Centering in Multilevel Regression

Centering is the rescaling of predictors by subtracting the mean (or sometimes another constant value). In OLS regression, rescaling using a linear transformation of a predictor (e.g., subtracting one value from every individual score) has no effect on the significance tests and does not alter the standardized slope values.[1] This is not the case with multilevel regression, as fixed or random effect and their significance tests can by altered by linear transformations that rescale the predictors. The effects of centering on multilevel regression are quite complex and deserve more consideration than is possible here (see references below for some good sources), but I would like to make a few general points.

Intercept Interpretation

With multilevel regression, intercepts and intercept variances are of interest and linear transformations impact these values as well as their significance tests. One can see in the formula for the intercept at level 1 (or similarly in OLS regression) that the intercept depends on the value of X at its mean.

$$\beta_{0j} = \overline{Y}_j - \beta_{1j}\overline{X}_j$$

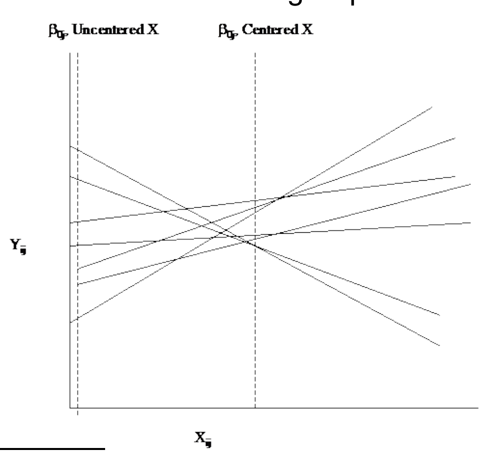
If we recompute the predictor by subtracting the mean from every score, the value of the intercept will change. The intercept is no longer the expected value of Y when X equals the original zero value but, instead, when the value when X is equal to its original mean. Because the γ_{00} is the average of all of the β_{0j} values, its value and interpretation also is affected by centering.

There are two different versions of centering in multilevel regression, *grand-mean centering* and *group-mean centering* (sometimes called "centering within context"). Grand-mean centering subtracts the grand-mean of the predictor using the mean from the full sample ($x_{ij} = X_{ij} - \overline{X}_{..}$). Group-mean centering subtracts the individual's group-mean ($x_{ij} = X_{ij} - \overline{X}_j$) from the individual's score (as would be the case in the intercept equation given above).

Generally, centering makes this value more interpretable, because the expected value of Y when x (centered X) is zero represents the expected value of Y when X is at its mean. In many cases, such as when age or a variable with no meaningful 0 value (e.g., a Likert scale from 1 to 7) is a predictor, the interpretation of the intercept will be unreasonable or undesirable (e.g., value of the outcome when age equals zero or when the Likert scale equals 0) without some type of centering. It thus appears that raw, uncentered predictors should not be the researcher's default scaling.

Effects of Centering on Intercept Variance

Importantly, not only is the average intercept impacted by centering, but the variance of the intercepts is affected by centering as well. The direction of the effect of centering and the degree of effect depends on the pattern of within-group slopes (when they vary). The picture below illustrates that moving the location of the Y-axis by rescaling X, would lead to a different estimate of the variance of the group intercepts in this case. One can imagine then how the covariance among slopes and intercepts will be affected too.



[1] Depending on the particular linear transformation (e.g., adding vs. multiplying by a constant), interpretation of the metric of the unstandardized slope may differ, but in an understandable way.

Centering as a Default

The effects of the scaling of the predictor on the variance of the intercept is a serious concern in multilevel models because the choice of scaling has the potential to substantially change important aspects of the results and the conclusions. This circumstance differs completely from the OLS model. Because the variability of the intercepts is a central component of multilevel model, the fixed and random coefficients can be affected by the variability. If we look back at the level-1 equation and the level-2 equation with β_{0j} as the outcome, it is clear that if the variability of β_{0j} changes then the values of the coefficients will change.

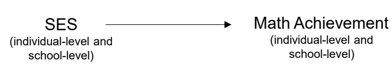
$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + R_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + U_{0j}$$

Thus, it is essential that the interpretation of the group intercepts make sense. For many or even most cases the 0 value for X or Z are not a desired value to use for defining the intercept, and it is therefore reasonable to think about using centered predictors as a default rather than uncentered predictors.

Centering also can function as a way to separate micro and macro level phenomena. Quite often a goal of multilevel modeling is to investigate micro- and macro-level processes. Examining the relationship between group-level predictors and outcomes (e.g., at the school level, neighborhood level, hospital level, companies) may miss phenomena at the individual level. Similarly examining the relationship between individual-level predictors and outcomes (e.g., students, residents, patients, employees) may miss phenomena at the group level. Attempting to extrapolate from one level to another can lead to an ecological (Robinson) or individualistic fallacy (Robinson, 1950; see Subramanian et al., 2009, for an interesting multilevel update).

Level-1 Predictors also Vary Across Groups

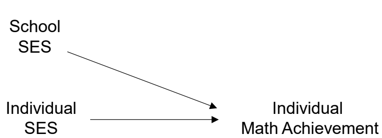
An important point to realize is that any individual-level predictor, such as student SES, may account for the individual variation in the outcome, such as student differences in math achievement, as well as account for some group-level variation in the outcome, such as school differences in math achievement. Including just individual student SES in the model will account for part of why schools vary in math achievement to the extent that schools also vary in SES.



Subtracting out the sample SES mean (grand-mean centering) from every individual student's SES does not change this fact. It simply adjusts the intercept value. However, subtracting the school average SES values from each individual student's SES (group-mean centering) will reduce the variability in math achievement across schools that is due to between-school variation in SES (because we have taken out the school-level mean of SES from the variable). So, the effect of the group-centered predictor can be thought of as reflecting more within-group effects (Hoffman & Walters, 2022).

Separating Individual-level and Group-level Effects: The Special Case of Reintroduced Means

One special case of a level-2 predictor is a variable that has been computed by averaging the responses for all cases in each group, thereby creating a group-level variable with one value for each group. An example might be calculating the average of the individual student SES values for each school and then using these averages as a level-2 predictor in a multilevel regression to represent differences in school SES. When the group-mean, level-2 variable (school average SES) and its corresponding level-1 predictor variable, such as individual student SES, that has been group-mean centered, are used in the model together, it is sometimes referred to as "reintroducing the mean" of the predictor (Kreft et al., 1995), because the group-mean is removed from the group-mean centered variable but added back in with the use of the group-mean as a level-2 predictor.



Typically the motivation for this type of model is to investigate separate within-group and between-group effects of the predictor. For example, what is the impact of individual-level SES on math achievement as compared with the effect of school-level SES? In this special type of model, when the level-1 and level-2 counterparts are included as predictors, different centering approaches provide certain interpretations of the coefficients (Raudenbush & Bryk, 2002, Table 5.11, p. 140, and Snijders & Bosker, 2012, Section 4.6, provide good explanations; refer also to the illustration “Overhead: Reintroduced Means and Compositional Effects” the class page).

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + U_{0j} + U_{1j}X_{ij} + R_{ij}$$

If group-mean centering of the level-1 predictor is used, the level-1 predictor coefficient, γ_{10}, will represent the within-group effect and the level-2 predictor, γ_{01}, will represent the between-group effect. In the case of grand-mean centering, however, γ_{10} and γ_{01} are estimates of the within-group effect and the *compositional effect* (difference between the within and between slopes), respectively. The compositional effect coefficient can be derived from the between- and within-group coefficients in the group-mean centered model. The compositional effect may be of interest in some cases, but I suspect that researchers are more often interested in estimating the independent within-group and between-group effects in order to obtain the independent micro and macro level contributions of a certain predictor. Thus, when the level-1 variable is used as a predictor along and its mean value is also used as a level-2 predictor, then the most likely desired centering method for the level-1 variable will be group-mean centering in order to provide the desired within-group and between-group effects interpretation.

For this within vs. between type of multilevel model, it is also important not to forget about the intercept interpretation that results from the scaling of the mean level-2 variable (e.g., average SES). In most instances, it will be desirable to center that variable as well (grand-mean centering is the only option). Without centering the level-2 variable also, the intercept will be equal to the value of the dependent variable when the predictor equals 0. For most variables, as in the case of age or a 1-to-7 Likert scale predictor, using the group-mean variable (average age or average Likert value) uncentered as a level-2 predictor will be less desirable than using it as grand-mean centered. Centering the level-2 variable does not change the within- and between-group interpretations just described.

Recommendations

The primary decisions about centering have to do with the scaling of level-1 variables. Because there is only one score per group. There is only one choice for centering of level-2 variables—grand-mean centering. Thus, the decision is simple for level-2 variables. In most cases, researchers would likely choose to grand-mean center level-2 variables to improve the interpretation of the intercept values. Of course, one should not blindly follow this recommendation, but there will be far more instances where centering the predictor makes more sense than not centering the predictor. Interpreting the intercept as the value when the level-2 predictor is equal to zero may be desirable in some cases, but I venture to guess not in most cases. It should also be said that there reason values other than the mean cannot be used to derive a different interpretation of the intercept (Hoffman & Walters, 2022).

Choices made regarding centering level-1 variables are much more difficult. Generally, in most if not nearly all circumstances, intercept interpretation will be more reasonable using some type of centered predictors as compared with using uncentered predictors.

The consequences of choosing grand- or group-mean centering are almost overwhelming and it is difficult for me to make global recommendations. I do urge a careful reading of the most thorough considerations of this topic (Algina & Swaminathan, 2011; Enders & Tofighi, 2007), but I provide a few thoughts and a summary of some of the general conclusions from these sources in the table below. At this point in the course we are considering group-nested designs only, but we will revisit the issue in the context of longitudinal (growth curve) models. Wang and Maxwell (2015) and Hoffman (2015, chapters 2, 7, & 8) are sources that discusses centering specifically in the longitudinal case.

Raw uncentered variables	Grand-mean centering	Group-mean centering
• Rarely makes sense unless there is a desire to estimate intercept and intercept variance when the predictor is equal to zero (e.g., mean or variance of group-means for females only). • Does not make sense when X = 0 is an unreasonable or undesirable value (e.g., age = 0) • Entails assumption that group-means are uncorrelated with the predictor (Algina & Swaminathan, 2011).	• For level-2 variables controlling for level-1 variables. • If adjusted group-mean is desired interpretation of average intercept, and variance of adjusted group-means is desired interpretation of intercept variance. • Interactions between level-2 variables. (Interaction estimate and test are unaffected, but lower order terms are). • Entails assumption that group-means are uncorrelated with predictor. • Inclusion of level-2 variables in a model without level-1 reintroduced mean variables will not fully control for the group-level effects of individual-level variables.	• Use if pure level-1 effect is desired without considering level-2 variables. Group-mean centered variables will be uncorrelated with level-2 variables and therefore estimates of the effect of the level-1 variable will not partial out level-2 variables. • When accurate slope variance estimates are desired • When relative between and within effects of the same construct at level 1 and level 2 is desired (e.g., SES and average class SES). • If effects of level-2 variables only of interest without regard to partialling level-1 vairability out. • When cross-level interactions are of interest and interpretation of "main effect" is of interest. • Inclusion of level-2 variables in a model without level-1 reintroduced mean variables will not fully control for the covariate.

Cross-level interactions. Cross-level interaction coefficients are relatively unaffected by centering decisions, although Algina and Swaminathan point out that this assumes no other covariates have confounding interactions with the independent variables. Lower order terms ("main effects" of the predictors involved in the interaction) are affected by centering, however. Bauer and Curran (2005) recommend group-mean centering level-1 predictors (and grand-mean centering level-2) to improve computation and interpretation of the "main effects" when cross-level interactions are tested.

Centering dichotomous predictors? It may seem odd to center a dichotomous predictor like an intervention variable, but if original coding of 0,1 is used, then the intercept and variance of the intercept represents the mean of the 0 group and the variance of the 0 group (e.g., mean and variance of the control group). There is nothing incorrect about this, but it may not be desirable to simply estimate the variance of the intercepts for the 0 group in many cases. It makes sense then to consider centering a binary variable, so that the mean represents the average of the two groups. Note that coding a binary predictor as 1,2 would rarely, if ever, make sense because the intercepts are interpreted as the 0 values of the predictor, which would be a group that does not exist. Deciding whether to group-mean or grand-mean center a binary level-1 predictor is complicated, however. Group-mean centering will produce intercepts weighted by the proportion of 1 to 0 values for each group, whereas grand-mean centering will produce intercepts weighted by the proportion of 1 to 0 in the entire sample. The grand-mean centering is analogous to using a sample weight adjustment to make the sample mean (here, each group's mean) be proportionate to the population mean (here, the full sample).

General comments. Most of the above conclusions are based on fairly simple models and the structure of the model, such as whether both level-1 and level-2 predictors are included and whether there are cross-level interactions, can make a difference on the consequences of centering choices. There are a number of other complexities that have not been thoroughly considered in the literature, such as the consequences of mixing different centering approaches and the impact of large variability of group sample sizes.

References

Algina, J., & Swaminathan, H. (2011). Centering in two-level nested designs. In. J. Hox, & K. Roberts (Eds.), *The Handbook of Advanced Multilevel Data Analysis* (pp 285-312).New York: Routledge.
Bauer, D.J., & Curran, P.J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40*, 373-400.
Enders, C.K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.
Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change.* Routledge.
Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology, 73*, 659-689.
Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-357.
Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting Robinson: the perils of individualistic and ecologic fallacy. *International Journal of Epidemiology, 38*(2), 342-360.
Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. Psychological methods, 20(1), 63.

Centering Example Using HSB Data
***Without* Reintroduced Means**

To provide some illustration of the impact of centering on coefficients and significance, I have tested two different models using the HSB data set. One set of models (p. 4) includes a Level-1 predictor, TSES, a transformed version of the SES variable, and a Level-2 variable, SECTOR, the variable for type of school (0=public, 1=catholic). The transformed variable TSES was used because the original SES variable was standardized with a mean of 0, which interferes with the ability to compare the effects of centering choices. Although results presented here are from the HLM package, the consequences of centering will not be different using SPSS, R, or other packages. The second set of models (p. 5) examines the effects of reintroducing the school mean of the socioeconomic variable, MEANTSES, into the model at Level-2. Centering effects are complex and the pattern with other models may differ. Centering was only used for this level-1 predictor to simplify the illustration.

Separate Equations (centering options for TSES are varied)

$$MATHACH = \beta_{0j} + \beta_{1j}(TSES) + R_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR) + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(SECTOR) + U_{1j}$$

Single Equation

$$MATHACH = \gamma_{00} + \gamma_{01}(SECTOR) + \gamma_{10}(TSES) + \gamma_{11}(TSES * SECTOR) + U_{0j} + U_{1j}(TSES) + R_{ij}$$

	Description	Uncentered	Grand-mean Centered	Group-mean Centered
γ_{00}	Adjusted grand-mean of MATHACH	- 3.044 (.701) , t = -4.340,p<.001	11.751 (.292), t = 50.596, p<.001	11.394 (.293), t = 38.915, p<.001
γ_{01}	Effect of SECTOR	8.694 (1.089), t = 7.982,p<.001	2.128 (.347), t = 6.140, p<.001	2.807 (.439), t = 6.392, p<.001
γ_{10}	Average effect of TSES	.296 (.015), t = 20.341,p<.001	.296 (.015), t = 20.341, p<.001	.280 (.016), t = 17.904, p<.001
γ_{11}	Interaction of SECTOR with TSES	-.131 (.022), t = -5.994,p<.001	-.131 (.022), t = -5.994, p<.001	-.134 (.024), t = -5.680, p <.001
τ_0^2	Variance of adjusted intercepts across schools	1.668, χ^2 = 164.874, p = .338	3.833, χ^2 = 756.043, p < .001	6.740, χ^2 = 1383.785, p < .001
τ_1^2	Variance of TSES slopes across schools	.001, χ^2 = 178.091, p = .131	.001, χ^2 = 178.091, p = .131	.003, χ^2 = 175.312, p = .164
σ^2	Variance within schools	36.763	36.763	36.690

Note: TSES has a mean of 50 and a standard deviation of 7.79; it is a transformed version of the SES variable found in the original Raudenbush and Bryk (2002) HSB data set using the T-score test scoring formula to better illustrate uncentered variables (i.e., so that the mean would not equal zero for uncentered scores). Level-2 predictors are entered as uncentered variables for all models to keep things slightly simpler (not usually recommended in practice!). Standard REML estimates (not using robust standard errors) are presented here.

```
VARIABLE NAME    N    MEAN     SD     MINIMUM   MAXIMUM
 MEANTSES     160   49.94    4.14    38.06     58.25
 TSES       7185   50.00    7.79    12.42     76.92
```

Centering Example Using HSB Data
With *Reintroduced Means*

Separate Equations (centering options for TSES are varied)

$$MATHACH = \beta_{0j} + \beta_{1j}(TSES) + R_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(MEANTSES) + \gamma_{02}(SECTOR) + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(MEANTSES) + \gamma_{12}(SECTOR) + U_{1j}$$

Single Equation

$$MATHACH = \gamma_{00} + \gamma_{01}(MEANTSES) + \gamma_{02}(SECTOR) + \gamma_{10}(TSES) + \gamma_{11}(TSES * MEANTSES) + \gamma_{12}(TSES * SECTOR) + U_{0j} + U_{1j}(TSES) + R_{ij}$$

	Description	**Uncentered**	**Grand-mean Centered**	**Group-mean Centered**
γ_{00}	Adjusted grand-mean of MATHACH	1.983 (6.562), $t = .302$, p=.763	-4.518 (1.911) , $t = -2.364$, p<.05	-14.537 (1.805) , $t = -8.053$, p<.001
γ_{01}	Effect of MEANTSES	-.088 (.136) , $t = -.648$, p=.518	.333 (.039) , $t = 8.546$, p<.001	.533 (.037) , $t = 14.446$, p<.001
γ_{02}	Effect of SECTOR	9.074 (1.150) , $t = 7.888$, p<.001	1.193 (.308) , $t = 3.870$, p<.001	1.227 (.306) , $t = 4.005$, p<.001
γ_{10}	Average effect of TSES	-.130 (.134) , $t = -.969$, p=.334	-.130 (.134) , $t = -.969$, p=.334	-.223 (.147) , $t = -1.512$, p=.132
γ_{11}	Interaction of MEANTSES with TSES	.008 (.002) , $t = 3.058$, p<.01	.008(.003) , $t = 3.058$, p<.01	.010 (.003) , $t = 3.420$, p<.01
γ_{12}	Interaction of SECTOR with TSES	-.158 (.023) , $t = -6.929$, p<.001	-.158 (.023) , $t = -6.929$, p<.001	-.164 (.024) , $t = -6.756$, p<.001
τ_0^2	Variance of adjusted intercepts across schools	1.922, $\chi^2 = 160.950$, p=.398	2.411, $\chi^2 = 573.179$, p < .001	2.380, $\chi^2 = 605.306$, p < .001
τ_1^2	Variance of TSES slopes across schools	.001, $\chi^2 = 162.623$, p = .363	.001, $\chi^2 = 162.623$, p = .362	.001, $\chi^2 = 162.302$, p = .369
σ^2	Variance within schools	36.740	36.740	36.703

Note: TSES has a mean of 50 and a standard deviation of 7.79; it is a transformed version of the SES variable found in the original Raudenbush and Bryk (2002) HSB data set using the T-score test scoring formula to better illustrate uncentered variables (i.e., so that the mean would not equal zero for uncentered scores). Level-2 predictors are entered as uncentered variables for all models to keep things slightly simpler (not usually recommended in practice!). Standard REML estimates (not using robust standard errors) are presented here.