Scientific Software International

Use of proportion variance explained statistics

While proportion variance explained statistics can be used to evaluate the impact of the addition of predictors to a hierarchical model on the amount of variation in the outcome that is explained by a model, use of these statistics is not always straightforward. When models grow more complex, particularly at the higher levels of the hierarchy, interpretation can cause confusion and estimation anomalies can occur.

For a detailed discussion on the reasons for and solutions to these problems, the user is referred to Raudenbush & Bryk (2002, pp. 149-152). Three model types are identified, these being random-intercept-only models, random-slopes and –intercepts models, and complex models with fixed and random level-1 coefficients and mixed forms of centering. A brief summary of this section of the text is given below. Note that, in general, comments on two-level models apply to three-level models too.

Random-intercept-only models

The only variance components to be considered in these models are the level-1 residual variance σ^2 and the random variation in intercepts over groups at level-2 τ_{00}. While the introduction of a level-1 predictor typically reduces the level-1 residual variance, it may also change the level-2 variance. as the meaning of the intercept β_{0j} may change when additional level-1 predictors are included. The level-2 variance may also change and will not necessarily be reduced: it can be smaller or larger than the level-2 variance obtained under an unconditional model.

They key point made by the authors is that the variance explained in a level-2 parameter such as β_{0j} is conditional on a fixed level-1 specification, and that the variance reduction statistics are only interpretable for models with the same level-1 model. As such, it is recommended that a user develop the level-1 model first, before adding level-2 predictors to the model. If this procedure is followed, no anomalies should arise in the computation of these statistics when level-2 predictors are added to the β_{0j} equation. It is noted that, if a truly nonsignificant variable enters the model, it is mathematically possible under maximum likelihood to observe a slight increase in the residual variance σ^2.

Random-slopes and -intercepts models

When both intercepts and slopes are allowed to vary randomly over higher level units, the level-2 random effects increase in number, and these effects may be correlated. As a result, predictors entered into one level-2 equation may affect variance estimates in another equation. If the introduction of a level-2 predictor in one equation seemingly explains the variance in another level-2 equation, it is an indication that the errors in the random effects from the two equations are correlated and that the predictor should be entered in both equations. It is also noted that in extreme cases, a predictor that appeared to have a significant effect in one level-2 equation may become insignificant once the model misspecification as described above has been rectified.

The authors also suggest that, when it is not possible to enter a common set of predictors in all the level-2 equations, the intercept model β_{0j} should be specified first. If additional predictors are entered in the slope equation(s), these predictors should also be included in the intercept equation.

Complex models with fixed and random level-1 coefficients and mixed forms of centering

The more complex a hierarchical model is, the more likely the occurrence of anomalies in variance-explained statistics is. While reasons for this are not entirely clear, it has been found that most of these problems can be avoided if, for each level-1 predictor (regardless of form of centering used), the mean of that level-1 predictor is included in the intercept equation at level-2. This allows representation for each level-1 predictor of two separate relationships, these being β_w and β_B (within and between). If β_w and β_B are different, the failure to represent both relationships in the model introduces model misspecification.