

Partitioning variation in multilevel models.

by

Harvey Goldstein, William Browne and Jon Rasbash

Institute of Education, London, UK.

Summary. In multilevel modelling, the residual variation in a response variable is split into component parts that are attributed to various levels. In applied work, much use is made of the percentage of variation that is attributable to the higher-level sources of variation. Such a measure however only makes sense in simple variance components Normal response models where it is often referred to as the ‘intra-unit correlation’. In this paper we describe how similar measures can be found for both more complex random variation in Normal response models and models with discrete responses. In these cases the variance partitions are dependent on predictors associated with the individual observation. We will compare several computational techniques to compute the variance partitions

Keywords: multilevel modelling, variance components models, MQL, PQL, intra-unit correlation, variance partition coefficient.

Address for Correspondence:

Professor Harvey Goldstein, Institute of Education, 20 Bedford Way, London, WC1H 0AL, UK.

Email: h.goldstein@ioe.ac.uk

1. Introduction

Over the past 15 years multilevel modelling (see for example, Goldstein 1995, Bryk and Raudenbush, 1992) has become used by researchers in many application areas in both the social and medical sciences. One of the main motivations of multilevel modelling is that generally there are inherently groups of observations within a data set that come from a common source, for example a school in educational research or a hospital in medical research. Then two observations chosen randomly from within this particular source are generally not independent and it is important to model this dependency.

The technique of multilevel modelling accounts for this dependency by partitioning the total variance in the data, having fitted any covariates, into variation due to these sources or ‘higher level units’ and the level 1 variation that remains. So for example in educational research we may consider exam results for schoolchildren and here we would partition the variation into variation between, and variation within the higher level units (the schools).

For illustration we shall consider an educational example of a dataset on 4059 children from 65 schools in the Inner London education authority (Goldstein et al. 1993). Here the response variable is the total score for the students in their examinations at age 16 with a predictor being the score on a reading test that each child took at age 11, both test scores having been Normalised.

We consider firstly the 2-level variance components model where the single reading test predictor is treated as a fixed effect. The model is then as follows:

$$\begin{aligned}y_{ij} &= \beta_{0ij} + \beta_1 x_{1ij} \\ \beta_{0ij} &= \beta_0 + u_{0j} + e_{0ij} \\ \text{var}(u_{0j}) &= \sigma_{u0}^2 \quad \text{var}(e_{0ij}) = \sigma_{e0}^2\end{aligned}\tag{1}$$

A convenient summary of the ‘importance’ of schools is the proportion of the total variance accounted for, which we will call the ‘variance partition coefficient’ (VPC) given by the formula $\tau = \sigma_{u0}^2 (\sigma_{u0}^2 + \sigma_{e0}^2)^{-1}$ and in the variance components model case

this is also a measure of the residual correlation between the responses from two students in the same school, hence the term ‘intra-unit correlation’ is often found with the use of the symbol ρ . Researchers involved in multilevel modelling often find it useful to quote an estimate of the VPC and discussions of its usefulness and interpretation often occur on the worldwide E-mail discussion list on multilevel models (www.jiscmail.ac.uk/multilevel/). This statistic is also known in the sample survey literature as the ‘intra class correlation’ and is commonly used as a measure of the extent of clustering.

In this paper we shall show how this summary statistic can be modified for more complex models when there is no simple decomposition of the overall variance and also in the case where the response variable is discrete.

Model 1 can be fitted in standard multilevel modelling software, for example MLwiN (Rasbash et al. 2000) and the results obtained give estimates of 0.092 for the between schools variance (σ_{u0}^2) and 0.563 for the residual variation (σ_{e0}^2). In our example $\tau = 0.14$ so that there is some clustering effect and consequently we gain more accurate estimates and confidence intervals by fitting a multilevel model. A 95% confidence interval, using 999 parametric bootstrap replications (Goldstein, 1995) is (0.092, 0.193). We can also find interval estimates from an MCMC output chain (see below).

2. General Normal response models

The model (1) is equivalent to fitting parallel lines to each of the 65 schools in the dataset. While the VPC is useful for such a model with a single source of variation at each level, it is less so for a random coefficient model, for example a random-slopes regression model where we allow the slopes of the 65 school regressions to vary as follows

$$\begin{aligned}
y_{ij} &= \beta_{0ij} + \beta_{1j}x_{1ij} \\
\beta_{0ij} &= \beta_0 + u_{0j} + e_{0ij} \\
\beta_{1j} &= \beta_1 + u_{1j} \\
\text{var}(u_{0j}) &= \sigma_{u0}^2, \quad \text{var}(u_{1j}) = \sigma_{u1}^2, \\
\text{cov}(u_{0j}u_{1j}) &= \sigma_{u01}, \quad \text{var}(e_{0ij}) = \sigma_{e0}^2
\end{aligned} \tag{2}$$

Here $\tau = (\sigma_{u0}^2 + 2\sigma_{u01}x_{1ij} + \sigma_{u1}^2x_{1ij}^2)(\sigma_{u0}^2 + 2\sigma_{u01}x_{1ij} + \sigma_{u1}^2x_{1ij}^2 + \sigma_{e0}^2)^{-1}$.

In this case the VPC is not the same as the intra-unit correlation since, for two children with scores x_{1i_1j}, x_{1i_2j} , the correlation is given by

$$\frac{(\sigma_{u0}^2 + \sigma_{u01}(x_{1i_1j} + x_{1i_2j}) + \sigma_{u1}^2x_{1i_1j}x_{1i_2j})}{\sqrt{(\sigma_{u0}^2 + 2\sigma_{u01}x_{1i_1j} + \sigma_{u1}^2x_{1i_1j}^2 + \sigma_{e0}^2)(\sigma_{u0}^2 + 2\sigma_{u01}x_{1i_2j} + \sigma_{u1}^2x_{1i_2j}^2 + \sigma_{e0}^2)}}$$

More generally the VPC may be a function of several predictor variables if these have random coefficients (at either level). Thus its simplicity as a measure of clustering is diminished, although for any combination of predictor variables in the model the VPC can be calculated. Figure 1 plots the VPC as a function of reading test score. Of particular interest here is the way in which the ‘importance’ of the school attended increases markedly with increasing reading test score above the mean.

(Figure 1 here)

3. Discrete Response models

We shall now consider a multilevel model with a binary response, but our remarks will apply more generally to models for proportions, for different non-identity link functions and also where the response is a count, in fact to any non-linear model. For a (0,1) response, the model that is analogous to (1) is

$$\begin{aligned}
E(y_{ij}) &= \pi_{ij} = g(\beta_0 + \beta_1x_{1ij} + u_{0j}) \\
y_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
u_{0j} &\sim N(0, \sigma_{u0}^2)
\end{aligned} \tag{3}$$

Remembering that the response is just (0, 1), we see that unlike in the Normal case here the level 1 variance depends on the expected value, $\text{var}(y_{ij}) = \pi_{ij}(1 - \pi_{ij})$ as the fixed predictor in the model depends on the reading test score. Therefore as we are considering a function of the predictor variable x_1 , once again a simple VPC is not available. Furthermore, the level 2 variance, $\sigma_{u_0}^2$, is measured on the logistic scale so is not directly comparable to this level 1 variance.

If we still wish to produce a measure, however, albeit dependent on x_1 , the following procedures will provide at least approximate estimates.

3.1 Model linearisation (Method A)

Using a first order Taylor expansion (see for example Goldstein and Rasbash, 1996) we can write (3) in the form

$$y_{ij} = (\beta_0 + \beta_1 x_{1ij})\pi'_{ij} + u_j \pi'_{ij} + e_{ij} \sqrt{\pi_{ij}(1 - \pi_{ij})}$$

$$\text{var}(e_{0ij}) = 1$$

where we evaluate π_{ij} at the mean of the distribution of the level 2 random effect, that is, for the logistic model

$$\pi_{ij} = \exp(\beta_0 + \beta_1 x_{1ij}) [1 + \exp(\beta_0 + \beta_1 x_{1ij})]^{-1}$$

$$\pi'_{ij} = \pi_{ij} [1 + \exp(\beta_0 + \beta_1 x_{1ij})]^{-1} \tag{4}$$

so that, for a given value of x_1 we have

$$\text{var}(y_{ij} | x_{1ij}) = \sigma_{u_0}^2 \pi_{ij}^2 ([1 + \exp(\beta_0 + \beta_1 x_{1ij})]^{-2} + \pi_{ij}(1 - \pi_{ij}))$$

and

$$\tau = \sigma_{u_0}^2 \pi_{ij}^2 ([1 + \exp(\beta_0 + \beta_1 x_{1ij})]^{-2} \{ \sigma_{u_0}^2 \pi_{ij}^2 ([1 + \exp(\beta_0 + \beta_1 x_{1ij})]^{-2} + \pi_{ij}(1 - \pi_{ij})) \}^{-1}$$

where sample estimates are substituted.

3.2 Simulation (Method B)

This method is general and can be applied to any non-linear model without the need to evaluate an approximating formula. It consists of the following steps:

1. From the fitted model (say (3)) simulate a large number m (say 5000) values for the level 2 residual from the distribution $N(0, \sigma_{u_0}^2)$, using the sample estimate of the variance.
2. For a particular chosen value(s) of x_1 compute the m corresponding values of π_{ij} (π_{ij}^*) using (4). For each of these values compute the level 1 variance $v_{1ij} = \pi_{ij}^*(1 - \pi_{ij}^*)$.
3. The coefficient is now estimated as

$$\tau = v_2(v_2 + v_1)^{-1}$$

$$v_2 = \text{var}(\pi_{ij}^*), \quad v_1 = E(v_{1ij})$$

3.3 A binary linear model (Method C)

As a very approximate indication for the VPC we can consider treating the (0, 1) response as if it were a Normally distributed variable as in (1) and estimate the VPC as in that case. This will generally be acceptable when the probabilities involved are not extreme, but if any of the underlying probabilities are close to 0 or 1, this model would not be expected to fit well, and may predict probabilities outside the (0, 1) range.

3.4 An Example

We illustrate the above three procedures using data on voting patterns (Heath et al, 1996).

The model response is whether or not, in a sample of 800 respondents, they expressed a preference for voting conservative ($y_{ij} = 1$) in the 1983 British general election. Several covariates were originally fitted but for simplicity we fit only an intercept model, namely

$$E(y_{ij}) = \pi_{ij} = \exp(\beta_0 + u_{0j})[1 + \exp(\beta_0 + u_{0j})]^{-1}$$

$$y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2)$$

The fitted model parameters are as follows (using PQL2 with IGLS (Goldstein and Rasbash, 1996)):

$$\hat{\beta}_0 = -0.256, \quad \hat{\sigma}_{u0}^2 = 0.142$$

The corresponding estimates are

(Table 1 here)

There is, as expected, good agreement for all three methods. Now, however, we choose an extreme value. In Chapter 8 of Rasbash et al. (2000) these data are fitted with 4 predictor variables measuring political attitudes. The scales of these predictor variables are constructed such that lower values correspond to left wing attitudes. If we take the set of values corresponding to the 5th percentile point of each of the four predictors, then the predicted value on the logit scale is approximately -2.5 , corresponding to a low probability of voting Conservative of 0.076. At this value of the predictor the respective VPCs for methods A and B are 0.0096 and 0.0111, which are again in reasonable agreement. For method C, fitting the four predictor variables, we obtain a value of 0.0257 which is very different.

3.5 A latent variable approach (Method D)

In some circumstances we may wish to think of an observed (0, 1) as arising from an underlying continuous variable so that a 1 is observed when a certain threshold is exceeded, otherwise a 0 is observed. For the logit model we have the underlying logistic distribution

$$f(x) = \frac{\exp(x)}{[1 + \exp(x)]^2} \quad (5)$$

with cumulative distribution function

$$\int_{-Y}^{\infty} f(x) dx = [1 + \exp(-Y)]^{-1} \quad (6)$$

The right hand side of equation (6) is simply the logit link function model with Y as the linear component incorporating level 2 variation. The variance for the standard

logistic distribution (5) is $\pi^2/3 = 3.29$ so we take this to be the level 1 variance and both the level 1 and level 2 variances are on a continuous scale. We now simply calculate the ratio of the level 2 variance to the sum of the level 1 and level 2 variances to obtain the VPC. Corresponding to the values in table 1 we obtain a VPC of 0.041, which is somewhat different to the others. A similar device can be used for the probit link function.

This approach may be reasonable where the (0, 1) response is, say, derived from a truncation of an underlying continuum such as a pass/fail response based upon a continuous mark scale, but would seem to have less justification when the response is truly discrete, such as mortality or voting. See also Snijders and Bosker (1999, Chapter 14) for a further discussion.

4. Conclusions

In summary, if a VPC is required for non-linear models, it should be computed for a range of values of the predictor variables. Methods A or B or, where appropriate method D can all be used. The advantage of method B is that it does not involve any approximation and is simple and fast to compute. We can readily extend the computations for random coefficient models; with method B this simply involves simulating from a multivariate Normal distribution using the estimated level-2 covariance matrix.

We can obtain an interval estimate for any of our estimates via the bootstrap or using the results from an MCMC estimation run. In the former case we would generate a suitable number (for example 999) bootstrap data sets with associated model fits, and then using one of the above procedures obtain a sample of (999) values from which quantiles can be estimated, as in the first example of the paper. In the MCMC case we would run a chain (of length say 5000) where each MCMC cycle in the chain generates a sample from the posterior distribution of the parameters, and using one of the above procedures this likewise will produce a sample of values from the posterior distribution of the chosen VPC (see Rasbash et al., 2000, Chapter 15).

References

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Newbury Park, California, Sage:

Goldstein, H (1995). *Multilevel Statistical Models*, Second Edition. London: Edward Arnold.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nutall, D., and Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education*, 19, 425-433.

Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, A*. 159: 505-13.

Heath, A., Yang, M. and Goldstein, H. (1996) Multilevel analysis of the changing relationship between class and party in Britain 1964-1992. *Quality and Quantity* 30: 389-404.

Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., et al. (2000). *A user's guide to MLwiN (Second Edition)*. London, Institute of Education:

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis*. London, Sage:

Appendix: Sample MLwiN Macros for methods A and B.

Methods (C) and (D) are trivial to evaluate. Below are macros for evaluating methods (A) and (B) for any two level binomial model using the MLwiN macro language. The macros require the list of x values for which the variance partition coefficient is to be calculated in c151. Column c152 contains the subset of c151 that have random effects. The text for these macros can be executed in an MLwiN macro window. The commands Print b7 or Print b8 display the results from methods A and B in the output window.

Macro for method A.

```
note c151 contains values for set of x variables for which
note: to calculate variance partition coefficient
note c152 contains subset of c151 with random effects at level 2
note calculate (XB) and store in b2 and pi=antilogit(XB) in b3
calc c153=(~c151)*.c98
pick 1 c153 b2
note calc. Le.1 2 variance for chosen vals. of expl. Vars: store in b4
calc c153=(~c152) *. omega(2) *. c152
pick 1 c153 b4
note pi^2*Su^2. Su is level 2 variance matrix
calc b3=alog(b2)
calc b5=b3^2*b4
calc b6=b5/(1+expo(b2))^2
calc b7=b6/(b6+b3*(1-b3))
Print b7
```

Macro for method B.

```
note c151 contains values for set of x variables for which
note: to calculate variance partition coefficient
note c152 contains subset of c151 with random effects at level 2
note calculate (XB) and store in b2
calc c153=(~c151)*.c98
pick 1 c153 b2
note calc. Le.1 2 variance for chosen vals. of expl. Vars: store in b4
calc c153=(~c152) *. omega(2) *. c152
pick 1 c153 b4
nran 5000 c154
calc c154=alog(c154*b4^0.5+b2)
aver c154 b1 b3 b2
calc c154=c154*(1-c154)
aver c154 b5 b1
calc b8=b2^2/(b1+b2^2)
Print b8
```

Table 1. Level 2 and level 1 estimated variances with variance partition coefficient (VPC)

	Method		
	A	B	C
Level 2 variance	0.0086	0.0083	0.0088
Level 1 variance	0.246	0.238	0.237
VPC	0.034	0.034	0.036

Figure 1: Plot of variance partition coefficient for different reading test scores

