EXPLORING THE IMPACT OF CLUSTER MEAN CENTERING ON DEVIANCE IN MULTI-LEVEL MODELS

Final Project Paper for PSYC 575

Meltem Ozcan

Department of Psychology

University of Southern California

2021-12-09

1. Introduction

1.1 Background

Multi-level models, or hierarchical linear models, are statistical models that can help account for the dependence between observations (and the resulting inflation of Type I errors) that arises when data are organized at more than one level [7] [4] (e.g., nesting of individuals within districts [2], nesting of measurements within individuals over time [6] etc.). While specifying a multi-level model, a number of decisions need to be made beyond deciding which variables should be used to predict the outcome, such as whether to include random effects, whether to include or omit cluster means in model equations, and whether level-1 variables should be centered. Each of these decisions need to be justified theoretically and/or empirically [4]. One such decision is whether to include the raw as opposed to the cluster mean centered version of a level-1 predictor in a model. Consider the following two multilevel models for data where observations i are grouped within $j \in n_j$ clusters:

Model 1: Contextual

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}\mathrm{x}_{1ij} + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = \gamma_{10} + u_{1j} \end{aligned}$$

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{01}\overline{\mathrm{x}_{1.j}} + \gamma_{10}\mathrm{x}_{1ij}}_{\text{fixed}} + \underbrace{u_{0j} + u_{1j}\mathrm{x}_{1ij} + e_{ij}}_{\text{random}} \qquad (1)$$

Model 2: Between-Within

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = \gamma_{10} + u_{1j} \end{aligned}$$

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{01}\overline{\mathrm{x}_{1.j}} + \gamma_{10}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}})}_{\text{fixed}} + \underbrace{u_{0j} + u_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + e_{ij}}_{\text{random}} \quad (2)$$

Here, x_{1ij} denotes the value x_1 takes for the i-th observation in the j-th cluster, $\overline{\mathrm{x}_{1.j}}$ denotes the j-th cluster mean, and $(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}})$ indicates the cluster mean centered x_1 variable. γs refer to the fixed effects, and γ_{00} in particular denotes the grand mean. Randomness is introduced into the models through e_{ij}, the random error for the i-th observation in the j-th cluster; u_{0j}, the random intercept; and u_{1j}, the random slope component for variable x_1.

For the purposes of this project, we will assume $e_{ij} \sim N(0, \sigma)$, and that u_{0j} and u_{1j} follow a bivariate normal distribution with mean zero and covariance matrix $[[\tau_0^2, \tau_{02}], [\tau_{02}, \tau_2^2]]$ where τ_0^2 denotes the random intercept variance, τ_2^2 denotes the random slope variance, and τ_{02} denotes the random intercept-slope covariance. Barring the contribution from u_{1j} (i.e., when no random slopes are included in the model), predicting Y_{ij} using x_{1ij} and $\overline{\mathrm{x}_{1.j}}$ (as in the contextual model formulation) is equivalent to predicting Y_{ij} using $(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}})$ and $\overline{\mathrm{x}_{1.j}}$ (as in the between-within formulation) ([7], [3]). Both models tease apart the individual-level and cluster-level effects, and differ mainly in the interpretation of the coefficient estimates. The contextual model formulation can allow us to compare two individuals with the same score on some level-1 variable but belong to different clusters that are on average one unit apart; while the between-within model can allow us to examine the level-1 and level-2 effects separately by removing level-2 effects from level-1. In this project, our main goal was to empirically investigate how the decision to include the raw (x_{1ij} and $\overline{\mathrm{x}_{1.j}}$) as opposed to the cluster mean centered version ($(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}})$ and $\overline{\mathrm{x}_{1.j}}$) of predictors in a model with random slopes affects the comparative fit of models when the data were generated using either the raw or the cluster mean centered version of the predictor. We were particularly interested in the deviance statistic as a measure of the lack of fit between the model and the data [7].

Denoting the maximum log likelihood achievable under the full model (where each data point has a corresponding parameter in the model, resulting in a perfect fit) as $\mathrm{LL}_{\text{full}}$, and the maximum log likelihood achieved under the model of interest as $\mathrm{LL}_{\text{fitted}}$, negative two times the difference between $\mathrm{LL}_{\text{fitted}}$ and $\mathrm{LL}_{\text{full}}$ is defined as *deviance*, or the likelihood ratio test statistic comparing the full model and the model of interest: $\mathrm{D}_{\text{fitted}} = -2 \times (\mathrm{LL}_{\text{fitted}} - \mathrm{LL}_{\text{full}})$ [5], [1]. Deviance computed for a particular model fit only gains meaning in comparison to another deviance value computed for a nested model (i.e., the comparison between deviances for two fitted models, $\mathrm{D}_{\mathrm{FM}1}$ and $\mathrm{D}_{\mathrm{FM}2}$). The difference between two deviance values when model parameters are estimated via maximum likelihood approximately follows a χ^2 distribution with degrees of freedom given by the number of additional parameters in the second model, and it can be used as a test statistic that can be used to compare two nested models fitted to the same data set [5]. A lower deviance value signals a better fit for the data.

1.2 Motivating Example

Our motivating example uses data from the High School and Beyond Survey (HSBS) conducted by the U.S. Department of Education in 1982 [8]. As part of HSBS, 7,185 10-12th grade students were sampled from 160 schools. Each student's achievement level in mathematics, minority status and socio-economic status (SES) was recorded along with information pertaining to the student's school, such as school size and whether over 40% of the student body identified as minority. Consider the following two models predicting students' math achievement using their SES and minority status:

Model 1: Contextual

$$
\begin{aligned}
\text{Level 1:}\quad & \text{math}_{ij} = \beta_{0j} + \beta_{1j}(\text{ses}_{ij} - \overline{\text{ses}}_{.j}) + \beta_{2j}\text{min}_{ij} + e_{ij} \\
\text{Level 2:}\quad & \beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\text{ses}}_{.j} + \gamma_{02}\overline{\text{min}}_{.j} + u_{0j} \\
& \beta_{1j} = \gamma_{10} \\
& \beta_{2j} = \gamma_{20} + u_{2j}
\end{aligned}
$$

Model 2: Between-within

$$
\begin{aligned}
\text{Level 1:}\quad & \text{math}_{ij} = \beta_{0j} + \beta_{1j}(\text{ses}_{ij} - \overline{\text{ses}}_{.j}) + \beta_{2j}(\text{min}_{ij} - \overline{\text{min}}_{.j}) + e_{ij} \\
\text{Level 2:}\quad & \beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\text{ses}}_{.j} + \gamma_{02}\overline{\text{min}}_{.j} + u_{0j} \\
& \beta_{1j} = \gamma_{10} \\
& \beta_{2j} = \gamma_{20} + u_{2j}
\end{aligned}
$$

Both models use the within-between formulation for the ses variable, and the main difference between these two models is whether the raw (min_{ij}, $\overline{\text{min}}_{.j}$) or the cluster mean centered minority variable (($\text{min}_{ij} - \overline{\text{min}}_{.j}$), $\overline{\text{min}}_{.j}$) and the associated random slopes are included in the model. Figure 1 illustrates the $anova$ output comparing these two models when used to predict math achievement for students from a random subset of 10 schools. We notice that there is a 1.7-point difference in deviance between the two formulations which signals that the between-within formulation was a better fit for the data at hand. As the ground truth is unknown for real-world data, it is not possible to determine whether the between-within formulation was a better fit for this particular data set or whether the between-within formulation would always produce a lower deviance score compared to the contextual formulation. This led us to the following research questions:

- Is the between-within formulation always favored over the contextual effects formulation when deviance is the metric of comparison? If not,

- Can deviance consistently pick out the data generating model (DGM) as the better model? If not,

- Are there specific circumstances that this holds true (i.e., that for certain parameter values we can confidently use deviance as a model comparison metric)?

2. Method

In order to answer these questions, we conducted two simulation studies in which we generated multi-level data using the contextual model (DGM=C) and compared model fit and deviance under the contextual (FM=C) versus the between-within (FM=BW) model for a large number of iterations (niter=1000 for all simulations), and then repeated this process for data generated using the between-within model (DGM=BW). For either DGM, we computed the proportions the following comparisons held true out of the 1000 iterations: $D_{FM=BW} < D_{FM=C}, D_{FM=BW} > D_{FM=C}, D_{FM=BW} == D_{FM=C}$, and repeated this process for various parameter specifications (see Figure 2 for a diagram of main steps in this process). In Study 1, we generated data that paralleled the structure of the HSBS data set with either a contextual or between-within formulation, and manipulated cluster size and number of cluster combinations. In the first part of Study 2, we generated data using a binary predictor, and examined the proportions under different random intercept and slope parameters. In the second part of Study 2, we repeated the analyses for data generated using a normally distributed predictor, under different random intercept and random slope parameters. In each study, parameters aside from those being manipulated were held constant.

All analyses were performed using R, and the data analytic scripts and supplemental materials for this project are available at `https://github.com/meltemozcan/575project`.

3. Study 1

3.1 Method

We first generated a data set where x_1 is a quantitative variable and x_2 is a binary variable, paralleling the ses and minority variables in HSBS. We assigned parameter values based on estimates computed using HSBS, and we used the same parameter estimates to generate the contextual data set and the between-within data set. The generation of the two data sets differed only in whether the binary variable x_2 and its random slopes were incorporated in the raw or cluster-mean-centered forms.

We started with the assumption that $\mathrm{x}_{1ij} \sim N(\mu_{x_1}, \sigma_{x_1})$ where μ_{x_1} is the mean of x_{1ij}, σ_{x_1} is the standard deviation of x_{1ij}. We assumed that the effect of x_1 varies randomly across clusters such that μ_{x_1} follows a normal distribution with a mean that equals the observed mean of ses_cm and standard deviation that equals the standard deviation of ses_cm: $\mu_{x_1} \sim N\left(\overline{\text{ses\_cm}}, \text{sd}(\text{ses\_cm})\right)$. Furthermore, we assumed that σ_{x_1} follows a normal distribution with a mean that equals the mean of the standard deviation of ses for each cluster and a standard deviation that equals the standard deviation of the cluster standard deviations of ses: $\sigma_{x_1} \sim N\left(\overline{\text{sd}(\text{ses}_{.j})}, \text{sd}(\text{sd}(\text{ses}_{.j}))\right)$. Plugging in the appropriate parameter estimates (see Figure 3), $\mathrm{x}_{1ij} \sim N(\mu_{x_1}, \sigma_{x_1})$ where $\mu_{x_1} \sim N(-0.01, 0.41),$ and $\sigma_{x_1} \sim N(.67, 0.1)$.

The probability that a student will identify as minority is dependent on their school. Taking a top-down approach, we first allocated schools to the categories of high or low minority through a third variable x_3 that parallels the *himnity* variable in HSBS, which classifies schools as high or low minority using a 40% threshold ($himnty = 1$ indicates a school where greater than 40% of the students identify as minority). As a school will either have over 40% minority identification in its student body or not, x_{3j} can be modelled with the Bernoulli distribution: $\mathrm{x}_{3j} \sim \text{Ber}\left(1 - \overline{\text{himnty}}\right)$ We further assumed that the effect of x_2 varies across clusters such that, for clusters where $\mathrm{x}_{3j} = 1$, $p_j \sim \text{Unif}(p_{high})$ and for clusters where $\mathrm{x}_{3j} = 0$, $p_j \sim \text{Unif}(p_{low})$. The ranges of p_{high} and p_{low} were set in accordance with the 40% threshold used for himnty. As a student will either identify as minority or not, x_{2ij} can be modelled using the Bernoulli distribution ($\mathrm{x}_{2ij} \sim \text{Ber}(p_j)$), or the Binomial distribution (($\mathrm{x}_{2j} \sim \text{Bin}(n_j, p_j)$ where n_j is the cluster size). Plugging in the relevant estimates from the HSB data set (Figure 3), we set the following distributional assumptions for x_2 and x_3:

$$\mathrm{x}_{3j} \sim \text{Bern}(0.28) \text{ or, equivalently, } \mathrm{x}_3 \sim \text{Bin}(n_j, 0.28).$$

$$\mathrm{x}_{2j} \sim \text{Bin}(n_j, p_j) \text{ where } p_j \sim \begin{cases} \text{Uniform}(0, .4) & \text{if} \quad \mathrm{x}_{3j} = 0 \\ \text{Uniform}(.4, 1) & \text{if} \quad \mathrm{x}_{3j} = 1. \end{cases}$$

We set the values of the remaining model parameters using rough estimates from *lmer* outputs on a subset of HSBS data: $\gamma_{00} = 14, \gamma_{01} = 2, \gamma_{02} = -0.5, \gamma_{10} = 1.5, \gamma_{20} = -3, \tau_{01} = 0.01, \tau_0^2 = 0.5, \tau_1^2 = 3.5$. Furthermore, we set the random error term mean at zero and the random error variance at 5 ($e_{ij} \sim N(0, 5)$).

Recapping the above, our data generating models were as follows:

Generating Model: Contextual

$$\begin{array}{ll} \text{Level 1:} & Y_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + \beta_{2j}\mathrm{x}_{2ij} + e_{ij} \\ \text{Level 2:} & \beta_{0j} = 14 + 2\overline{\mathrm{x}_{1.j}} - 0.5\overline{\mathrm{x}_{2.j}} + u_{0j} \\ & \beta_{1j} = 1.5 \\ & \beta_{2j} = -3 + u_{2j} \end{array}$$

Generating Model: Between-Within

$$\begin{array}{ll} \text{Level 1:} & Y_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + \beta_{2j}(\mathrm{x}_{2ij} - \overline{\mathrm{x}_{2.j}}) + e_{ij} \\ \text{Level 2:} & \beta_{0j} = 14 + 2\overline{\mathrm{x}_{1.j}} - 0.5\overline{\mathrm{x}_{2.j}} + u_{0j} \\ & \beta_{1j} = 1.5 \\ & \beta_{2j} = -3 + u_{2j} \end{array}$$

For both models, we assumed that $e_{ij} \sim N(0, 5)$, and $\begin{pmatrix} u_{0i} \\ u_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 0.5 & 0.01 \\ 0.01 & 3.5 \end{bmatrix}\right)$. We varied cluster size and number of clusters as below:

$$n_j = 100, N \in \{10, 25, 50, 75, 100\}$$
$$N = 100, n_j \in \{10, 25, 50, 75, 100\}$$

3.2 Results

3.2.1 Example

We start by exploring two data sets generated using the process outlined above as an example of the type of data generated in Study 1. Figures 4 and 5 illustrate the distribution of the y variable for subset of 10 clusters. For this subset of the data, we see that the cluster means for the data generated using the between-within model fall closer to the grand mean of $y = 14$ compared to the cluster means for the data generated using the contextual model, suggesting that the random slopes may have more weight in the contextual model when the distributional parameters are held constant across the conditions. This is expected, as the random slope is multiplied by the non-centered x_2 variable in the contextual model, producing a larger value overall. We next fit the contextual and between-within models to the two subsets (see Table 1 for a table of the model coefficients). An examination of Figures

6 and 7, which illustrate the association between x_1 and y under each DGM-FM combination, shows that there is little difference between FM=C and FM=BW, and we observe less variance between the random slopes of models fitted to data where DGM=BW.Table 2 illustrates the deviance values comparing the contextual model with the between-within model under the two data conditions. The deviance values are very close across the models for both DGM conditions, and suggest a better fit to data using the contextual model.

3.3 Simulations

Tables 3 and 4 illustrate the proportions out of the 1000 the comparisons noted above held true. Looking at Table 3, we observe that the between-within model produced a lower deviance than the contextual model for all $N \in \{10, 25, 50, 75, 100\}$ when $n_j = 100$. Similarly, we found that the between-within model produced a lower deviance than the contextual model for all $n_j \in \{10, 25, 50, 75, 100\}$ when $N = 100$.

4. Study 2 Part 1: Binary predictor

4.1 Method

In Study 2, we moved away from the HSBS data set and parameter specifications, and simplified our data generating process and model parameters to more clearly examine how including the raw versus cluster mean centered version of a variable influences model fit. Models 1 and 2 predicted the outcome Y_{ij} using a single binary variable, x_1, which was distributed with $\mathrm{x}_{1j} \sim \mathrm{Bin}(n_j, p_j)$ where $p_j \sim \begin{cases} \mathrm{Uniform}(0, .4) & \text{if} \quad \mathrm{x}_{2j} = 0 \\ \mathrm{Uniform}(.4, 1) & \text{if} \quad \mathrm{x}_{2j} = 1. \end{cases}$
and $\mathrm{x}_2 \sim \mathrm{Bin}(n_j, 0.20)$.

We decided on the following fixed parameter values for both the contextual and the between-within models: $\gamma_{00} = 10, \gamma_{01} = 1, \gamma_{10} = 5, \sigma = 1$. Plugging in these values,

Model 1: Contextual

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}\mathrm{x}_{1ij} + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = 10 + \overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = 5 + u_{1j} \end{aligned}$$

Model 2: Between-Within

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = 10 + \overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = 5 + u_{1j}. \end{aligned}$$

For both models, we assumed $e_{ij} \sim N(0,1)$ and $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right)$.

Holding the number of clusters and cluster size constant at $N = 10$ and $n_j = 50$, we investigated the impact of

1) varying τ_1^2 and τ_0^2 for a fixed τ_{01} ($\tau_1^2 \in \{1, 5, 10\}, \tau_0^2 \in \{0.1, 1\}, \tau_{01} = 0$),

2) varying τ_1^2 and τ_{01} for a fixed τ_0^2 ($\tau_1^2 \in \{1, 5, 10\}, \tau_0^2 = 0, \tau_{01} \in \{0, 0.5\}$), and

3) varying the ratio of τ_1^2 to τ_0^2 ($\tau_1^2/\tau_0^2 \in \{10, 100\}$)

$(\tau_0^2, \tau_1^2 \in \{(0.01, 1), (0.1, 10), (1, 100), (0.1, 1), (1, 10), (10, 100)\}, \tau_{01} = 0)$

on the proportions the following comparisons hold true out of the 1000 iterations: $D_{FM=BW} < D_{FM=C}, D_{FM=BW} > D_{FM=C}, D_{FM=BW} == D_{FM=C}$ for data generated using the BW versus the C formulation.

4.2 Results

5. Study 2 Part 2: Normal predictor

5.1 Method

In Study 2 Part 2, we repeated the analyses in Study 2 Part 1 with a normally distributed variable x_1 rather than a binary variable. We assumed that $\mathrm{x}_{1ij} \sim N(\mu_{x_1}, \sigma_{x_1})$ where μ_{x_1} is the mean of x_{1ij}, σ_{x_1} is the standard deviation of x_{1ij}. We assumed that the effect of x_1 varies randomly across clusters such that $\mu_{x_1} \sim N(1, 5)$ and $\sigma_{x_1} \sim N(1, 1)$.

We decided on the following fixed parameter values for both the contextual and the between-within models: $\gamma_{00} = 10, \gamma_{01} = 1, \gamma_{10} = 5, \sigma = 1$. Plugging in these values,

Model 1: Contextual

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}\mathrm{x}_{1ij} + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = 10 + \overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = 5 + u_{1j} \end{aligned}$$

Model 2: Between-Within

$$\begin{aligned} \text{Level 1:} \quad & Y_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{x}_{1ij} - \overline{\mathrm{x}_{1.j}}) + e_{ij} \\ \text{Level 2:} \quad & \beta_{0j} = 10 + \overline{\mathrm{x}_{1.j}} + u_{0j} \\ & \beta_{1j} = 5 + u_{1j}. \end{aligned}$$

For both models, we assumed $e_{ij} \sim N(0,1)$ and $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right)$.

We investigated the impact of

1) varying τ_1^2 and τ_0^2 for a fixed τ_{01} ($\tau_1^2 \in \{1,5,10\}, \tau_0^2 \in \{0.1,1\}, \tau_{01} = 0$),

2) varying τ_1^2 and τ_{01} for a fixed τ_0^2 ($\tau_1^2 \in \{1,5,10\}, \tau_0^2 = 0, \tau_{01} \in \{0, 0.5\}$), and

3) varying the ratio of τ_1^2 to τ_0^2 ($\tau_1^2/\tau_0^2 \in \{10, 100\}$)

$(\tau_0^2, \tau_1^2 \in \{(0.01,1), (0.1,10), (1,100), (0.1,1), (1,10), (10,100)\}, \tau_{01} = 0)$

on the proportions the following comparisons hold true out of the 1000 iterations: $D_{FM=BW} < D_{FM=C}, D_{FM=BW} > D_{FM=C}, D_{FM=BW} == D_{FM=C}$ for data generated using the BW versus the C formulation. As before, we held the number of clusters and cluster size constant at $N = 10$ and $n_j = 50$.

5.2 Results

6. Discussion

The goal of this project was investigate whether the difference in deviance computed for multi-level models with the contextual versus between-within formulation fitted to the same data set can provide useful information for model specification and model comparison, or whether the between-within model will always produce a smaller deviance compared to the contextual model, and if so, why. We were specifically interested in exploring whether the direction and magnitude of the difference in deviation could be used to help guide model specification decisions such as whether the cluster-mean centered variable or the original variable should be included as random slopes.

We demonstrated that the between-within formulation is not always favored over the contextual effects formulation when deviance is the metric of comparison. We illustrated that deviance does not consistently pick out the data generating model as the better model. It appears from the results in Study 2 that when the random slope estimate is larger in magnitude deviance may be a more reliable metric for determining the DGM when the underlying model is contextual.

In future investigations, different fixed parameter values should be studied as well as different ratios for the random variables, and cluster size-number of cluster combinations. A better method may be to also put priors on these parameter values rather than fixing them arbitrarily. Furthermore, examining ICC and varying ICC may be an important next step. In addition, the effect size of the comparison between proportions, and the magnitude of the differences should be investigated.

References

[1] A. Agresti. *An Introduction to Categorical Data Analysis, Second Edition*. Wiley Series in Probability and Statistics. 2007. ISBN: 9780471226185.

[2] D. Ballas and M. Tranmer. "Happy People or Happy Places? A Multilevel Modeling Approach to the Analysis of Happiness and Well-Being." In: *International Regional Science Review* 35(1) (2012). DOI: `https://doi.org/10.1177/0160017611403737`.

[3] I.G. Kreft, J. de Leeuw, and L.S. Aiken. "The Effect of Different Forms of Centering in Hierarchical Linear Models." In: *Multivariate Behavioral Research* 30(1) (1995). DOI: `https://doi.org/10.1207/s15327906mbr3001_1`.

[4] D.B. McCoach. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. 2019. Chap. Multilevel Modeling. ISBN: 9781315755649.

[5] P. McCullagh and J.A. Nelder. *Generalized Linear Models, (2nd ed.)* Vol. 37. Monographs on Statistics and Applied Probability. Chapman and Hall, 1983. ISBN: 0412238500. DOI: `https://doi.org/10.1201/9780203753736`.

[6] J.S. Simons, T.A. Wills, and D.J. Neal. "The many faces of affect: a multilevel model of drinking frequency /quantity and alcohol dependence symptoms among young adults." In: *Journal of Abnormal Psychology* 123(3) (2014). DOI: `https://doi.org/10.1037/a0036926`.

[7] T. A. B. Snijders and R.J. Bosker. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, (2nd ed.)* London: Sage Publications, 2012.

[8] National Center for Education Statistics U.S. Department of Education Institute of Education Sciences. *High School & Beyond Survey*. URL: `https://nces.ed.gov/surveys/hsb/`.

7. Tables

Table 1: Table of model coefficients for Example 1. The first two columns refer to FM=C, FM=BW when DGM=C. The last two columns refer to FM=C, FM=BW when DGM=BW.

	contextual.ex1	between_within.ex1	contextual.ex2	between_within.ex2
(Intercept)	14.297	14.273	13.878	13.899
	(0.325)	(0.287)	(0.307)	(0.302)
x1_cmc_ij	1.591	1.587	1.624	1.626
	(0.262)	(0.262)	(0.237)	(0.237)
x1_cm_j	2.483	2.582	1.965	2.079
	(0.404)	(0.408)	(0.565)	(0.548)
x2_ij	-2.290		-2.851	
	(0.518)		(0.397)	
x2_cm_j	-1.337	-3.780	2.743	-0.188
	(0.791)	(0.620)	(0.897)	(0.785)
SD (Intercept)	0.416	0.159	0.000	0.000
SD (x2_ij)	0.809		0.272	
Cor (Intercept~x2_ij × cluster)	-1.000			
SD (Observations)	5.222	5.224	5.104	5.106
x2_cmc_ij		-2.251		-2.855
		(0.534)		(0.387)
SD (x2_cmc_ij)		0.887		0.008
Cor (Intercept~x2_cmc_ij × cluster)		-1.000		
Num.Obs.	1000	1000	1000	1000
R2 Marg.	0.167	0.174	0.098	0.099
R2 Cond.				
AIC	6166.6	6166.6	6116.6	6116.7
BIC	6210.7	6210.7	6160.8	6160.9
RMSE	5.21	5.22	5.10	5.11

Table 2: Deviance values by FM and DGM for Example 1.

FM \ DGM	Contextual	Between-within
Contextual	6148.551	6098.649
Between-within	6148.553	6098.686

7.1 Tables for Study 1

Table 3: $n_j = 100$, $N \in \{10, 25, 50, 75, 100\}$, niter=1000

DGM	Proportion	$N = 10$	$N = 25$	$N = 50$	$N = 75$	$N = 100$
BW	P(D$_{BW}$ <D$_C$)	$\mathbf{0.598}$ $\mathbf{[0.568, 0.628]}$	$\mathbf{0.558}$ $\mathbf{[0.527, 0.589]}$	$\mathbf{0.605}$ $\mathbf{[0.575, 0.635]}$	$\mathbf{0.584}$ $\mathbf{[0.553, 0.615]}$	$\mathbf{0.585}$ $\mathbf{[0.554, 0.616]}$
	P(D$_{BW}$ $>$ D$_C$)	0.385 [0.355,0.415]	0.433 [0.402,0.464]	0.395 [0.365,0.425]	0.415 [0.384,0.446]	0.414 [0.383, 0.445]
	P(D$_{BW}$ $==$ D$_C$)	0.017 [0.009, 0.025]	0.009 [0.003,0.015]	0 [0,0]	0.001 [-0.001,0.003]	0.001 [-0.001, 0.003]
C	D$_{BW}$ $<$ D$_C$	$\mathbf{0.588}$ $\mathbf{[0.557, 0.619]}$	$\mathbf{0.566}$ $\mathbf{[0.535, 0.597]}$	$\mathbf{0.543}$ $\mathbf{[0.512, 0.574]}$	$\mathbf{0.542}$ $\mathbf{[0.511, 0.573]}$	$\mathbf{0.548}$ $\mathbf{[0.517, 0.579]}$
	P(D$_{BW}$ $>$ D$_C$)	0.395 [0.365, 0.425]	0.429 [0.398,0.460]	0.453 [0.422,0.484]	0.453 [0.422,0.484]	0.451 [0.420,0.482]
	P(D$_{BW}$ $==$ D$_C$)	0.017 [0.009, 0.025]	0.005 [0.001,0.009]	0.004 [0,0.008]	0.005 [0.001,0.009]	0.001 [-0.001,0.003]

Table 4: $N = 100$, $n_j \in \{10, 25, 50, 75, 100\}$, niter=1000

DGM	Proportion	$n_j = 10$	$n_j = 25$	$n_j = 50$	$n_j = 75$	$n_j = 100$
BW	D$_{BW}$ <D$_C$	$\mathbf{0.629}$ $\mathbf{[0.599, 0.659]}$	$\mathbf{0.623}$ $\mathbf{[0.593, 0.653]}$	$\mathbf{0.551}$ $\mathbf{[0.520, 0.582]}$	$\mathbf{0.585}$ $\mathbf{[0.554, 0.616]}$	$\mathbf{0.585}$ $\mathbf{[0.554, 0.616]}$
	D$_{BW}$ $>$ D$_C$	0.368 [0.338, 0.398]	0.373 [0.343,0.403]	0.447 [0.416,0.478]	0.415 [0.384,0.446]	0.414 [0.383,0.445]
	D$_{BW}$ $==$ D$_C$	0.003 [0,0.006]	0.004 [0,0.008]	0.002 [-0.001, 0.005]	0 [0,0]	0.001 [-0.001,0.003]
C	D$_{BW}$ $<$ D$_C$	$\mathbf{0.580}$ $\mathbf{[0.549, 0.611]}$	$\mathbf{0.595}$ $\mathbf{[0.565, 0.625]}$	$\mathbf{0.549}$ $\mathbf{[0.518, 0.580]}$	$\mathbf{0.561}$ $\mathbf{[0.530, 0.592]}$	$\mathbf{0.548}$ $\mathbf{[0.517, 0.579]}$
	D$_{BW}$ $>$ D$_C$	0.417 [0.386,0.448]	0.402 [0.373,0.432]	0.450 [0.419,0.481]	0.438 [0.407,0.469]	0.451 [0.420,0.482]
	D$_{BW}$ $==$ D$_C$	0.003 [0,0.006]	0.003 [0,0.006]	0.001 [-0.001,0.003]	0.001 [-0.001,0.003]	0.001 [-0.001,0.003]

7.2 Tables for Study 2 Part 1

Table 5: $N = 10, n_j = 50$, $\tau_1^2 \in \{1, 5, 10\}$ $\tau_0^2 \in \{0.1, 1\}$, $\tau_{01} = 0$, niter=1000

DGM	Proportion	$\tau_0^2 = 0.1$ $\tau_1^2 = 1$	$\tau_0^2 = 0.1$ $\tau_1^2 = 5$	$\tau_0^2 = 0.1$ $\tau_1^2 = 10$	$\tau_0^2 = 1$ $\tau_1^2 = 1$	$\tau_0^2 = 1$ $\tau_1^2 = 5$	$\tau_0^2 = 1$ $\tau_1^2 = 10$
BW	$D_{BW} < D_C$	**0.654** **[0.625, 0.683]**	**0.761** **[0.735, 0.787]**	**0.770** **[0.744, 0.796]**	**0.581** **[0.550, 0.612]**	**0.710** **0.682, 0.738**	**0.733** **[0.706, 0.760]**
	$D_{BW} > D_C$	0.336 [0.307,0.365]	0.235 [0.209,0.261]	0.227 [0.201,0.253]	0.406 [0.376, 0.436]	0.279 [0.251,0.307]	0.264 [0.237, 0.291]
	$D_{BW} == D_C$	0.010 [0.004, 0.016]	0.004 [0,0.008]	0.003 [0,0.006]	0.013 [0.006,0.020]	0.011 [0.005,0.017]	0.003 [0,0.006]
C	$D_{BW} < D_C$	**0.618** **[0.588, 0.648]**	**0.513** **[0.482, 0.544]**	0.463 [0.432,0.494]	**0.584** **[0.553, 0.615]**	**0.558** **0.527, 0.589**	0.493 [0.462, 0.524]
	$D_{BW} > D_C$	0.375 [0.345,0.405]	0.483 [0.452,0.514]	**0.532** **[0.501, 0.563]**	0.408 [0.378,0.438]	0.428 [0.397,0.459]	**0.498** **[0.467, 0.529]**
	$D_{BW} == D_C$	0.007 [0.002,0.012]	0.004 [0,0.008]	0.005 [0.001,0.009]	0.008 [0.002,0.014]	0.014 [0.007,0.021]	0.009 [0.003, 0.015]

Table 6: $N = 10, n_j = 50$, $\tau_1^2 \in \{1, 5, 10\}$ $\tau_0^2 = 1$, $\tau_{01} \in \{0, 0.5\}$, niter=1000

DGM	Proportion	$\tau_{01} = 0$ $\tau_1^2 = 1$	$\tau_{01} = 0$ $\tau_1^2 = 5$	$\tau_{01} = 0$ $\tau_1^2 = 10$	$\tau_{01} = 0.5$ $\tau_1^2 = 1$	$\tau_{01} = 0.5$ $\tau_1^2 = 5$	$\tau_{01} = 0.5$ $\tau_1^2 = 10$
BW	$D_{BW} < D_C$	**0.581** **[0.625, 0.683]**	**0.710** **[0.735, 0.787]**	**0.733** **[0.744, 0.796]**	**0.595** **[0.565, 0.625]**	**0.688** **[0.659, 0.717]**	**0.729** **[0.701, 0.757]**
	$D_{BW} > D_C$	0.406 [0.307,0.365]	0.279 [0.209,0.261]	0.264 [0.201,0.253]	0.393 [0.363,0.423]	0.305 [0.276,0.334]	0.266 [0.239,0.293]
	$D_{BW} == D_C$	0.013 [0.006,0.020]	0.011 [0.005,0.017]	0.003 [0,0.006]	0.012 [0.005,0.019]	0.007 [0.002,0.012]	0.005 [0.001,0.009]
C	$D_{BW} < D_C$	**0.584** **[0.553, 0.615]**	**0.558** **0.527, 0.589**	0.493 [0.462, 0.524]	**0.587** **[0.556, 0.618]**	**0.558** **[0.537, 0.589]**	**0.499** **[0.468, 0.530]**
	$D_{BW} > D_C$	0.408 [0.378,0.438]	0.428 [0.397,0.459]	**0.498** **[0.467, 0.529]**	0.403 [0.373,0.433]	0.430 [0.399,0.461]	0.493 [0.462,0.524]
	$D_{BW} == D_C$	0.008 [0.002,0.014]	0.014 [0.007,0.021]	0.009 [0.003, 0.015	0.010 [0.004, 0.016]	0.012 [0.005,0.019]	0.008 [0.002,0.014]

Table 7: $\tau_0^2, \tau_1^2 \in \{(0.01, 1), (0.1, 10), (1, 100), (0.1, 1), (1, 10), (10, 100)\}$, $N = 10$, $n_j = 50$, $\tau_{01} = 0$, niter=1000

DGM	Proportion	$\tau_1^2/\tau_0^2 = 10$ $\tau_1^2 = 1$ $\tau_0^2 = 0.01$	$\tau_1^2/\tau_0^2 = 10$ $\tau_1^2 = 10$ $\tau_0^2 = 0.1$	$\tau_1^2/\tau_0^2 = 10$ $\tau_1^2 = 100$ $\tau_0^2 = 1$	$\tau_1^2/\tau_0^2 = 100$ $\tau_1^2 = 1$ $\tau_0^2 = 0.1$	$\tau_1^2/\tau_0^2 = 100$ $\tau_1^2 = 10$ $\tau_0^2 = 1$	$\tau_1^2/\tau_0^2 = 100$ $\tau_1^2 = 100$ $\tau_0^2 = 10$
BW	$D_{BW} < D_C$	**0.649** **[0.619, 0.679]**	**0.770** **0.744, 0.796**	**0.730** **[0.702, 0.758]**	**0.654** **[0.625, 0.683]**	**0.733** **[0.706, 0.760**	**0.754** **[0.727, 0.781]**
	$D_{BW} > D_C$	0.333 [0.304,0.362]	0.227 [0.201, 0.253]	0.266 [0.239,0.293]	0.336 [0.307,0.365]	0.264 [0.237,0.291]	0.240 [0.214,0.266]
	$D_{BW} == D_C$	0.018 [0.010,0.026]	0.003 [0,0.006]	0.004 [0,0.008]	0.010 [0.004,0.016]	0.003 [0,0.006]	0.006 [0.001,0.011]
C	$D_{BW} < D_C$	**0.577** **[0.546, 0.608]**	0.436 [0.432,0.494]	0.380 [0.350,0.410]	**0.618** **[0.588, 0.648]**	0.493 [0.462,0.524]	0.450 [0.419,0.481]
	$D_{BW} > D_C$	0.498 [0.378,0.438]	**0.532** **[0.501, 0.563]**	**0.612** **[0.582, 0.642]**	0.375 [0.345,0.405]	**0.498** **[][0.467, 0.529]**	**0.538** **0.507, 0.569**
	$D_{BW} == D_C$	0.015 [0.007,0.023]	0.005 [0.001,0.009]	0.008 [0.002,0.014]	0.007 [0.002,0.012]	0.009 [0.003,0.015]	0.012 [0.005,0.019]

7.3 Tables for Study 2 Part 2

Table 8: $N = 10, n_j = 50$, $\tau_1^2 \in \{1, 5, 10\}$ $\tau_0^2 \in \{0.1, 1\}$, $\tau_{01} = 0$, niter=1000

		$\tau_0^2 = 0.1$			$\tau_0^2 = 1$		
DGM	Proportion	$\tau_1^2 = 1$	$\tau_1^2 = 5$	$\tau_1^2 = 10$	$\tau_1^2 = 1$	$\tau_1^2 = 5$	$\tau_1^2 = 10$
BW	$D_{BW} < D_C$	**0.863** **[0.842, 0.884]**	**0.887** **[0.867, 0.907]**	**0.886** **[0.866, 0.906]**	**0.823** **[0.799, 0.847]**	**0.870** **[0.849, 0.891]**	**0.881** **[0.861, 0.901]**
	$D_{BW} > D_C$	0.133 [0.112,0.154]	0.110 [0.091,0.129]	0.113 [0.093, 0.133]	0.174 [0.151,0.197]	0.127 [0.106,0.148]	0.115 [0.095, 0.135]
	$D_{BW} == D_C$	0.004 [0,0.008]	0.003 [0,0.006]	0.001 [-0.001,0.003]	0.003 [0,0.006]	0.003 [0,0.006]	0.004 [0,0.008]
C	$D_{BW} < D_C$	**0.511** **[0.480, 0.542]**	0.471 [0.440,0.502]	0.471 [0.440,0.502]	**0.507** **[0.476, 0.538]**	0.485 [0.477,0.539]	0.484 [0.453,0.515]
	$D_{BW} > D_C$	0.479 [0.448,0,510]	**0.523** **[0.440, 0.554]**	**0.519** **[0.488, 0.550]**	0.487 [0.456,0.518]	**0.508** **[0.477, 0.539]**	**0.506** **[0.475, 0.537]**
	$D_{BW} == D_C$	0.010 [0.004,0.016]	0.006 [0.001,0.011]	0.010 [0.004,0.016]	0.006 [0.001,0.011]	0.007 [0.002,0.012]	0.010 [0.004,0.016]

Table 9: $N = 10, n_j = 50$, $\tau_1^2 \in \{1, 5, 10\}$ $\tau_0^2 = 1$, $\tau_{01} \in \{0, 0.5\}$, niter=1000

		$\tau_{01} = 0$			$\tau_{01} = 0.5$		
DGM	Proportion	$\tau_1^2 = 1$	$\tau_1^2 = 5$	$\tau_1^2 = 10$	$\tau_1^2 = 1$	$\tau_1^2 = 5$	$\tau_1^2 = 10$
BW	$D_{BW} < D_C$	**0.864** **[0.843, 0.885]**	**0.886** **[0.866, 0.906]**	**0.892** **[0.873, 0.911]**	**0.834** **[0.811, 0.857]**	**0.8754** **[0.853, 0.895]**	**0.881** **[0.861, 0.901]**
	$D_{BW} > D_C$	0.134 [0.113,0.155]	0.113 [0.093,0.133]	0.106 [0.087,0.125]	0.164 [0.141,0.187]	0.125 [0.105,0.145]	0.117 [0.097,0.137]
	$D_{BW} == D_C$	0.002 [-0.001,0.005]	0.001 [-0.001,0.003]	0.002 [-0.001,0.005]	0.002 [-0.001,0.005]	0.001 [-0.001,0.003]	0.002 [-0.001,0.005]
C	$D_{BW} < D_C$	**0.515** **[0.484, 0.546]**	0.471 [0.440,0.502]	0.483 [0.452,0.514]	0.482 [0.451,0.513]	0.483 [0.452,0.483]	0.494 [0.463,0.525]
	$D_{BW} > D_C$	0.479 [0.448,0.0.550]	**0.519** [**[0.488, 0.550]**]	**0.511** **0.480, 0.542**	**0.510** **[0.479, 0.541]**	**0.514** **[0.483, 0.545]**	**0.497** **[0.466, 0.528]**
	$D_{BW} == D_C$	0.006 [0.001,0.011]	0.010 [0.004,0.016]	0.006 [0.001,0.011]	0.008 [0.002,0.014]	0.003 [0,0.006]	0.009 [0.003,0.015]

Table 10: $\tau_0^2, \tau_1^2 \in \{(0.01, 1), (0.1, 10), (1, 100), (0.1, 1), (1, 10), (10, 100)\}$, $N = 10$, $n_j = 50$, $\tau_{01} = 0$, niter=1000

		$\tau_1^2/\tau_0^2 = 10$			$\tau_1^2/\tau_0^2 = 100$		
DGM	Proportion	$\tau_1^2 = 1$ $\tau_0^2 = 0.01$	$\tau_1^2 = 10$ $\tau_0^2 = 0.1$	$\tau_1^2 = 100$ $\tau_0^2 = 1$	$\tau_1^2 = 1$ $\tau_0^2 = 0.1$	$\tau_1^2 = 10$ $\tau_0^2 = 1$	$\tau_1^2 = 100$ $\tau_0^2 = 10$
BW	$D_{BW} < D_C$	**0.864** **[0.84, 0.884]**	**0.886** **[0.866.0.906]**	**0.892** **[0.873, 0.911]**	**0.863** **[0.842, 0.884]**	**0.881** **[0.861, 0.901]**	**0.891** **[0.872, 0.910]**
	$D_{BW} > D_C$	0.134 [0.112,0.154]	0.113 [0.093,0.133]	0.106 [0.087,0.125]	0.133 [0.112,0.154]	0.115 [0.095,0.135]	0.107 [0.088,0.126]
	$D_{BW} == D_C$	0.002 [0,0.008]	0.001 [-0.001,0.003]	0.002 [-0.001,0.005]	0.004 [0,0.008]	0.004 [0,0.008]	0.001 [-0.001,0.005]
C	$D_{BW} < D_C$	**0.515** **[0.480, 0.542]**	0.471 [0.440,0.502]	0.483 [0.452,0.514]	**0.511** **[0.480, 0.542]**	0.484 [0.453,0.515]	0.470 [0.439,0.501]
	$D_{BW} > D_C$	0.479 [0.448,0.510]	**0.519** **[0.488, 0.550]**	**0.511** **[0.480, 0.542]**	0.479 [0.448,0.510]	**0.506** **[0.475, 0.537]**	**0.513** **[0.482, 0.544]**
	$D_{BW} == D_C$	0.006 [0.001,0.011]	0.010 [0.004,0.016]	0.006 [0.001,0.011]	0.010 [0.004,0.016]	0.010 [0.004,0.016]	0.017 [0.009,0.025]

8. Figures

```r
# Figure 1, the observed difference in deviance
set.seed(2008)
random_ids <- sample(unique(hsb$id), size = 10)
hsbsub <- hsb %>% filter(id %in% random_ids)

between_within <- lmer(mathach ~ ses_cmc + ses_cm + minority_cmc + minority_cm +
             (minority_cmc | id), REML = FALSE, data = hsbsub)
contextual <- lmer(mathach ~ ses_cmc + ses_cm +   minority   + minority_cm +
             (minority | id), REML = FALSE, data = hsbsub)
anova(between_within, contextual)
```

```
## Data: hsbsub
## Models:
## between_within: mathach ~ ses_cmc + ses_cm + minority_cmc + minority_cm + (minority_cmc | id)
## contextual: mathach ~ ses_cmc + ses_cm + minority + minority_cm + (minority | id)
##                npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## between_within    9 3064.5 3102.0 -1523.2   3046.5
## contextual        9 3066.2 3103.8 -1524.1   3048.2     0  0
```

Figure 1: R output illustrating the difference in deviation for the between-within and the contextual models.

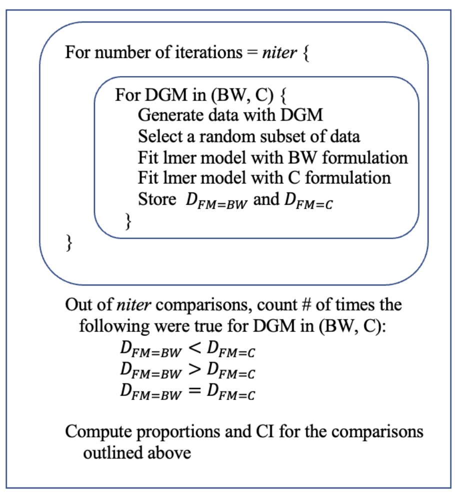

Figure 2: General steps for simulations studies in this project.

```r
# Read in the data
hsb <- read_sav(here("data_files", "hsball.sav"))
hsb <- hsb %>%
  group_by(id) %>%
  mutate(ses_cm = mean(ses),
         minority_cm = mean(minority), ses_cmc = ses - ses_cm,
         minority_cmc = minority - minority_cm) %>% ungroup()
#himinty 1 means > 40% minority,  0 means < 40% minority
dfhm <- data.frame(table(hsb$himinty))
round(dfhm$Freq[1]/(dfhm$Freq[1] + dfhm$Freq[2]),2)
```

```
## [1] 0.72
```

```r
# 0.72 is the proportion of schools with less than 40% minority
# 0.28 is the proportion of schools with more than 40% minority

ses_group_mn_sd <- hsb %>%
  group_by(id) %>%
  summarise(ses_cm = mean(ses), ses_sd = sd(ses))

round(mean(ses_group_mn_sd$ses_cm), 2) # cluster means have mean -0.01
```

```
## [1] -0.01
```

```r
round(sd(ses_group_mn_sd$ses_cm), 2) # cluster means have sd 0.41
```

```
## [1] 0.41
```

```r
round(mean(ses_group_mn_sd$ses_sd), 2) # cluster sds have mean 0.67
```

```
## [1] 0.67
```

```r
round(sd(ses_group_mn_sd$ses_sd), 2) # cluster sds have sd 0.1
```

```
## [1] 0.1
```

Figure 3: R code to obtain parameter estimates.

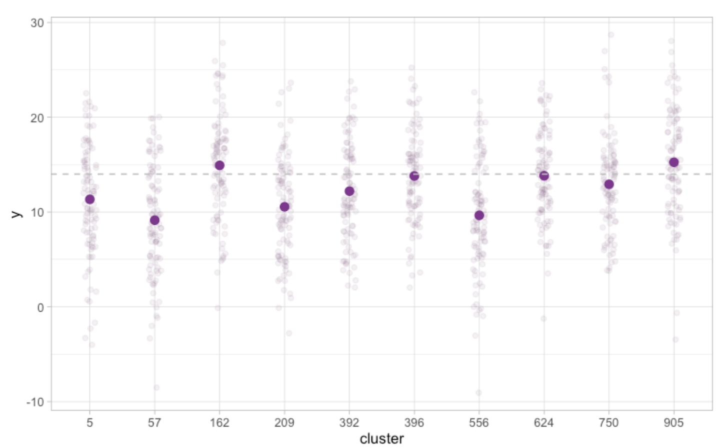

Figure 4: Distribution of y for a random subset of 10 clusters from data generated using the contextual model. The dashed line indicates the grand mean of 14, and circles indicate the cluster means.

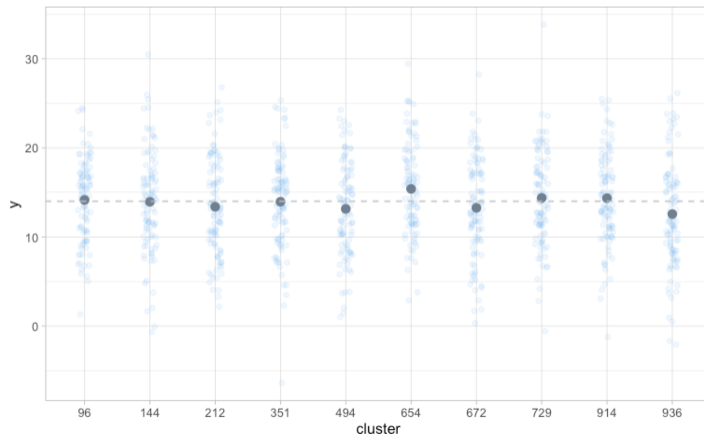

Figure 5: Distribution of y for a random subset of 10 clusters from data generated using the between-within model. The dashed line indicates the grand mean of 14, and circles indicate the cluster means.

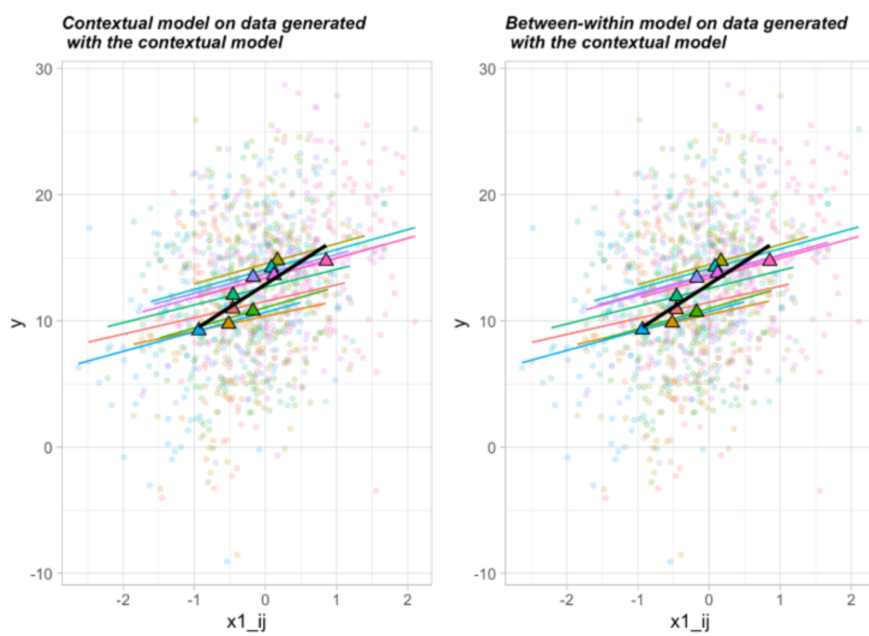

Figure 6: Models fitted to data generated using the contextual model (DGM=C).

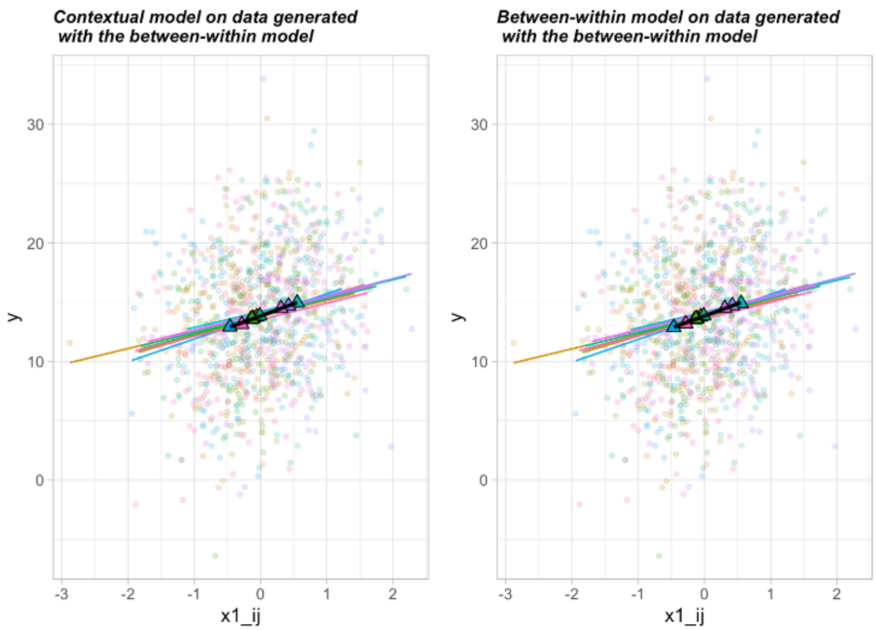

Figure 7: Models fitted to data generated using the between-within model (DGM=BW).