Centering in Linear Multilevel Models

Jan de Leeuw

Volume 1, pp. 247–249

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Centering in Linear Multilevel Models

Consider the situation in which we have m groups of individuals, where group j has n_j members. We consider a general multilevel model, (*see* **Linear Multilevel Models**) that is, a random coefficient model for each group of the form

$$y_{ij} = \beta_{0j} + \sum_{s=1}^{p} x_{ijs}\beta_{sj} + \epsilon_{ij}, \tag{1}$$

where the coefficients are the outcomes of a second regression model

$$\beta_{sj} = \gamma_{0s} + \sum_{r=1}^{q} z_{jr}\gamma_{rs} + \delta_{sj}. \tag{2}$$

Both the $n_j \times (p+1)$ matrices X_j of first-level predictors and the $p \times (q+1)$ matrix Z of second-level predictors have a leading column with all elements equal to $+1$, corresponding with the intercepts of the regression equations. To single out the intercepts in our formulas, our indices for both first- and second-level predictors start at zero. Thus, $0 \le s \le p$ and $0 \le r \le q$ and $x_{ij0} = z_{j0} = 1$ for all i and j. Observe that the predictors X_j and Z are both nonrandom, either because they are fixed by design or because we condition our model on their observed values.

The disturbance vectors ϵ_j and δ_j have zero expectations. They are uncorrelated with each other and have covariance matrices $V(\epsilon_j) = \sigma^2 I$, where I is the $n_j \times n_j$ identity matrix, and where $V(\delta_j) = \Omega$. The $p \times p$ matrix Ω has elements ω_{st}. It follows that the expectations are

$$E(y_{ij}) = \gamma_{00} + \sum_{r=1}^{q} z_{jr}\gamma_{r0} + \sum_{s=1}^{p} x_{ijs}\gamma_{0s} + \sum_{s=1}^{p}\sum_{r=1}^{q} x_{ijs}z_{jr}\gamma_{rs}, \tag{3}$$

and the covariances are

$$C(y_{ij}, y_{kj}) = \omega_{00} + \sum_{s=1}^{p}(x_{ijs} + x_{kjs})\omega_{0s} + \sum_{s=1}^{p}\sum_{t=1}^{p} x_{ijs}x_{kjt}\omega_{st} + \delta^{ik}\sigma^2. \tag{4}$$

Here δ^{ik} is Kronecker's delta, that is, it is equal to one if $i = k$ and equal to zero otherwise. Typically, we define more restrictive models for the same data by requiring that some of the regression coefficients γ_{rs} and some of the variance and covariance components ω_{st} are zero.

In multilevel analysis, the scaling and centering of the predictors is often arbitrary. Also, there are sometimes theoretical reasons to choose a particular form of centering. See Raudenbush and Bryk 2, p. 31–34 or Kreft et al. [1]. In this entry, we consider the effect on the model of translations. Suppose we replace x_{is} by $\tilde{x}_{ijs} = x_{ijs} - a_s$. Thus, we subtract a constant from each first-level predictor, and we use the same constant for all groups. If the a_s are the predictor means, this means *grand mean centering* of all predictor variables. Using grand mean centering has some familiar interpretational advantages. It allows us to interpret the intercept, for instance, as the expected value if all predictors are equal to their mean value. If we do not center, the intercept is the expected value if all predictors are zero, and for many predictors used in behavioral science zero is an arbitrary or impossible value (think of a zero IQ, a zero income, or a person of zero height).

After some algebra, we see that

$$\gamma_{00} + \sum_{r=1}^{q} z_{jr}\gamma_{r0} + \sum_{s=1}^{p} x_{ijs}\gamma_{0s} + \sum_{s=1}^{p}\sum_{r=1}^{q} x_{ijs}z_{jr}\gamma_{rs} = \tilde{\gamma}_{00} + \sum_{r=1}^{q} z_{jr}\tilde{\gamma}_{r0} + \sum_{s=1}^{p} \tilde{x}_{ijs}\gamma_{0s} + \sum_{s=1}^{p}\sum_{r=1}^{q} \tilde{x}_{ijs}z_{jr}\gamma_{rs}, \tag{5}$$

with

$$\tilde{\gamma}_{r0} = \gamma_{r0} + \sum_{s=1}^{p} \gamma_{rs}a_s \tag{6}$$

for all $0 \le r \le q$. Thus, the translation of the predictor can be compensated by a linear transformation of the regression coefficients, and any vector of expected values generated by the untranslated model can also be generated by the translated model. This is a useful type of invariance. But it is important to observe that if we restrict our untranslated model, for instance, by requiring one or more γ_{r0} to be zero, then those same γ_{r0} will no longer be zero in the corresponding translated model. We have invariance of the expected values under translation if the regression coefficients of the group-level predictors are nonzero.

In the same way, we can see that

$$\omega_{00} + \sum_{s=1}^{p}(x_{ijs} + x_{kjs})\omega_{0s} + \sum_{s=1}^{p}\sum_{t=1}^{p} x_{ijs}x_{kjt}\omega_{st} = \tilde{\omega}_{00} + \sum_{s=1}^{p}(\tilde{x}_{ijs} + \tilde{x}_{kjs})\tilde{\omega}_{0s} + \sum_{s=1}^{p}\sum_{t=1}^{p} \tilde{x}_{ijs}\tilde{x}_{kjt}\tilde{\omega}_{st} \tag{7}$$

if

$$\tilde{\omega}_{00} = \omega_{00} + 2\sum_{s=1}^{p}\omega_{0s}a_s + \sum_{s=1}^{p}\sum_{t=1}^{p}\omega_{st}a_s a_t,$$
$$\tilde{\omega}_{0s} = \omega_{0s} + \sum_{t=1}^{p}\omega_{st}a_t. \tag{8}$$

Thus, we have invariance under translation of the variance and covariance components as well, but, again, only if we do not require the ω_{0s}, that is, the covariances of the slopes and the intercepts, to be zero. If we center by using the grand mean of the predictors, we still fit the same model, at least in the case in which we do not restrict the γ_{r0} or the ω_{s0} to be zero.

If we translate by $\tilde{x}_{ijs} = x_{ijs} - a_{js}$ and thus subtract a different constant for each group, the situation becomes more complicated. If the a_{js} are the group means of the predictors, this is *within-group centering*. The relevant formulas are derived in [1], and we will not repeat them here. The conclusion is that separate translations for each group cannot be compensated for by adjusting the regression coefficients and the variance components. In this case, there is no invariance, and we are fitting a truly different model. In other words, choosing between a translated and a nontranslated model becomes a matter of either theoretical or statistical (goodness-of-fit) considerations.

From the theoretical point of view, consider the difference in meaning of a grand mean centered and a within-group mean centered version of a predictor such as grade point average. If two students have the same grade point average (GPA), they will also have the same grand mean centered GPA. But GPA in deviations from the school mean defines a different variable, in which students with high GPAs in good schools have the same corrected GPAs as students with low GPAs in bad schools. In the first case, the variable measures GPA; in the second case, it measures how good the student is in comparison to all students in his or her school. The two GPA variables are certainly not monotonic with each other, and if the within-school variation is small, they will be almost uncorrelated.

References

[1] Kreft, G.G., de Leeuw, J. & Aiken, L.S. (1995). The effects of different forms of centering in hierarchical linear models, *Multivariate Behavioral Research* **30**, 1–21.

[2] Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods*, 2nd Edition, Sage Publications, Newbury Park.

JAN DE LEEUW