Scientific Software International

Proportion variance explained: three-level models

This document illustrates the use of estimated variance components to calculate various proportions explaining variation. The following topics are discussed here:

- The total variance potentially to be explained at all levels (Model 1)
- Proportion of variance explained at level-1 after addition of a level-2 predictor (Model 2)
- Proportion of variance between level-3 units in π s (Model 2)
- Proportion of variance explained for random coefficients from level-1 model (Model 3)
- Proportion of variance explained for random effects at level-3 (Model 3)
- Incremental variance explained by additional level-2 predictors
- Proportion of variance explained at level-3 after addition of a level-3 predictor (Model 4)

1. Discussion of models specified

The unconditional three-level model

$$y_{ijk} = \pi_{0jk} + e_{ijk}$$
$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\beta_{00k} = \gamma_{000} + u_{00k}$$

allows us to partition the total variability in the outcome into three distinct components:

- $\sigma^2 = \mathrm{var}\left(e_{ijk}\right)$, representing the variability of level-1 units in outcome
- $\tau_\pi = \mathrm{var}\left(r_{0jk}\right)$, representing the variability of level-2 units in outcome
- $\tau_\beta = \mathrm{var}\left(u_{00k}\right)$, which is the variance component associated with variability on outcome over level-3 units.

Apart from providing a useful baseline for purposes of comparison with subsequent, more complex models, we can also use these variance components to estimate the proportion of variance in outcome at each of the levels (see Raudenbush & Bryk, pp. 230).

$$\textit{The proportion of variance over level} - 1 \textit{ units} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\tau}_\pi + \hat{\tau}_\beta}$$

$$The\ proportion\ of\ variance\ over\ level-3\ units = \frac{\hat{\tau}_{\beta}}{\hat{\sigma}^2 + \hat{\tau}_{\pi} + \hat{\tau}_{\beta}}$$

$$The\ proportion\ of\ variance\ over\ level-3\ units = \frac{\hat{\tau}_{\pi}}{\hat{\sigma}^2 + \hat{\tau}_{\pi} + \hat{\tau}_{\beta}}$$

For the EG data supplied with the program four models were fitted, these being

Model 1: a fully unconditional model (with student math achievement as outcome y_{ijk}).

$$\begin{aligned} y_{ijk} &= \pi_{0jk} + e_{ijk} \\ \pi_{0jk} &= \beta_{00k} + r_{0jk} \\ \beta_{00k} &= \gamma_{000} + u_{00k} \end{aligned}$$

Model 2: a model in which the level-1 predictor YEAR was added and allowed to vary randomly over level 2 and 3:

$$\begin{aligned} y_{ijk} &= \pi_{0jk} + \pi_{1jk}\left(YEAR_{ijk}\right) + e_{ijk} \\ \pi_{0jk} &= \beta_{00k} + r_{0jk} \\ \pi_{1jk} &= \beta_{10k} + r_{1jk} \\ \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{10k} &= \gamma_{100} + u_{10k} \end{aligned}$$

Model 3: for this model, the level-2 model for model (2) was extended to

$$\begin{aligned} y_{ijk} &= \pi_{0jk} + \pi_{1jk}\left(YEAR_{ijk}\right) + e_{ijk} \\ \pi_{0jk} &= \beta_{00k} + \beta_{01k}\left(BLACK\right)_{jk} + \beta_{02k}\left(HISPANIC\right)_{jk} + r_{0jk} \\ \pi_{1jk} &= \beta_{10k} + \beta_{11k}\left(BLACK\right)_{jk} + \beta_{12k}\left(HISPANIC\right)_{jk} + r_{1jk} \\ \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{01k} &= \gamma_{010} \\ \beta_{02k} &= \gamma_{020} \\ \beta_{10k} &= \gamma_{100} + u_{00k} \\ \beta_{11k} &= \gamma_{110} \\ \beta_{12k} &= \gamma_{120} \end{aligned}$$

Model 4: In the final model, model 3 were extended to include the level-3 predictor LOWINC.

$$y_{ijk} = \pi_{0jk} + \pi_{1jk}\left(YEAR_{ijk}\right) + e_{ijk}$$
$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\left(BLACK\right)_{jk} + \beta_{02k}\left(HISPANIC\right)_{jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + \beta_{11k}\left(BLACK\right)_{jk} + \beta_{12k}\left(HISPANIC\right)_{jk} + r_{1jk}$$
$$\beta_{00k} = \gamma_{000} + \gamma_{001}\left(LOWINC\right)_{k} + u_{00k}$$
$$\beta_{01k} = \gamma_{010}$$
$$\beta_{02k} = \gamma_{020}$$
$$\beta_{10k} = \gamma_{100} + \gamma_{101}\left(LOWINC\right)_{k} + u_{00k}$$
$$\beta_{11k} = \gamma_{110}$$
$$\beta_{12k} = \gamma_{120}$$

Results of these models are summarized in table 1.

Table1: Results for four HLM3 models based on the EG data

Summary of the Four HLM3 Models				
	Model 1 (unconditional)	**Model 2 (YEAR added)**	**Model 3 (Level-2 variables added)**	**Model 4 (final)**
Level-1:				
$\hat{\sigma}^2$	1.52393	0.30148	0.30155	0.30162
Level-2:				
$\hat{\tau}_{\pi_{00}}$	0.57038	0.64049	0.62257	0.62231
$\hat{\tau}_{\pi_{01}}$		0.04676	0.04661	0.04657
$\hat{\tau}_{\pi_{11}}$		0.01122	0.01107	0.01106
Level-3:				
$\hat{\tau}_{\beta_{00}}$	0.31767	0.16531	0.11379	0.07808
$\hat{\tau}_{\beta_{01}}$		0.01705	0.00764	0.00082
$\hat{\tau}_{\beta_{11}}$		0.01102	0.00941	0.00798
Deviance	25305.9808	16326.2313	16260.2269	16239.21
Parameters	4	9	13	15

While the set of models presented here have been built according to the well-known guidelines of building a model in stages (see Raudenbush & Bryk, Chapter 5), it should be noted that it becomes progressively more difficult to compute the various reduction-in-variance statistics for more complex models. When random-slope(s)-and-intercept models or models with both fixed and random effects and mixed forms of centering are considered, anomalies in these statistics may occur. One of the reasons for this is the presence of correlated intercept and slopes at, say, level-2; another, the introduction of a non-significant predictor; finally, anomalies may be indicative of model misspecification.

2. Total variance potentially to be explained at all levels (Model 1)

Using equations and the results for the unconditional model given above, we find that the

$$\text{The proportion of variance over level}-1\,\text{units} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\tau}_\pi + \hat{\tau}_\beta} = \frac{1.52393}{2.41198} = 0.6312$$

Similarly, the

$$\text{The proportion of variance over level}-2\,\text{units} = \frac{\hat{\tau}_\pi}{\hat{\sigma}^2 + \hat{\tau}_\pi + \hat{\tau}_\beta} = \frac{0.57038}{2.41198} = 0.2365$$

and the

$$\text{The proportion of variance over level}-3\text{units} = \frac{\hat{\tau}_\beta}{\hat{\sigma}^2 + \hat{\tau}_\pi + \hat{\tau}_\beta} = \frac{0.31767}{2.41198} = 0.1317$$

We conclude that 63% of the variability in math achievement scores is over measurements within students (i.e., over level-1 units), while only 13% of the variability in outcome is at school level (level-3).

For model 2, the level-1 model has been extended to include the predictor YEAR. In this random intercepts-and-slopes model, both intercept and YEAR slope are allowed to vary randomly at level-2 and 3.

As a consequence, γ_{00} is now the estimated average outcome when YEAR assumes a value of 0, that is, at the beginning of the observation/testing process. In contrast to model 1, τ_π is now a 2 x 2 matrix of form

$$\tau_\pi = \begin{bmatrix} \tau_{\pi_{00}} & \tau_{\pi_{01}} \\ \tau_{\pi_{10}} & \tau_{\pi_{11}} \end{bmatrix}$$

where $\tau_{\pi_{00}}$ and $\tau_{\pi_{11}}$ are the variance components associated with the random intercept and random slope at level-2 respectively. Analogous to this, we also have

$$\tau_\beta = \begin{bmatrix} \tau_{\beta_{00}} & \tau_{\beta_{01}} \\ \tau_{\beta_{10}} & \tau_{\beta_{11}} \end{bmatrix}$$

When the model is written in a mixed formulation

$$y_{ijk} = \gamma_{00k} + \gamma_{100}\left(YEAR_{ijk}\right) + e_{ijk} + \left[r_{0jk} + r_{1jk}\left(YEAR_{ijk}\right)\right] + \left[u_{00k} + u_{10k}\left(YEAR_{ijk}\right)\right]$$

we can conclude that the total variation in outcome for model 2 is

$$\operatorname{var}\left(y_{ijk}\right)=\operatorname{var}\left(e_{ijk}\right)+\operatorname{var}\left[r_{0jk}+r_{1jk}\left(YEAR_{ijk}\right)\right]+\operatorname{var}\left[u_{00k}+u_{10k}\left(YEAR_{ijk}\right)\right]$$

where

$$\begin{aligned}\operatorname{var}\left[r_{0jk}+r_{1jk}\left(YEAR_{ijk}\right)\right]&=\operatorname{var}\left(r_{0jk}\right)+\operatorname{var}\left(r_{1jk}\left(YEAR_{ijk}\right)\right)+2\operatorname{cov}\left(r_{0jk},r_{1jk}\left(YEAR_{ijk}\right)\right)\\&=\tau_{\pi_{00}}+\left(YEAR_{ijk}\right)^2\tau_{\pi_{11}}+2\left(YEAR_{ijk}\right)\tau_{\pi_{01}}\end{aligned}$$

which depends on the value of YEAR.

3. Proportion of variance explained at level-1 after addition of a level-2 predictor (Model 2)

The proportion variance explained at level-1 can be calculated as

$$\begin{aligned}The\ proportion\ of\ variance\ at\ level-1&=\frac{\hat{\sigma}^2\left(unconditional\right)-\hat{\sigma}^2\left(model\ 2\right)}{\hat{\sigma}^2\left(unconditional\right)}\\&=\frac{1.52393-0.30148}{1.52393}\\&=0.8022\end{aligned}$$

As $\hat{\sigma}^2\left(unconditional\right)$ represents the total within-student variance that can be explained by any level-1 model, it is the appropriate base for this ratio. We see that 80.22% of the level-1 variance in math achievement is accounted for by the predictor YEAR, that is, the year of study -3.5.

4. Proportion of variance between level-3 units (Model2)

In model 2, the estimated variance components can be used to compute the percentage of variation that lies between school for both the intercept and YEAR slope.

The percentage variance between schools on intercept is

$$\frac{\hat{\tau}_{\beta_{00}}}{\hat{\tau}_{\beta_{00}}+\hat{\tau}_{\pi_{00}}}\times 100=\frac{0.16531}{0.16531+0.64049}\times 100=20.52\%$$

and for the YEAR slope (learning rate)

$$\frac{\hat{\tau}_{\beta_{11}}}{\hat{\tau}_{\beta_{11}}+\hat{\tau}_{\pi_{11}}}\times 100=\frac{0.01102}{0.01102+0.01122}\times 100=49.55\%$$

.

Most of the variance between schools is in the YEAR slope.

5. Proportion of variance explained for random coefficients from level-1 model (Model 3)

In the third model, the level-1 model given in model 2 was retained, and level-2 model was extended. While model 2 had the level-2 model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$

we now consider the level-2 model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\left(BLACK\right)_{jk} + \beta_{02k}\left(HISPANIC\right)_{jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + \beta_{11k}\left(BLACK\right)_{jk} + \beta_{12k}\left(HISPANIC\right)_{jk} + r_{1jk}$$

The amount of variance explained in $\tau_{\pi_{00k}}$ and $\tau_{\pi_{11k}}$ due to the introduction of the predictors BLACK and HISPANIC can now be investigated.

The proportion of variance explained in $\tau_{\pi_{00}}$ is calculated as

$$\frac{\hat{\tau}_{\pi_{00k}}\left(Model\ 2\right) - \hat{\tau}_{\pi_{00k}}\left(Model\ 3\right)}{\hat{\tau}_{\pi_{00k}}\left(Model\ 2\right)} = \frac{0.64049 - 0.62257}{0.64049} = 0.0280$$

The predictors BLACK and HISPANIC explained approximately 3% of the variance in the coefficient $\tau_{\pi_{00k}}$.

Similarly, the proportion of variance explained in $\tau_{\pi_{11k}}$ is

$$\frac{\hat{\tau}_{\pi_{11k}}\left(Model\ 2\right) - \hat{\tau}_{\pi_{11k}}\left(Model\ 3\right)}{\hat{\tau}_{\pi_{11k}}\left(Model\ 2\right)} = \frac{0.01122 - 0.01107}{0.01122} = 0.0134$$

Only 1% of variance on outcome over YEAR slopes is explained by the predictors BLACK and HISPANIC. We conclude that these predictors are more helpful in explaining differences in intercepts between students than in explaining difference in the YEAR slope.

6. Proportion of variance explained for random effects at level-3 (Model 3)

We can also evaluate the proportion variance explained in the intercept over schools at level-3 due to the introduction of the extended level-1 and level-2 models.

The introduction of the level-1 predictor YEAR (Model 2) led to a reduction in the variation in intercepts from 0.31767 to 0.16531, i.e.

$$\frac{\hat{\tau}_{\beta_{00k}}\left(Model\ 1\right)-\hat{\tau}_{\beta_{00k}}\left(Model\ 2\right)}{\hat{\tau}_{\beta_{11k}}\left(Model\ 1\right)}=\frac{0.31767-0.16531}{0.31767}=0.4796$$

When the level-2 model is extended (see Model 3) a further reduction in the estimated $\hat{\tau}_{\beta_{00k}}$ is noted: from 0.16531 to 0.11379. It can be concluded that the level-2 model explained an additional

$$\frac{\hat{\tau}_{\beta_{00k}}\left(Model\ 2\right)-\hat{\tau}_{\beta_{00k}}\left(Model\ 3\right)}{\hat{\tau}_{\beta_{11k}}\left(Model\ 2\right)}=\frac{0.16531-0.11379}{0.16531}=0.3117$$

31% of the variation in intercepts over schools.

It should be noted that these statistics are not as trustworthy as those discussed in previous sections, as they may be influence by model misspecification and correlations between the random effects.

7. Incremental variance explained by additional level-2 predictors (Model 5)

The impact of the level-2 predictor HISPANIC can be evaluated by comparing the level-2 model

$$\pi_{0jk}=\beta_{00k}+\beta_{01k}\left(BLACK\right)_{jk}+r_{0jk}$$
$$\pi_{1jk}=\beta_{10k}+\beta_{11k}\left(BLACK\right)_{jk}+r_{1jk}$$

to the level-2 model fitted in model 3,

$$\pi_{0jk}=\beta_{00k}+\beta_{01k}\left(BLACK\right)_{jk}+\beta_{02k}\left(HISPANIC\right)_{jk}+r_{0jk}$$
$$\pi_{1jk}=\beta_{10k}+\beta_{11k}\left(BLACK\right)_{jk}+\beta_{12k}\left(HISPANIC\right)_{jk}+r_{1jk}$$

which contains the additional effects $\beta_{02k}\left(HISPANIC\right)_{jk}$ and $\beta_{12k}\left(HISPANIC\right)_{jk}$. If the model described in and is fitted, we obtain the following estimates for τ_{π}.

$$\hat{\tau}_{\pi}=\begin{bmatrix}\hat{\tau}_{\pi_{00}} & \hat{\tau}_{\pi_{01}}\\ \hat{\tau}_{\pi_{10}} & \hat{\tau}_{\pi_{11}}\end{bmatrix}=\begin{bmatrix}0.62834 & 0.04574\\ 0.04574 & 0.01124\end{bmatrix}$$

The proportion of variance explained in $\tau_{\beta_{00k}}$ due to the predictor HISPANIC is

$$\frac{\hat{\tau}_{\beta_{00k}}\left(current\ model\right)-\hat{\tau}_{\beta_{00k}}\left(Model\ 3\right)}{\hat{\tau}_{\beta_{00k}}\left(current\ model\right)}=\frac{0.62834-0.62257}{0.62834}=0.0092$$

and in $\tau_{\beta_{11k}}$

$$\frac{\hat{\tau}_{\beta_{11k}}\left(current\ model\right)-\hat{\tau}_{\beta_{11k}}\left(Model\ 3\right)}{\hat{\tau}_{\beta_{11k}}\left(current\ model\right)}=\frac{0.01124-0.01102}{0.01124}=0.0196$$

The variance HISPANIC thus contributed an additional 1% to explain the variance in $\tau_{\beta_{00k}}$ and $\tau_{\beta_{11k}}$.

8. Proportion of variance explained at level-3 after addition of a level-3 predictor (Model 4)

The final predictor added to the model is a school-level variable, LOWINC. Results for this model is given in the table under the heading Model 4. While variance components at levels 1 and 2 are unaffected by the addition of the variable, there is a reduction in the estimated variation in outcome over the school level intercept and YEAR slope.

For the unconditional level-3 intercept model used in Model 3 equation

$$\beta_{00k}=\gamma_{000}+u_{00k}$$

variance over school intercept in outcome was estimated at 0.11379. The $\tau_{\beta_{00}}$ obtained from model 4 represent the conditional variance over intercepts, after controlling for the effect of LOWINC:

$$\beta_{00k}=\gamma_{000}+\gamma_{001}\left(LOWINC\right)_k+u_{00k}$$

A reduction of

$$\frac{\hat{\tau}_{\beta_{00}}(Model\ 3)-\hat{\tau}_{\beta_{00}}(Model\ 4)}{\hat{\tau}_{\beta_{00}}(Model\ 3)}=\frac{0.11379-0.07808}{0.11379}=0.3138$$

or 31%, in variation of average math scores over schools, can be ascribed to the predictor LOWINC.

Similarly, when the YEAR slope equations at this level for model 3 and 4 are compared, we see that $\hat{\tau}_{\beta_{11}}(Model\ 3)=0.00941$ and $\hat{\tau}_{\beta_{11}}(Model\ 4)=0.00798$. The addition of LOWINC explained an additional

$$\frac{0.00941-0.00798}{0.00798}\times 100\%=15.20\%$$

of the school-level variation in YEAR slope.