summary
summary(m1)

# Output
# Call:
# mixed.sdf(formula = composite ~ dsex + b017451 + (1 | scrpsu),
#   data = lsdf, weightVars = c("pwt1", "pwt2"),
#   verbose = 1, fast = TRUE, nQuad = 7)
#
# Formula: composite ~ dsex + b017451 + (1 | scrpsu)
#
#
# Plausible Values: 5
# Number of Groups:
#   Group Var Observations Level
# 1   scrpsu           22      2
# 2 Residual           492      1
#
# Variance terms:
#
#               variance Std. Error Std.Dev.
# scrpsu:(Intercept) 661.9544  248.80786 25.72847
# Residual           874.0728   93.27785 29.56472
#
# Fixed Effects:
#
#               Estimate Std. Error t value
# (Intercept)    263.56137    7.73858 34.0581
```

```
# dsexFemale          -0.79063    3.15192 -0.2508
# b017451Once every few weeks  1.83987    6.47586  0.2841
# b017451About once a week    10.17583   4.67462  2.1768
# b0174512 or 3 times a week   9.81466   5.54558  1.7698
# b017451Every day           7.08593   6.38682  1.1095
#
# Intraclass Correlation= 0.430
```

## Appendix B: Examples of different mixed-model formulas

The `mixed.sdf` function in the `EdSurvey` package (and the `mix` function in the `WeMix` package) use the same formula syntax as in `lme4` package. The following table specifies all formulas currently supported by `EdSurvey 2.2`. Note that  $g$  denotes the grouping factor and  $x$  denotes a covariate or predictor. In R formulas, 1 often denotes *Intercept*. Because this appendix focuses on the formula specification, the *Example* does not show the full code of `mixed.sdf`. Please refer to the function documentation or package manual to fill in the missing part.

Formula	Alternative	Meaning	Example
$(1 \mid g)$	$1 + (1 \mid g)$	Random intercept at level 2 for group variable $g$ (null model)	<code>mixed.sdf(composite ~ (1   scrpsu), ...)</code>
$x + (1 \mid g)$	$x + 1 + (1 \mid g)$	Random intercept with fixed slope for predictor $x$	<code>mixed.sdf(composite ~ dsex + (1   scrpsu), ...)</code>
$x + (x \mid g)$	$1 + x + (1 + x \mid g)$	Random intercept and random slope for predictor $x$ (coefficients associated with intercept and random-effect predictor $x$ are correlated)	<code>mixed.sdf(composite ~ dsex + (dsex   scrpsu), ...)</code>
$x + (x \parallel g)$	$1 + x + (1 \mid g) + (0 + x \mid g)$	Random intercept and random slope for predictor $x$ (coefficients associated with intercept and random-effect predictor $x$ are uncorrelated)	<code>mixed.sdf(composite ~ dsex + (dsex    scrpsu), ...)</code>

## References

- Carstens, R. (Ed.). (2009). *TALIS user guide for international database*. Paris, France: Organization for Economic Cooperation and Development.
- Jung, M. & Carstens, R. (Eds.). (2015). *ICILS 2013 user guide for the international database*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.
- Köhler, H., Weber, S., Brese, F., Schulz, W., & Carsten, R. (Eds.). (2016). *ICCS 2016 user guide for the international database*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.
- Koretz, D., McCaffrey, D., & Sullivan, T. (2001). Using TIMSS to analyze correlates of performance variation in mathematics (Working Paper 2001–05). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- McNeish, D., Stapleton, L. M., Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modelling. *Psychological Methods*, 22(1), 114–140
- Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 169(4), 805–827.
- Schulz, W., & Sibberns, H. (2004). *IEA civic education study technical report*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.
- TIMSS. (2015). *Assessment Frameworks*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.