**Supplemental materials: Partitioning variation in multilevel models for count data**

# S1. Review of two-level models for continuous responses

In this section we review two-level variance-components, random-intercept and random-coefficient models for continuous responses and their associated VPC and ICC expressions. The variance components model is simply the special case of the random-intercept model where we include no covariates (the “null” or “empty” model). To aid the discussion of each model, we present the conditional (given the covariates and random effects) and marginal (given the covariates but averaged over the random effects) expectations, variances, covariances and correlations of the continuous response. Reviewing this material will help readers better understand the more complex derivations we present in the article for multilevel count response models. When reading this section, readers may find it helpful to have an example application in mind such as the study of student math scores and student and school characteristics that we described in the Introduction.

## S1.1. Variance-components model for continuous responses

 The two-level variance components model for continuous responses simply decomposes the total response variance into separate variance components at level-1 and level-2 in the data hierarchy.

### Model

Let denote the continuous response for unit () in cluster (). In terms of the math score study, the units would be students, the clusters schools, and the continuous response would be the student math score. We can then write the model for as follows

 (S1.1)

where denotes the intercept, is the cluster random intercept effect, assumed normally distributed with zero mean and between-cluster variance , and is the residual, assumed normally distributed with zero mean and within-cluster variance . The response variance (total, overall or marginal) is then given by the summation .

 The VPC measures the proportion of response variance which lies between clusters and is calculated as the ratio of the between-cluster variance to the total response variance

 (S1.2)

The ICC measures the expected correlation between two units from the same cluster and is again calculated as the ratio of the between-cluster variance to the total response variance

 (S1.3)

The above expressions for the VPC and ICC are widely used by applied researchers. However, they are typically presented with little explanation.

### Conditional inference

The conditional expectation of (given ) is given by

 (S1.4)

and measures the mean response in each cluster .

 The conditional variance is given by

 (S1.5)

and measures the variance of the responses around the mean response in each cluster.

The conditional covariance and correlation between the response measurements on two units and for the same cluster , and , are zero and so the responses are assumed independent in each cluster. This assumption is sometimes referred to as the conditional independence or local independence assumption. The conditional covariance and correlation between response measurements on two units from two different clusters are also assumed equal to zero.

### Marginal inference

The marginal expectation of (now averaged over ) is given by

 (S1.6)

and measures the overall mean response.

The marginal variance of is given by

 (S1.7)

and measures the total or overall response variance and is equal to the sum of the variance components.

The marginal covariance between and is given by

 (S1.8)

The marginal correlation between and can then be calculated as

 (S1.9)

This marginal correlation can be interpreted as the ICC as it is the response correlation between two units in the same cluster. In contrast, the marginal covariance and correlation between response measurements on two units from two different clusters is equal to zero as the units do not share a cluster random effect.

 The VPC is defined as the proportion of the marginal response variance which lies between clusters. Thus, we can only calculate the VPC after we have partitioned the marginal variance into level-specific components. The expression for the marginal variance (Equation S1.7) does just that. The expression for the VPC is therefore given by

 (S1.10)

and this expression is identical to that for the marginal correlation or ICC given in Equation S1.9. We see the usual result whereby the higher the between-cluster variance , the higher the VPC, but the higher the within-cluster variance , the lower the VPC. This VPC should strictly be referred to as the level-2 VPC. The level-1 VPC – the proportion of the marginal response variance which lies within clusters – is then equal to one minus the level-2 VPC.

## S1.2. Two-level random-intercept models with covariates for continuous responses

So far, we have reviewed the two-level variance-component model for continuous responses and we have shown how to calculate the VPC and ICC expressions for this model. We now extend this to the two-level random-intercept model with covariates and discuss the implications this has for calculating and interpreting the VPC and ICC.

### Model

The model can be written as follows

 (S1.11)

where denotes the vector of unit- and cluster-level covariates (including the intercept and any cross-level interactions), and is the associated vector of regression coefficients. In terms of our math score study, the unit- and cluster-level covariates would be student and school characteristics believed to be relevant for explaining the variation in students’ scores.

It may help the reader to consider the special case of a model with a random intercept and one covariate . In terms of our math score study this might be student SES. Such a model can then be written as

 (S1.12)

where denotes the intercept and denotes the slope coefficient on .

The expressions for the VPC and ICC in the random-intercept model with covariates are the same as in the variance-components model. Thus, they again coincide to give

 (S1.13)

Here, the interpretation of these statistics is in terms of the response variation which remains after adjusting for the covariate. Thus, the VPC and ICC now answer the questions: What proportion of the residual math score variation lies between schools? And: What is the residual math score correlation between two students from the same school. As with the variance-components model, applied researchers widely report the VPC and ICC after fitting random-intercept models with covariates. However, they again typically present these statistics without derivation and so we review their derivation here.

The introduction of covariates mean that we must update the VPC and ICC expressions ( as well as the various other conditional and marginal statistics) given above for the variance-components model so that they now additionally condition on . This then leads us to replace with in any expression where was shown previously. As we shall see below, this affects only the expressions for the conditional and marginal expectations as does not appear in any of the other expressions. This is why the expressions for the VPC and ICC remain the same as those for the variance-components model (Equation S1.13 takes the same form as Equation S1.3). The interpretation of each statistic, however, does change upon the introduction of covariates as indicated above. Specifically, the interpretation of each statistic is now in terms of making statements about the variation in the response conditional on the covariates. Thus, these statistics now allow us to make statements about the response variation which remains once we have adjusted for the covariates (i.e., the unexplained or residual variation).

### Conditional statistics

The conditional expectation of (now given and ) is given by

 (S1.14)

and so is now a function of the covariates . The conditional variance is given by

 (S1.15)

and so takes the same form as before. The conditional covariance and correlation between and are again assumed to be zero.

### Marginal statistics

The marginal expectation of (again given but now averaged over ) is given by

 (S1.16)

and so is now a function of the covariates . The marginal variance of is given by

 (S1.17)

and so is again the sum of the variance components. Note, however, these variance components no longer sum to give the response variance. Rather they sum to give the response variance having adjusted for the covariates. The marginal covariance between and (given and but averaged over ) is given by

 (S1.18)

and so also takes the same form as before. The marginal correlation between and can then be calculated as

 (S1.19)

and so takes the same form as before.

 The VPC (given ) is derived in the usual way, as the ratio of the level-2 component of the marginal variance divided by the summation of the level-2 and -1 components. The expression for the marginal variance takes the same form as in the variance-component case and so the expression for the VPC also takes the same form and is given again by

 (S1.20)

Thus, the expressions for the VPC and the ICC are the same as one another and are also the same in both the variance-components model and in the random-intercept model with covariates.

## S1.3 Two-level random-coefficient models

Next, we consider the two-level random-coefficient model. We then discuss the implications this has for the calculation and interpretation of the VPC and ICC. The two-level random-coefficient model extends the two-level random-intercept model by allowing not just the intercept, but one or more of the slope coefficients to vary randomly across clusters.

### Model

The model can be written as follows

 (S1.21)

where denotes a vector of unit- and cluster-level covariates (typically an intercept and a subset of the unit-level covariates in ) and is the associated vector of cluster random coefficient effects, assumed multivariate normally distributed with zero mean vector and covariance matrix .

It may help the reader to consider the special case of a model with a random-intercept and one covariate and where we enter that covariate with a random coefficient. Thus, we enter the covariate both in the fixed part of the model as and in the random part of the model as . In terms of our math score example application this might be student SES. The model can be written as

 (S1.22)

where denotes the intercept and denotes the slope coefficient on , and are the cluster random intercept and slope effects, assumed bivariate normally distributed with zero means and between-cluster intercept and slope variances and covariance , and , and is the usual residual, assumed normally distributed with zero mean and within-cluster variance .

As we shall show below, the expressions for the VPC and ICC can also be derived for this more complex model.

### Conditional statistics

The conditional expectation of (now given , and ) is given by

 (S1.23)

 The conditional variance is given by

 (S1.24)

and so takes the same form as in the simple variance-components and random-intercept models. The conditional covariance and correlation between and are again assumed zero.

### Marginal statistics

The marginal expectation of (again given and but now averaged over ) is given by

 (S1.25)

which is the same as in the random-intercept case. The marginal variance of is given by

 (S1.26)

which now takes a more complex form than that seen previously. The first term is referred to as a cluster-level variance function as the response variance between clusters is now modelled as a function of the covariates (heteroskedasticity). It follows that the marginal variance is also a function of the covariates since to obtain this we simply add on the second term .

The marginal covariance between and (given , , , but averaged over ) is given by

 (S1.27)

The marginal correlation between and can then be calculated in the usual way to give

 (S1.28)

The ICC is therefore a function of and . To simplify matters, consider the correlation between two units and within the same cluster who have the same covariate values . The expression for the ICC then simplifies to

 (S1.29)

Now reconsider our example model with only one covariate . For this model, the ICC (Equation S1.28) becomes

 (S1.30)

The magnitude of the residual response correlation (i.e., the response correlation having adjusted for the covariate) between units and for the same cluster therefore depends on the value of the covariate with the random coefficient for the two units being compared, and . Restricting our attention to the special case when leads this ICC expression (Equation S1.29) to simplify to

 (S1.31)

The ICC is then a function of just one variable and this can be more easily plotted and evaluated.

The VPC (given and ) is derived again as the ratio of the level-2 component of the marginal variance divided by the summation of the level-2 and -1 components. The expression for the marginal variance is more complex than that for the variance-component and random-intercept models leading the expression for the VPC to also be more complex. The VPC is given by

 (S1.32)

and this expression is identical to that for the simplified marginal correlation or ICC (where ) given in Equation S1.29. This VPC is the level-2 VPC. The level-1 VPC – the proportion of the marginal response variance which lies within clusters – is simply equal to one minus the level-2 VPC.

Reconsider, once again, our example model with only one covariate . For this model, the VPC becomes

 (S1.33)

and this expression is identical to that for the simplified marginal correlation or ICC (where ) given in Equation S1.31.

# S2. Textbook and other coverage of conditional and marginal statistics for the two-level random-intercept Poisson model

In this section we review the presentation of the two-level random-intercept Poisson model in the standard multilevel modelling textbooks by Goldstein (2011b), Hox et al. (2017), Raudenbush and Bryk (2002), and Snijders and Bosker (2012). We focus on the extent to which these textbooks present and explain the conditional and marginal statistics implied by the model. We also include in our review the article by Austin et al. (2017) since this is the only article which presents the VPC and ICC for this model. One further resource we consider here as we have found it particularly useful is the advanced statistics text by McCulloch et al. (2008).

## S2.1 Review

Table S2.1 indicates the conditional and marginal statistics presented in each resource. We see that all resources present the conditional expectation and variance. The model is a conditional model and so this is expected. Where the resources differ is in their presentation of the marginal statistics: the marginal expectation, variance, covariance, correlation, including the ICC, level-2 and level-1 components of the marginal variance, and the VPC. Few of the resources present any of these statistics. Out of the standard textbooks on multilevel modelling, only Raudenbush and Bryk (2002) present any of these marginal statistics and even then they only present the marginal expectation.

Austin et al. (2017) do of course present the VPC. They also make clear that this expression coincides with that for the ICC (when we restrict comparisons to units with the same covariate values). They do not however present the four other marginal statistics reviewed here: the marginal expectation, variance, covariance, and correlation.

The best resource we have found is the advanced statistics text by McCulloch et al. (2008) which does presents the marginal expectation, variance, covariance and correlation for the two-level random-intercept Poisson model. However, this textbook does not additionally discuss the VPC and ICC. The ICC is simply the marginal correlation, but the authors do not make this connection clear to the reader. The expression for the VPC coincides with that for the ICC and therefore the marginal correlation, but neither do the authors make this link clear. We have shown that the VPC is derived by partitioning the marginal variance into variance components operating at the cluster and unit level. McCulloch et al. (2008) do in fact derive each of these level-specific terms prior to summing them to derive the marginal variance. The authors do not however make clear that these two terms can be interpreted as level-specific components of the marginal variance.

In sum, there is very little presentation or explanation in standard multilevel textbooks and other resources reader might turn to of the marginal statistics implied by the two-level random-intercept Poisson model. In the few cases where there is some coverage of the marginal statistics, this coverage ceases once we turn to the negative binomial, three-level and random-coefficient extensions to the two-level random-intercept Poisson model which are the focus of this article.

## S2.1 References not found in main reference list

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models (2nd ed.)*. New York, USA: John Wiley & Since, Inc.

*Table S2.1*. Textbook and other coverage of conditional and marginal statistics for the two-level random-intercept Poisson model by resource

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Description | Notation | Goldstein (2011b) | Hox et al. (2017) | Raudennbush and Bryk (2002) | Snijders and Bosker (2012) | Austin et al. (2017) | McCulloch et al. (2008). |
| Conditional expectation |  | X | X | X | X | X | X |
| Conditional variance |  | X | X | X | X | X | X |
| Marginal expectation |  |  |  | X |  |  | X |
| Marginal variance |  |  |  |  |  |  | X |
| Marginal covariance |  |  |  |  |  |  | X |
| Marginal correlation |  |  |  |  |  |  | X |
| Intraclass correlation coefficient (ICC) |  |  |  |  |  | X |  |
| Marginal variance: Level-2 component |  |  |  |  |  |  |  |
| Marginal variance: Level-1 component |  |  |  |  |  |  |  |
| Variance partition coefficient (VPC) |  |  |  |  |  | X |  |

# S3. Review of two further two-level random-intercept models for count data

In the article, we review the two most widely applied multilevel models for count responses, the Poisson model and the negative binomial model (mean dispersion or NB2 version). In this section, we review two further count models, the Poisson model with an overdispersion random effect and the constant dispersion or NB1 version of the negative binomial model. Table S4.1 presents a summary table allowing readers to compare the conditional and marginal variances across all four models.

## S3.1 Poisson model with overdispersion random effect

The Poisson random-intercept model (Equation 1 where we substitute for ) includes a normally distributed cluster random intercept effect to account for the clustering in the data. A natural way to deal with overdispersion in this model is to therefore add a normally distributed unit-level overdispersion random effect to represent omitted unit-level variables that are envisaged to be driving any overdispersion. In contrast to the conventional cluster random intercept effect, this does not induce any dependence among the units.

### Model

The new model can be written as

 (S3.1)

where denotes the overdispersion random effect. Thus, in this model, two units with the same covariate and random intercept effect value may nonetheless differ in their expected counts with such differences attributed to the two units differing in terms of their values on omitted unit-level variables. This is also a feature of the two negative binomial models considered in this article (Equation 9 and Equation S3.9). The overdispersion random effect is assumed normally distributed with zero mean and variance or overdispersion parameter . The larger is, the greater the overdispersion. When , the model simplifies to the Poisson model (Equation 1) permitting a likelihood-ratio test for whether any estimated overdispersion is statistically significant.

### Conditional statistics

In this model, we can again calculate the conditional expectation and variance of the response. However, here we must integrate out the overdispersion random effect since this is not typically of substantive interest. To do this, we exploit the fact that is log normally distributed and so its expectation and variance have known forms. Thus, the conditional expectation of (given and but averaged over ) is now given by

 (S3.2)

and we see that, in contrast to the Poisson model (Equation 2), . The conditional variance is then given by

 (S3.3)

Thus, the conditional variance is now a quadratic function of the conditional expectation and is larger than the conditional expectation if . Therefore, the usual variance-mean relationship for the Poisson model (Equation 3) is relaxed, producing overdispersion with respect to the conditional expectation ().

To help see the similarities between the mean dispersion negative binomial model (Equation 9) and the current model, recall that in the former we assume while in the latter we assume , implying . Thus, the two models differ only in the distribution they assume for the exponentiated overdispersion random effect (gamma vs. log-normal).

### Marginal statistics

The marginal expectation of (given but now averaged over as well as ) is given by

 (S3.4)

which differs from that for the Poisson and mean dispersion negative binomial models (Equations 4 and 12) via the inclusion of the additional term .

The marginal variance of is given by

 (S3.5)

which differs from that for the Poisson model (Equation 5) via the inclusion of the additional multiplicative term . Thus, in this model, the marginal variance is larger than the marginal expectation if there is clustering  *or* overdispersion . Note that this expression takes the same form as that for the mean dispersion negative binomial model where takes the place of .

The marginal covariance of and (given and but averaged over , and ) is given by

 (S3.6)

and is the same as that for the Poisson and mean dispersion negative binomial models (Equations 6 and 14). The marginal correlation of and is then given by

  (S3.7)

and as with the Poisson and mean dispersion negative binomial models this statistic can be interpreted as the ICC. The expression differs from that for the Poisson model (Equation 7) only in the inclusion of the additional multiplicative term in the denominator. Here too, this expression takes the same form as that for the mean dispersion negative binomial model where takes the place of .

 As with the Poisson and mean dispersion negative binomial models, we can partition the marginal variance (Equation S3.5) into level-specific components which capture the within- and between-cluster variance in (Supplemental materials S4.2). The resulting level-2 VPC is given by

 (S3.8)

and this expression is identical (after rearranging terms) to that for the marginal correlation or ICC given in Equation S3.7.

Studying Equation S3.8, we see that, as in the Poisson case (Equation 8), the VPC is an increasing function of both the marginal expectation and the cluster variance . However, the VPC is now also a decreasing function of the overdispersion parameter . This makes sense. As the overdispersion increases, all else equal, the more unmodelled variation there is at level-1 and so the VPC decreases. Here too, this expression takes the same form as that for the mean dispersion negative binomial model where takes the place of .

Comparing the three VPC expressions (Equations 8, 16 and S3.8), we see that the expression for the level-2 component of the marginal variance is the same and so it is only the expression for the level-1 component which varies across models. This makes sense as the models differ only in their treatment of overdispersion, which is viewed as a level-1 phenomenon. The overdispersion parameter in the current model (and the mean dispersion negative binomial model) leads the expression for the level-1 component of the marginal variance to exceed that of the Poisson model. The marginal variance is simply the summation of the level-2 and -1 variances and so is also expected to be higher in the current model (and mean dispersion negative binomial model) compared to that of the Poisson model.

## S3.2 Negative binomial model: Constant dispersion version

The constant dispersion or NB1 version of the negative binomial model cannot be derived from a version of the Poisson model with a unit-level overdispersion random effect.

### Model

The model is instead written as

 (S3.9)

where we now assume the expected count has a conditional gamma distribution (given and ) with shape and scale parameters and and therefore expectation and variance where is the overdispersion parameter. The larger is, the greater the overdispersion. When , the variance of this gamma distribution is equal to zero and the model simplifies to the Poisson model (Equation 1), once again permitting a likelihood-ratio test for whether the estimated overdispersion is statistically significant.

### Conditional statistics

It follows that the conditional expectation of (given and ) is again the same as in the Poisson and mean dispersion negative binomial models (Equations 2 and 10), with

 (S3.10)

However, the conditional variance of is now

 (S3.11)

Thus, in contrast to the quadratic form seen in the two unit-level random effect overdispersion models (Equation 11 and S3.3), the variance in the current model is a constant multiple of the conditional expectation.

 To help see the similarities between the two versions of the negative binomial model, note that we can rewrite the mean dispersion model (Equation 9) as follows

 (S3.12)

Thus, both versions of the negative binomial model assume the expected count follows a conditional gamma distribution, but the models differ in terms of the shape and scale parameters of this distribution. In the mean dispersion model (Equation 9), the shape and scale parameters are chosen such that the conditional expectation of is and conditional variance is whereas in the constant dispersion model (Equation S3.12) the shape and scale parameters are chosen such that the conditional expectation of is and conditional variance is . Thus, the difference between the two versions of the negative binomial model can be viewed as a difference in the relationship between the conditional variance and conditional expectation of . This difference then leads the models to differ in terms of the conditional variances of the response (Equation 11 vs. S3.11).

### Marginal statistics

The marginal expectation of (given but averaged over ) is given by

 (S3.13)

which is the same as that for the Poisson and mean dispersion negative binomial models (Equation 4 and 12).

The marginal variance of is given by

 (S3.14)

which differs from that for the Poisson model (Equation 5) in that the first term is now multiplied by . Thus, in this model, the marginal variance is larger than the marginal expectation if there is clustering  *or* overdispersion . Note that this expression takes a diferent form to that for the Poisson model with overdispersion random effect or the mean dispersion negative binomial model.

The marginal covariance of and (given and but averaged over ) is given by

 (S3.15)

and is the same as that for the Poisson, mean dispersion negative binomial, and Poisson model with overdispersed random effect models (Equations 6, 14 and S3.6). The marginal correlation of and is then given by

 (S3.16)

and as with the other three models can be interpreted as the ICC. The expression differs from that for the Poisson model (Equation 7) only in the inclusion of the additional multiplicative term in the denominator.

 As with the other three models, we can partition the marginal variance (Equation S3.14) into level-specific components which capture the within- and between-cluster variance in (Supplemental materials S4.2). The resulting level-2 VPC is given by

 (S3.17)

and this expression is identical (after rearranging terms) to that for the marginal correlation or ICC given in Equation S3.16.

Studying Equation S3.17, we see that, as with the other models which allow for overdispersion, the VPC is an increasing function of both the marginal expectation and the cluster variance , but is a decreasing function of the overdispersion parameter, here .

Comparing all four VPC expressions (Equations 8, 16, S3.8 and S3.17), we see that the expression for the level-2 component of the marginal variance is always the same and so it is only the expression for the level-1 component which varies across models. This makes sense as the models differ only in their treatment of overdispersion, which is viewed as a level-1 phenomenon. The overdispersion parameter in all three models which allow for overdispersion lead the expression for the level-1 component of the marginal variance to exceed that of the Poisson model. The marginal variance is simply the summation of the level-2 and -1 variances and so is also expected to be higher in all these models compared to that of the Poisson model.

# S4. Derivation of the marginal statistics in two-level random-intercept models: Marginal expectation, variance, covariance, correlation, VPCs and ICC

We now present general derivations for the marginal expectation, variance, covariance and correlation for two-level random-intercept models (McCulloch et al., 2008). These apply to the continuous response models reviewed in Supplemental materials S1 as well as all the count models explored in our work. We then show that the ICC is simply the marginal correlation. We then decompose the marginal variance into components of variance at each level and show that the VPC can be calculated as the ratio of the level-2 component to the sum of the level-1 and level-2 components. All the expressions for the marginal statistics presented in the article have been obtained using these general derivations. Table S4.1 presents these expressions in tabular form to facilitate comparisons across models.

## S4.1 Marginal expectation

The marginal expectation of (given but averaged over ) can be derived by exploiting the law of total expectations (; law of iterated expectations; McCulloch et al., 2008)

 (S4.1)

## S4.2 Marginal variance

The marginal variance of (given but averaged over ) can be derived by exploiting the law of total variance (; McCulloch et al., 2008)

 (S4.2)

It is instructive to realize that the law of total variance decomposes the marginal variance into separate level-2 and level-1 specific variance components. The level-2 variance component captures between cluster variation in the cluster specific expected counts attributable to the cluster random intercept effect (i.e., given ). The level-1 variance component captures within cluster variation in around these cluster specific expected counts (i.e., given ) averaged across all clusters (as given still varies across clusters as function of ).

## S4.3 Marginal covariance

To derive the marginal covariance of and (given and but averaged over ) , we exploit the law of total covariance (; McCulloch et al., 2008).

 (S4.3)

The second term equals zero due to the assumption of conditional independence among the responses for the same cluster given the covariates and random intercept effect. Namely, and are assumed conditionally independent (given , and ) (i.e., assuming independent Poisson sampling variation). This is sometimes referred to as the local independence assumption.

## S4.4 Marginal correlation

The marginal correlation or ICC can then be calculated in the usual way

 (S4.3)

## S4.5 Intraclass correlation coefficients (ICCs)

 We focus on the special case where the two units have the same covariate values in which case and and so the ICC simplifies to

 (S4.4)

## S4.5 Variance partition coefficients (VPCs)

Variance partition coefficients report the proportion of the total variation in the observed counts (given the covariates) which lies at each level of analysis. These can be calculated in the usual way, as ratios of each variance component to the marginal variance. The level-2 VPC is therefore calculated as

 (S4.5)

This is the same expression as that given above for the ICC (Equation S4.4).

The level-1 VPC is calculated as 1 minus the level-2 VPC

 (S4.5)

**S4.8 References not found in main reference list**

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models (2nd ed.)*. New York, USA: John Wiley & Since, Inc.

Table S4.1.

Two-level count response models: Expressions for the conditional and marginal expectations and variances of the response and the level-1 and -2 components of the marginal variance used in the calculation of the VPCs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Description | Notation | Poisson model | Poisson modelwith overdispersion random effect | Negative binomial model:Mean dispersion or NB2 version | Negative binomial model:Constant dispersion or NB1 version |
| Conditional expectation |  |  |  |  |  |
| Conditional variance |  |  |  |  |  |
| Marginal expectation |  |  |  |  |  |
| Marginal variance |  |  |  |  |  |
|  Level-2 component |  |  |  |  |  |
|  Level-1 component |  |  |  |  |  |

Note.

The above expressions are for two-level random-intercept models. The corresponding expressions for the two-level models with random coefficients are obtained by replacing and in all expressions with and .

# S5. Derivation of the marginal statistics in three-level random-intercept models: Marginal expectation, variance, covariance, correlation, VPCs and ICCs

We now present general derivations for the marginal expectation, variance, covariance and correlation for three-level random-intercept models. We then decompose the marginal variance into components of variance at each level and show that the different VPCs can be calculated as ratios of the level-specific variance components. All the expressions for the marginal statistics presented in the article have been obtained using these general derivations. Table S5.1 presents these expressions in tabular form to facilitate comparisons across models.

## S5.1 Marginal expectation

The marginal expectation of (given but averaged over and ) can be derived by exploiting the law of total expectations (; law of iterated expectations)

 (S5.1)

## S5.2 Marginal variance

The marginal variance of (given but averaged over and ) can be derived by repetitively exploiting the law of total variance (

First decompose the marginal variance of (given but averaged over and ) into a between supercluster and within supercluster components

 (S5.2)

Next decompose the conditional variance of (given and but averaged over ) into a between cluster and within cluster component (i.e., decompose the contents of the expectation in the second term of the above expressions).

 (S5.3)

Substitute Equation S5.3 into Equation S5.2 to give

 (S5.4)

This expression can be written more concisely as

 (S5.5)

The level-3 variance component captures between supercluster variation in the supercluster specific expected counts attributable to the supercluster random intercept effect (i.e., given ). The level-2 variance component captures the within supercluster, between cluster variation in the cluster specific expected counts attributable to the cluster random intercept effect (i.e., given ) and averaged across all superclusters. The level-1 variance component captures within cluster variation in around these cluster specific expected counts (i.e., given ) averaged across all clusters (as given still varies across clusters as function of and ).

## S5.3 Variance partition coefficients (VPCs)

Variance partition coefficients report the proportion of the total variation in the observed counts (given the covariates) which lies at each level of analysis. These can be calculated in the usual way, as ratios of each variance component to the marginal variance. The level-3 VPC is therefore calculated as

 (S5.5)

The level-2 VPC is calculated as

 (S5.6)

The level-1 VPC is calculated as

 (S5.7)

Table S5.1.

Three-level count response models: Expressions for the conditional and marginal expectations and variances of the response and the level-1, -2 and -3 components of the marginal variance used in the calculation of the VPCs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Description | Notation | Poisson model | Poisson modelwith overdispersion random effect | Negative binomial model:Mean dispersion or NB2 version | Negative binomial model:Constant dispersion or NB1 version |
| Conditional expectation (level-2) |  |  |  |  |  |
| Conditional variance (level-2) |  |  |  |  |  |
| Conditional expectation (level-3) |  |  |  |  |  |
| Conditional variance (level-3) |  |  |  |  |  |
| Marginal expectation |  |  |  |  |  |
| Marginal variance |  |  |  |  |  |
|  Level-3 component |  |  |  |  |  |
|  Level-2 component |  |  |  |  |  |
|  Level-1 component |  |  |  |  |  |

Note.

The above expressions are for three-level random-intercept models. The corresponding expressions for the three-level models with random coefficients are obtained by replacing and in all expressions with and and by replacing and in all expressions with and .

# S6. Derivation of the marginal expectation, variance, covariance and correlation in two- and three-level random-coefficient models

All expressions derived in Supplemental materials S4 were for two-level random-intercept models. The corresponding expressions for two-level models with random coefficients are obtained by replacing and in all expressions with and .

All expressions derived in Supplemental materials S5 were for three-level random-intercept models. The corresponding expressions for three-level models with random coefficients are obtained by replacing and in all expressions with and and by replacing and in all expressions with and .

# S7. Additional analyses for student absenteeism application

Table S7.1 presents variable definitions and summary statistics for the student covariates used in the models.

*Table S7.1*. Covariate distributions, including mean days absent.

|  |  |  |  |
| --- | --- | --- | --- |
| Student characteristic | N | Percent | Mean days absent |
| Prior attainment (quintile) |  |  |  |
|  1 (lowest prior attainment) | 14366 | 21% | 10.1 |
|  2 | 13288 | 20% | 9.0 |
|  3 | 12703 | 19% | 8.3 |
|  4 | 16000 | 24% | 7.4 |
|  5 (highest prior attainment) | 10598 | 16% | 7.0 |
| Age |  |  |  |
|  Summer (youngest in year) | 17032 | 25% | 8.1 |
|  Spring | 16495 | 25% | 8.2 |
|  Winter | 16551 | 25% | 8.5 |
|  Autumn (oldest in year) | 16877 | 25% | 8.8 |
| Gender |  |  |  |
|  Male | 33628 | 50% | 8.0 |
|  Female | 33327 | 50% | 8.8 |
| Ethnicity |  |  |  |
|  White | 27803 | 41% | 9.6 |
|  Mixed |  6181 | 09% | 9.6 |
|  Asian | 13914 | 21% | 7.1 |
|  Black | 14409 | 22% | 7.0 |
|  Other |  4648 | 7% | 7.9 |
| Language  |  |  |  |
|  English | 40529 | 61% | 9.2 |
|  Not English | 26426 | 39% | 7.2 |
| SEN |  |  |  |
|  Not SEN | 57157 | 85% | 7.9 |
|  SEN |  9798 | 15% | 11.4 |
| FSM |  |  |  |
|  Not FSM | 41606 | 62% | 7.3 |
|  FSM | 25349 | 38% | 10.2 |

*Note*. Number of school districts: ; number of schools: ; number of students: . Prior attainment quintiles are based on an average test score across separate tests in English and maths taken five years earlier at the end of primary schooling, just before the start of secondary schooling.

# S8. Simulation method for calculating the VPC

Until now the only way to calculate the VPC and ICC (and other marginal statistics) after fitting multilevel count models was to use the simulation method. The simulation method can be used to approximate the unknown expressions for calculating the VPC and ICC. Now that we have derived these expressions, the simulation method is redundant. In this section, we confirm that our VPC and ICC derivations are correct by showing that had we used the simulation method rather than our expressions, we would have calculated the same values for the VPC and ICC in each model in our application.

 The simulation method works by treating the fitted model as a data generating process. We then then simulate a single very large dataset. The marginal statistics are then estimated by forming different summaries of the simulated data. For example, in a model with no covariates, the marginal mean and variance are simple the mean and variance of the simulated counts.

## S8.1 Model 1: Two-level variance-components Poisson model for count responses

Model 1 is a two-level variance-components Poisson model for count responses (Equation 1).The model, written out again for convenience, is as follows.

 (S8.1)

The full results are presented in Table 1.

 The parameter estimates are , and the VPCs and other marginal statistics are presented in Table S8.1 under the column “exact method”. The simulation method for calculating these statistics consists of the following steps:

1. Create a new two-level dataset with schools and students per school
2. Assign each school a random effect value from the estimated distribution
3. Calculate the expected count
4. Simulate the observed counts
5. Calculate the marginal expectation as the sample mean of
6. Calculate the marginal variance as the sample variance of
7. Calculate the school (level-2) component of the marginal variance as the variance of the school means of
8. Calculate the student (level-1) component of the marginal variance as the variance of the student observed count deviated from their school means
9. Calculate the school (level-2) VPC as the ratio of the school component of the marginal variance to the overall marginal variance
10. Calculate the student (level-1) VPC as the ratio of the student component of the marginal variance to the overall marginal variance

where and should be set to large values to minimise Monte Carlo error in the resulting marginal statistics.

Applying this method gives the VPCs (and other marginal statistics) presented under the column “Simulation method” in Table S8.1. As expected, these statistics effectively identical to those based on the exact method confirming that our derivations are correct.

*Table S8.1*. Comparison of the estimated marginal statistics for the two-level variance-components Poisson model (Model 1) based on the exact method and the simulation method.

|  |  |
| --- | --- |
|  | Marginal statistics |
|  | Exact method | Simulation method |
| Marginal expectation | 8.46 | 8.42 |
| Marginal variance | 15.98 | 16.02 |
|  School (level-2) component | 7.52 | 7.61 |
|  Student (level-1) component | 8.46 | 8.84 |
|  School (level-2) VPC | 0.47 | 0.48 |
|  Student (level-1) VPC | 0.53 | 0.52 |

*Note*. Number of simulated schools . Number of simulated students per school . The estimates presented in the “Exact method” column replicate those presented in Table 1 (Model 1).

## S8.2 Model 2: Two-level variance-components negative binomial model for count responses

Model 2 is a two-level variance-components negative binomialmodel for count responses (Equation 9).The model, written out again for convenience, is as follows.

 (S8.2)

The full results are presented in Table 1.

 The parameter estimates are , , and the VPCs and other marginal statistics are presented in Table S8.2 under the column “exact method”. The simulation method for calculating these statistics consists of the following steps:

1. Create a new two-level dataset with schools and students per school
2. Assign each school a random effect value from the estimated distribution
3. Assign each student an exponentiated overdispersion random effect value from the estimated distribution
4. Calculate the expected count
5. Simulate the observed counts
6. Calculate the marginal expectation as the sample mean of
7. Calculate the marginal variance as the sample variance of
8. Calculate the school (level-2) component of the marginal variance as the variance of the school means of
9. Calculate the student (level-1) component of the marginal variance as the variance of the student observed count deviated from their school means
10. Calculate the school (level-2) VPC as the ratio of the school component of the marginal variance to the overall marginal variance
11. Calculate the student (level-1) VPC as the ratio of the student component of the marginal variance to the overall marginal variance

Applying this method gives the VPCs (and other marginal statistics) presented under the column “Simulation method” in Table S8.2. As expected, these statistics are effectively identical to those based on the exact method confirming that our derivations are correct.

*Table S8.2*. Comparison of the estimated marginal statistics for the two-level variance-components negative binomial model (Model 2) based on the exact method and the simulation method.

|  |  |
| --- | --- |
|  | Marginal statistics |
|  | Exact method | Simulation method |
| Marginal expectation | 8.45 | 8.41 |
| Marginal variance | 84.10 | 88.25 |
|  School (level-2) component | 6.95 | 6.92 |
|  Student (level-1) component | 77.15 | 76.33 |
|  School (level-2) VPC | 0.08 | 0.08 |
|  Student (level-1) VPC | 0.92 | 0.92 |

*Note*. Number of simulated schools . Number of simulated students per school . The estimates presented in the “Exact method” column replicate those presented in Table 1 (Model 2).

## S8.3 Model 3: Three-level variance-components negative binomial model for count responses

Model 3 is a three-level variance-components negative binomialmodel for count responses (Equation 17).The model, written out again for convenience, is as follows.

 (S8.3)

The full results are presented in Table 1.

 The parameter estimates are , , , and the VPCs and other marginal statistics are presented in Table S8.3 under the column “exact method”. The simulation method for calculating these statistics consists of the following steps:

1. Create a new three-level dataset with districts, schools per district, and students per school
2. Assign each district a random effect value from the estimated distribution
3. Assign each school a random effect value from the estimated distribution
4. Assign each student an exponentiated overdispersion random effect value from the estimated distribution
5. Calculate the expected count
6. Simulate the observed counts
7. Calculate the marginal expectation as the sample mean of
8. Calculate the marginal variance as the sample variance of
9. Calculate the district (level-3) component of the marginal variance as the variance of the district means of
10. Calculate the school (level-2) component of the marginal variance as the variance of the school means of deviated from their district means
11. Calculate the student (level-1) component of the marginal variance as the variance of the of the student observed count deviated from their school means
12. Calculate the district (level-3) VPC as the ratio of the district component of the marginal variance to the overall marginal variance
13. Calculate the school (level-2) VPC as the ratio of the school component of the marginal variance to the overall marginal variance
14. Calculate the student (level-1) VPC as the ratio of the student component of the marginal variance to the overall marginal variance

Applying this method gives the VPCs (and other marginal statistics) presented under the column “Simulation method” in Table S8.3. As expected, these statistics are effectively identical to those based on the exact method confirming that our derivations are correct.

*Table S8.3*. Comparison of the estimated marginal statistics for the three-level variance-components negative binomial model (Model 3) based on the exact method and the simulation method.

|  |  |
| --- | --- |
|  | Marginal statistics |
|  | Exact method | Simulation method |
| Marginal expectation | 8.44 | 8.53 |
| Marginal variance | 83.79 | 85.32 |
|  District (level-3) component | 0.42 | 0.42 |
|  School (level-2) component | 6.50 | 6.58 |
|  Student (level-1) component | 76.87 | 78.32 |
|  District (level-3) VPC | 0.005 | 0.005 |
|  School (level-2) VPC | 0.08 | 0.08 |
|  Student (level-1) VPC | 0.92 | 0.92 |

*Note*. Number of simulated districts . Number of simulated schools per district . Number of simulated students per school . The estimates presented in the “Exact method” column replicate those presented in Table 1 (Model 3).

## S8.4 Model 4: Two-level random-intercept negative binomial model for count responses

Model 4 is a two-level random-intercept negative binomialmodel for count responses (Equation 17 where we have substituted for ).The model, written out again for convenience, is as follows.

 (S8.2)

The full results are presented in Table 2.

 In models with covariates, the VPCs and marginal statistics vary as a function of these covariates. For simplicity, we focus first on the reference student (a student who takes a zero value for each covariate). The only relevant parameters estimates for calculating the VPC for this student are , , . The marginal statistics calculated using our expressions are presented in Table S8.2 under the column “exact method”. The simulation method for calculating these statistics consists of the following steps:

1. Create a new two-level dataset with schools and students per school
2. Assign each school a random effect value from the estimated distribution
3. Assign each student an exponentiated overdispersion random effect value from the estimated distribution
4. Calculate the expected count
5. Simulate the observed counts
6. Calculate the marginal expectation as the sample mean of
7. Calculate the marginal variance as the sample variance of
8. Calculate the school (level-2) component of the marginal variance as the variance of the school means of
9. Calculate the student (level-1) component of the marginal variance as the variance of the student observed count deviated from their school means
10. Calculate the school (level-2) VPC as the ratio of the school component of the marginal variance to the overall marginal variance
11. Calculate the student (level-1) VPC as the ratio of the student component of the marginal variance to the overall marginal variance

Next, we contrast this reference student to an otherwise equivalent student who is eligible for free school meals ( now takes the value 1, all other covariates continue to be held at 0). Now we must additionally consider . Thus we repeat the above steps, but where we now replace Step 4 above with:

1. Calculate the expected count

Applying this method for these two hypothetical students gives the VPCs (and other marginal statistics) presented under the column “Simulation method” in Table S8.4. As expected, these statistics are effectively identical to those based on the exact method confirming that our derivations are correct.

*Table S8.4*. Comparison of the estimated marginal statistics for the two-level random-intercept negative binomial model (Model 4) based on the exact method and the simulation method.

|  |  |
| --- | --- |
|  | Marginal statistics |
|  | Exact method | Simulation method |
|  | Non-FSM student |
| Marginal expectation | 8.82 | 8.80 |
| Marginal variance | 84.77 | 84.24 |
|  School (level-2) component | 8.45 | 8.45 |
|  Student (level-1) component | 76.32 | 75.79 |
|  School (level-2) VPC | 0.10 | 0.10 |
|  Student (level-1) VPC | 0.90 | 0.90 |
|  | FSM student |
| Marginal expectation | 12.86 | 12.83 |
| Marginal variance | 174.29 | 173.20 |
|  School (level-2) component | 17.96 | 17.94 |
|  Student (level-1) component | 156.33 | 155.26 |
|  School (level-2) VPC | 0.10 | 0.10 |
|  Student (level-1) VPC | 0.90 | 0.90 |

*Note*. Number of simulated schools . Number of simulated students per school .

## S8.5 Model 5: Two-level random-coefficient negative binomial model for count responses

Model 5 is a two-level random-coefficient negative binomialmodel for count responses (Equation 21 where we have substituted for ). Recall that the model simply extends the previous random-intercept model by allowing the coefficient on the binary indicator of free school meal (FSM) eligibility to vary randomly across schools.The model, written out again for convenience, is as follows.

 (S8.2)

where (the binary indicator of FSM eligibility). The full results are presented in Table 2.

 As in the previous random-intercept model, we first focus on the reference student (a student who takes a zero value for each covariate). The only relevant parameters estimates for calculating the VPC for this student are , , . The marginal statistics calculated using our expressions are presented in Table S8.2 under the column “exact method”. The simulation method for calculating these statistics consists of the following steps:

1. Create a new two-level dataset with schools and students per school
2. Assign each school a random effect value from the estimated distribution
3. Assign each student an exponentiated overdispersion random effect value from the estimated distribution
4. Calculate the expected count
5. Simulate the observed counts
6. Calculate the marginal expectation as the sample mean of
7. Calculate the marginal variance as the sample variance of
8. Calculate the school (level-2) component of the marginal variance as the variance of the school means of
9. Calculate the student (level-1) component of the marginal variance as the variance of the student observed count deviated from their school means
10. Calculate the school (level-2) VPC as the ratio of the school component of the marginal variance to the overall marginal variance
11. Calculate the student (level-1) VPC as the ratio of the student component of the marginal variance to the overall marginal variance

Next, we contrast this reference student to an otherwise equivalent student who is eligible for free school meals ( now takes the value 1, all other covariates continue to be held at 0). Now we must additionally consider , , . Thus we repeat the above steps, but where we now replace Steps 2 and 4 above with:

1. Assign each school a random effect value from the estimated distribution

and

1. Calculate the expected count

Applying this method for these two hypothetical students gives the VPCs (and other marginal statistics) presented under the column “Simulation method” in Table S8.5. As expected, these statistics are effectively identical to those based on the exact method confirming that our derivations are correct.

*Table S8.5*. Comparison of the estimated marginal statistics for the two-level random-coefficient negative binomial model (Model 4) based on the exact method and the simulation method.

|  |  |
| --- | --- |
|  | Marginal statistics |
|  | Exact method | Simulation method |
|  | Non-FSM student |
| Marginal expectation | 8.88 | 8.85 |
| Marginal variance | 87.24 | 86.68 |
|  School (level-2) component | 9.70 | 9.71 |
|  Student (level-1) component | 77.54 | 76.97 |
|  School (level-2) VPC | 0.11 | 0.11 |
|  Student (level-1) VPC | 0.89 | 0.89 |
|  | FSM student |
| Marginal expectation | 12.76 | 12.77 |
| Marginal variance | 168.44 | 168.84 |
|  School (level-2) component | 16.59 | 16.79 |
|  Student (level-1) component | 154.85 | 152.05 |
|  School (level-2) VPC | 0.10 | 0.10 |
|  Student (level-1) VPC | 0.90 | 0.90 |

*Note*. Number of simulated schools . Number of simulated students per school .

# S9. Stata syntax and output for student absenteeism application

In this section we present Stata syntax and output to replicate the results for models 1-5 presented in Tables 1 and 2 of the paper. The syntax is annotated to help the reader.

## S9.1 Stata syntax

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Variance partitioning in multilevel count models

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Stata do-file to replicate results presented in the article

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Basics

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Set model results display format to three decimal places

set cformat %9.3f

\* Load the data

use "absence.dta", clear

\* Figure 1: Spikeplot of days absent

spikeplot y

\* Figure 2: Bar chart of district means

graph bar (mean) y, over(district, sort(y) label(nolabel))

\* Figure 2: Bar chart of school means

graph bar (mean) y, over(school, sort(y) label(nolabel))

\* Table S7.1 Covariate distributions, including mean days absent

tabstat y, statistics(count mean) by(quintile) nototal

tabstat y, statistics(count mean) by(season) nototal

tabstat y, statistics(count mean) by(female) nototal

tabstat y, statistics(count mean) by(ethnicity) nototal

tabstat y, statistics(count mean) by(notenglish) nototal

tabstat y, statistics(count mean) by(sen) nototal

tabstat y, statistics(count mean) by(fsm) nototal

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table 1: Model 1: Two-level variance-components Poisson model

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Clear memory

clear all

\* Load the data

use "absence.dta", clear

\* Fit model

mepoisson y || school:, startvalues(iv)

\* Deviance

scalar define deviance = -2\*e(ll)

scalar list deviance

\* Intercept

scalar define beta0 = \_b[\_cons]

scalar list beta0

\* Cluster variance

scalar define sigma2u = \_b[/var(\_cons[school])]

scalar list sigma2u

\* Marginal expectation

scalar define expectation = exp(beta0 + sigma2u/2)

scalar list expectation

\* Marginal variance

scalar define variance = expectation + expectation^2\*(exp(sigma2u) - 1)

scalar list variance

\* Marginal variance: Level-2 component

scalar define variance2 = expectation^2\*(exp(sigma2u) - 1)

scalar list variance2

\* Marginal variance: Level-1 component

scalar define variance1 = expectation

scalar list variance1

\* Level-2 VPC

scalar define vpc2 = variance2/(variance2 + variance1)

scalar list vpc2

\* Level-1 VPC

scalar define vpc1 = variance1/(variance2 + variance1)

scalar list vpc1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table 1: Model 2: Two-level variance-components negative binomial model

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Clear memory

clear all

\* Load the data

use "absence.dta", clear

\* Fit model

menbreg y || school:, startvalues(iv)

\* Deviance

scalar define deviance = -2\*e(ll)

scalar list deviance

\* Intercept

scalar define beta0 = \_b[\_cons]

scalar list beta0

\* Cluster variance

scalar define sigma2u = \_b[/var(\_cons[school])]

scalar list sigma2u

\* Overdispersion parameter

scalar define alpha = exp(\_b[/lnalpha])

scalar list alpha

\* Marginal expectation

scalar define expectation = exp(beta0 + sigma2u/2)

scalar list expectation

\* Marginal variance

scalar define variance = expectation ///

 + expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

scalar list variance

\* Marginal variance: Level-2 component

scalar define variance2 = expectation^2\*(exp(sigma2u) - 1)

scalar list variance2

\* Marginal variance: Level-1 component

scalar define variance1 = expectation + expectation^2\*exp(sigma2u)\*alpha

scalar list variance1

\* Level-2 VPC

scalar define vpc2 = variance2/(variance2 + variance1)

scalar list vpc2

\* Level-1 VPC

scalar define vpc1 = variance1/(variance2 + variance1)

scalar list vpc1

\* Predict cluster random intercept effects

predict u, reffects

\* Keep the cluster identifier and the predicted cluster effect

keep school u

\* Collapse the data down to one observation per cluster

duplicates drop

\* Save the predicted cluster effects

save "model2.dta", replace

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table 1: Model 3: Three-level variance-components negative binomial model

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Clear memory

clear all

\* Load the data

use "absence.dta", clear

\* Fit model

menbreg y || district: || school:, startvalues(iv)

\* Deviance

scalar define deviance = -2\*e(ll)

scalar list deviance

\* Intercept

scalar define beta0 = \_b[\_cons]

scalar list beta0

\* Supercluster variance

scalar define sigma2v = \_b[/var(\_cons[district])]

scalar list sigma2v

\* Cluster variance

scalar define sigma2u = \_b[/var(\_cons[school<district])]

scalar list sigma2u

\* Overdispersion parameter

scalar define alpha = exp(\_b[/lnalpha])

scalar list alpha

\* Marginal expectation

scalar define expectation = exp(beta0 + sigma2v/2 + sigma2u/2)

scalar list expectation

\* Marginal variance

scalar define variance = expectation ///

 + expectation^2\*(exp(sigma2v + sigma2u)\*(1 + alpha) - 1)

scalar list variance

\* Marginal variance: Level-3 component

scalar define variance3 = expectation^2\*(exp(sigma2v) - 1)

scalar list variance3

\* Marginal variance: Level-2 component

scalar define variance2 = expectation^2\*exp(sigma2v)\*(exp(sigma2u) - 1)

scalar list variance2

\* Marginal variance: Level-1 component

scalar define variance1 = expectation ///

 + expectation^2\*exp(sigma2v + sigma2u)\*alpha

scalar list variance1

\* Level-3 VPC

scalar define vpc3 = variance3/(variance3 + variance2 + variance1)

scalar list vpc3

\* Level-2 VPC

scalar define vpc2 = variance2/(variance3 + variance2 + variance1)

scalar list vpc2

\* Level-1 VPC

scalar define vpc1 = variance1/(variance3 + variance2 + variance1)

scalar list vpc1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table 2: Model 4: Two-level random-intercept negative binomial model

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Clear memory

clear all

\* Load the data

use "absence.dta", clear

\* Fit model

menbreg y x || school:, startvalues(iv)

\* Deviance

scalar define deviance = -2\*e(ll)

scalar list deviance

\* Linear predictor

predict xb, xb

\* Cluster variance

scalar define sigma2u = \_b[/var(\_cons[school])]

scalar list sigma2u

\* Overdispersion parameter

scalar define alpha = exp(\_b[/lnalpha])

scalar list alpha

\* Marginal expectation

generate expectation = exp(xb + sigma2u/2)

\* Marginal variance

generate variance = expectation + expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

\* Marginal variance: Level-2 component

generate variance2 = expectation^2\*(exp(sigma2u) - 1)

\* Marginal variance: Level-1 component

generate variance1 = expectation + expectation^2\*exp(sigma2u)\*alpha

\* Level-2 VPC

generate vpc2 = variance2/(variance2 + variance1)

\* Level-1 VPC

generate vpc1 = variance1/(variance2 + variance1)

\* Summarize marginal statistics

summarize expectation variance variance2 variance1 vpc2 vpc1

\* Figure 3: Line plot of Level-2 VPC against the marginal expectation

line vpc2 expectation, sort

\* Figure 3: Spikeplot of marginal expectation

spikeplot expectation

\* Predict cluster random intercept effects

predict u, reffects

\* Keep the cluster identifier and the predicted cluster effect

keep school u

\* Collapse the data down to one observation per cluster

duplicates drop

\* Save the predicted cluster effects

save "model4.dta", replace

\* Load model 2 predicted cluster random effects

use "model2.dta", clear

rename u model2u

\* Merge in model 4 predicted cluster random effects

merge 1:1 school using "model4"

rename u model4u

\* Figure 4: Scatterplot of model 4 vs. model 2 predicted cluster random effects

scatter model4u model2u

correlate model4u model2u

\* Rank the model 2 predicted cluster random effects

sort model2u

generate model2urank = \_n

\* Rank the model 4 predicted cluster random effects

sort model4u

generate model4urank = \_n

\* Figure 4: Scatterplot of ranks of model 4 vs. model 2 predicted effects

scatter model4urank model2urank

correlate model4urank model2urank

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table 2: Model 5: Two-level random-coefficient negative binomial model

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Clear memory

clear all

\* Load the data

use "absence.dta", clear

\* Fit model

menbreg y x ///

 || school: x, covariance(unstructured) ///

 startvalues(iv)

\* Deviance

scalar define deviance = -2\*e(ll)

scalar list deviance

\* Linear predictor

predict xb, xb

\* Cluster intercept variance

scalar define sigma2u0 = \_b[/var(\_cons[school])]

scalar list sigma2u0

\* Cluster slope variance

scalar define sigma2u15 = \_b[/var(fsm[school])]

scalar list sigma2u15

\* Cluster intercept-slope covariance

scalar define sigmau015 = \_b[/cov(fsm[school],\_cons[school])]

scalar list sigmau015

\* Overdispersion parameter

scalar define alpha = exp(\_b[/lnalpha])

scalar list alpha

\* Cluster-level variance function

generate zomegauz = sigma2u0 + 2\*sigmau015\*fsm + sigma2u15\*fsm^2

\* Marginal expectation

generate expectation = exp(xb + zomegauz/2)

\* Marginal variance

generate variance = expectation + expectation^2\*(exp(zomegauz)\*(1 + alpha) - 1)

\* Marginal variance: Level-2 component

generate variance2 = expectation^2\*(exp(zomegauz) - 1)

\* Marginal variance: Level-1 component

generate variance1 = expectation + expectation^2\*exp(zomegauz)\*alpha

\* Level-2 VPC

generate vpc2 = variance2/(variance2 + variance1)

\* Level-1 VPC

generate vpc1 = variance1/(variance2 + variance1)

\* Summarize marginal statistics

summarize expectation variance variance2 variance1 vpc2 vpc1

\* Figure 5: Line plot of Level-2 VPC against the marginal expectation by FSM status

line vpc2 expectation, sort by(fsm)

\* Figure 5: Spikeplot of marginal expectation by FSM status

spikeplot expectation, by(fsm)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

exit

## S9.2 Stata output

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Variance partitioning in multilevel count models

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Stata do-file to replicate results presented in the article

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Basics

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Set model results display format to three decimal places

. set cformat %9.3f

.

. \* Load the data

. use "absence.dta", clear

.

. \* Figure 1: Spikeplot of days absent

. spikeplot y

.

. \* Figure 2: Bar chart of district means

. graph bar (mean) y, over(district, sort(y) label(nolabel))

.

. \* Figure 2: Bar chart of school means

. graph bar (mean) y, over(school, sort(y) label(nolabel))

.

. \* Table S7.1 Covariate distributions, including mean days absent

. tabstat y, statistics(count mean) by(quintile) nototal

Summary for variables: y

 by categories of: quintile (5 quantiles of x)

quintile | N mean

---------+--------------------

 1 | 14366 10.05972

 2 | 13288 8.992023

 3 | 12703 8.334567

 4 | 16000 7.421375

 5 | 10598 7.029251

------------------------------

. tabstat y, statistics(count mean) by(season) nototal

Summary for variables: y

 by categories of: season

season | N mean

-------+--------------------

Summer | 17032 8.120538

Spring | 16495 8.225523

Winter | 16551 8.540995

Autumn | 16877 8.755407

----------------------------

. tabstat y, statistics(count mean) by(female) nototal

Summary for variables: y

 by categories of: female (Female)

 female | N mean

---------+--------------------

 0 | 33628 8.025009

 1 | 33327 8.799202

------------------------------

. tabstat y, statistics(count mean) by(ethnicity) nototal

Summary for variables: y

 by categories of: ethnicity (Ethnicity)

ethnicity | N mean

----------+--------------------

 White | 27803 9.608711

 Mixed | 6181 9.639541

 Asian | 13914 7.072229

 Black | 14409 7.013672

 Other | 4648 7.943201

-------------------------------

. tabstat y, statistics(count mean) by(notenglish) nototal

Summary for variables: y

 by categories of: notenglish (English as an additional language)

notenglish | N mean

-----------+--------------------

 0 | 40529 9.231439

 1 | 26426 7.151101

--------------------------------

. tabstat y, statistics(count mean) by(sen) nototal

Summary for variables: y

 by categories of: sen (SEN)

 sen | N mean

---------+--------------------

 0 | 57157 7.892612

 1 | 9798 11.4307

------------------------------

. tabstat y, statistics(count mean) by(fsm) nototal

Summary for variables: y

 by categories of: fsm (FSM)

 fsm | N mean

---------+--------------------

 0 | 41606 7.311253

 1 | 25349 10.21437

------------------------------

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Table 1: Model 1: Two-level variance-components Poisson model

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Clear memory

. clear all

.

. \* Load the data

. use "absence.dta", clear

.

. \* Fit model

. mepoisson y || school:, startvalues(iv)

Fitting fixed-effects model:

Iteration 0: log likelihood = -761235.06

Iteration 1: log likelihood = -421985.3

Iteration 2: log likelihood = -419174.27

Iteration 3: log likelihood = -419168.08

Iteration 4: log likelihood = -419168.08

Refining starting values:

Grid node 0: log likelihood = -393452.09

Refining starting values (unscaled likelihoods):

Grid node 0: log likelihood = -393459.83

Fitting full model:

Iteration 0: log likelihood = -393459.83 (not concave)

Iteration 1: log likelihood = -393002.29

Iteration 2: log likelihood = -392866.51

Iteration 3: log likelihood = -392768.39

Iteration 4: log likelihood = -392698.35

Iteration 5: log likelihood = -392649.56

Iteration 6: log likelihood = -392616.78

Iteration 7: log likelihood = -392595.86

Iteration 8: log likelihood = -392583.38

Iteration 9: log likelihood = -392572.82

Iteration 10: log likelihood = -392571.14

Iteration 11: log likelihood = -392571.07

Iteration 12: log likelihood = -392571.07

Mixed-effects Poisson regression Number of obs = 66,955

Group variable: school Number of groups = 434

 Obs per group:

 min = 31

 avg = 154.3

 max = 315

Integration method: mvaghermite Integration pts. = 7

 Wald chi2(0) = .

Log likelihood = -392571.07 Prob > chi2 = .

------------------------------------------------------------------------------

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 \_cons | 2.085 0.015 136.72 0.000 2.055 2.115

-------------+----------------------------------------------------------------

school |

 var(\_cons)| 0.100 0.007 0.087 0.114

------------------------------------------------------------------------------

LR test vs. Poisson model: chibar2(01) = 53194.02 Prob >= chibar2 = 0.0000

.

. \* Deviance

. scalar define deviance = -2\*e(ll)

. scalar list deviance

 deviance = 785142.14

.

. \* Intercept

. scalar define beta0 = \_b[\_cons]

. scalar list beta0

 beta0 = 2.0852543

.

. \* Cluster variance

. scalar define sigma2u = \_b[/var(\_cons[school])]

. scalar list sigma2u

 sigma2u = .09998112

.

. \* Marginal expectation

. scalar define expectation = exp(beta0 + sigma2u/2)

. scalar list expectation

expectation = 8.4591173

.

. \* Marginal variance

. scalar define variance = expectation + expectation^2\*(exp(sigma2u) - 1)

. scalar list variance

 variance = 15.983304

.

. \* Marginal variance: Level-2 component

. scalar define variance2 = expectation^2\*(exp(sigma2u) - 1)

. scalar list variance2

 variance2 = 7.5241871

.

. \* Marginal variance: Level-1 component

. scalar define variance1 = expectation

. scalar list variance1

 variance1 = 8.4591173

.

. \* Level-2 VPC

. scalar define vpc2 = variance2/(variance2 + variance1)

. scalar list vpc2

 vpc2 = .47075291

.

. \* Level-1 VPC

. scalar define vpc1 = variance1/(variance2 + variance1)

. scalar list vpc1

 vpc1 = .52924709

.

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Table 1: Model 2: Two-level variance-components negative binomial model

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Clear memory

. clear all

.

. \* Load the data

. use "absence.dta", clear

.

. \* Fit model

. menbreg y || school:, startvalues(iv)

Fitting fixed-effects model:

Iteration 0: log likelihood = -216299.9

Iteration 1: log likelihood = -213455.56

Iteration 2: log likelihood = -213346.63

Iteration 3: log likelihood = -213346.62

Refining starting values:

Grid node 0: log likelihood = -211608.67

Fitting full model:

Iteration 0: log likelihood = -211608.67

Iteration 1: log likelihood = -211226.32

Iteration 2: log likelihood = -211072.56

Iteration 3: log likelihood = -211029.19

Iteration 4: log likelihood = -211023.25

Iteration 5: log likelihood = -211023.09

Iteration 6: log likelihood = -211023.09

Mixed-effects nbinomial regression Number of obs = 66,955

Overdispersion: mean

Group variable: school Number of groups = 434

 Obs per group:

 min = 31

 avg = 154.3

 max = 315

Integration method: mvaghermite Integration pts. = 7

 Wald chi2(0) = .

Log likelihood = -211023.09 Prob > chi2 = .

------------------------------------------------------------------------------

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 \_cons | 2.088 0.015 137.30 0.000 2.058 2.118

-------------+----------------------------------------------------------------

 /lnalpha | -0.132 0.006 -21.24 0.000 -0.144 -0.120

-------------+----------------------------------------------------------------

school |

 var(\_cons)| 0.093 0.007 0.080 0.107

------------------------------------------------------------------------------

LR test vs. nbinomial model: chibar2(01) = 4647.07 Prob >= chibar2 = 0.0000

.

. \* Deviance

. scalar define deviance = -2\*e(ll)

. scalar list deviance

 deviance = 422046.17

.

. \* Intercept

. scalar define beta0 = \_b[\_cons]

. scalar list beta0

 beta0 = 2.0878598

.

. \* Cluster variance

. scalar define sigma2u = \_b[/var(\_cons[school])]

. scalar list sigma2u

 sigma2u = .09284542

.

. \* Overdispersion parameter

. scalar define alpha = exp(\_b[/lnalpha])

. scalar list alpha

 alpha = .876623

.

. \* Marginal expectation

. scalar define expectation = exp(beta0 + sigma2u/2)

. scalar list expectation

expectation = 8.4509804

.

. \* Marginal variance

. scalar define variance = expectation ///

> + expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

. scalar list variance

 variance = 84.098316

.

. \* Marginal variance: Level-2 component

. scalar define variance2 = expectation^2\*(exp(sigma2u) - 1)

. scalar list variance2

 variance2 = 6.9485112

.

. \* Marginal variance: Level-1 component

. scalar define variance1 = expectation + expectation^2\*exp(sigma2u)\*alpha

. scalar list variance1

 variance1 = 77.149805

.

. \* Level-2 VPC

. scalar define vpc2 = variance2/(variance2 + variance1)

. scalar list vpc2

 vpc2 = .08262367

.

. \* Level-1 VPC

. scalar define vpc1 = variance1/(variance2 + variance1)

. scalar list vpc1

 vpc1 = .91737633

.

. \* Predict cluster random intercept effects

. predict u, reffects

(calculating posterior means of random effects)

(using 7 quadrature points)

.

. \* Keep the cluster identifier and the predicted cluster effect

. keep school u

.

. \* Collapse the data down to one observation per cluster

. duplicates drop

Duplicates in terms of all variables

(66,521 observations deleted)

.

. \* Save the predicted cluster effects

. save "model2.dta", replace

file model2.dta saved

.

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Table 1: Model 3: Three-level variance-components negative binomial model

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Clear memory

. clear all

.

. \* Load the data

. use "absence.dta", clear

.

. \* Fit model

. menbreg y || district: || school:, startvalues(iv)

Fitting fixed-effects model:

Iteration 0: log likelihood = -216299.9

Iteration 1: log likelihood = -213455.56

Iteration 2: log likelihood = -213346.63

Iteration 3: log likelihood = -213346.62

Refining starting values:

Grid node 0: log likelihood = -211551.43

Fitting full model:

Iteration 0: log likelihood = -211551.43

Iteration 1: log likelihood = -211198.25

Iteration 2: log likelihood = -211061.47

Iteration 3: log likelihood = -211024.08

Iteration 4: log likelihood = -211019.52

Iteration 5: log likelihood = -211019.44

Iteration 6: log likelihood = -211019.44

Mixed-effects nbinomial regression Number of obs = 66,955

Overdispersion: mean

-------------------------------------------------------------

 | No. of Observations per Group

 Group Variable | Groups Minimum Average Maximum

----------------+--------------------------------------------

 district | 32 665 2,092.3 3,182

 school | 434 31 154.3 315

-------------------------------------------------------------

Integration method: mvaghermite Integration pts. = 7

 Wald chi2(0) = .

Log likelihood = -211019.44 Prob > chi2 = .

---------------------------------------------------------------------------------

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

----------------+----------------------------------------------------------------

 \_cons | 2.086 0.020 103.17 0.000 2.046 2.126

----------------+----------------------------------------------------------------

 /lnalpha | -0.132 0.006 -21.24 0.000 -0.144 -0.120

----------------+----------------------------------------------------------------

district |

 var(\_cons)| 0.006 0.003 0.002 0.017

----------------+----------------------------------------------------------------

district>school |

 var(\_cons)| 0.087 0.007 0.075 0.101

---------------------------------------------------------------------------------

LR test vs. nbinomial model: chi2(2) = 4654.36 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

.

. \* Deviance

. scalar define deviance = -2\*e(ll)

. scalar list deviance

 deviance = 422038.88

.

. \* Intercept

. scalar define beta0 = \_b[\_cons]

. scalar list beta0

 beta0 = 2.0860497

.

. \* Supercluster variance

. scalar define sigma2v = \_b[/var(\_cons[district])]

. scalar list sigma2v

 sigma2v = .00582819

.

. \* Cluster variance

. scalar define sigma2u = \_b[/var(\_cons[school<district])]

. scalar list sigma2u

 sigma2u = .08692447

.

. \* Overdispersion parameter

. scalar define alpha = exp(\_b[/lnalpha])

. scalar list alpha

 alpha = .8766216

.

. \* Marginal expectation

. scalar define expectation = exp(beta0 + sigma2v/2 + sigma2u/2)

. scalar list expectation

expectation = 8.4353062

.

. \* Marginal variance

. scalar define variance = expectation ///

> + expectation^2\*(exp(sigma2v + sigma2u)\*(1 + alpha) - 1)

. scalar list variance

 variance = 83.788592

.

. \* Marginal variance: Level-3 component

. scalar define variance3 = expectation^2\*(exp(sigma2v) - 1)

. scalar list variance3

 variance3 = .41591198

.

. \* Marginal variance: Level-2 component

. scalar define variance2 = expectation^2\*exp(sigma2v)\*(exp(sigma2u) - 1)

. scalar list variance2

 variance2 = 6.4996057

.

. \* Marginal variance: Level-1 component

. scalar define variance1 = expectation ///

> + expectation^2\*exp(sigma2v + sigma2u)\*alpha

. scalar list variance1

 variance1 = 76.873075

.

. \* Level-3 VPC

. scalar define vpc3 = variance3/(variance3 + variance2 + variance1)

. scalar list vpc3

 vpc3 = .00496383

.

. \* Level-2 VPC

. scalar define vpc2 = variance2/(variance3 + variance2 + variance1)

. scalar list vpc2

 vpc2 = .07757149

.

. \* Level-1 VPC

. scalar define vpc1 = variance1/(variance3 + variance2 + variance1)

. scalar list vpc1

 vpc1 = .91746469

.

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Table 2: Model 4: Two-level random-intercept negative binomial model

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Clear memory

. clear all

.

. \* Load the data

. use "absence.dta", clear

.

. \* Fit model

. menbreg y quintile2 quintile3 quintile4 quintile5 spring winter autumn ///

> female mixed asian black other notenglish sen fsm || school:, startvalues(iv)

Fitting fixed-effects model:

Iteration 0: log likelihood = -213128.96

Iteration 1: log likelihood = -210651.83

Iteration 2: log likelihood = -210545.23

Iteration 3: log likelihood = -210545.17

Iteration 4: log likelihood = -210545.17

Refining starting values:

Grid node 0: log likelihood = -208325.29

Fitting full model:

Iteration 0: log likelihood = -208325.29

Iteration 1: log likelihood = -207903.92

Iteration 2: log likelihood = -207750.23

Iteration 3: log likelihood = -207720.21

Iteration 4: log likelihood = -207718.81

Iteration 5: log likelihood = -207718.8

Mixed-effects nbinomial regression Number of obs = 66,955

Overdispersion: mean

Group variable: school Number of groups = 434

 Obs per group:

 min = 31

 avg = 154.3

 max = 315

Integration method: mvaghermite Integration pts. = 7

 Wald chi2(15) = 6724.95

Log likelihood = -207718.8 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 quintile2 | -0.051 0.012 -4.34 0.000 -0.074 -0.028

 quintile3 | -0.118 0.012 -9.71 0.000 -0.142 -0.094

 quintile4 | -0.222 0.012 -18.77 0.000 -0.245 -0.199

 quintile5 | -0.330 0.014 -23.60 0.000 -0.358 -0.303

 spring | 0.026 0.011 2.50 0.013 0.006 0.047

 winter | 0.077 0.011 7.33 0.000 0.057 0.098

 autumn | 0.112 0.011 10.59 0.000 0.091 0.132

 female | 0.122 0.009 13.83 0.000 0.104 0.139

 mixed | -0.073 0.014 -5.28 0.000 -0.100 -0.046

 asian | -0.194 0.013 -15.25 0.000 -0.219 -0.169

 black | -0.422 0.011 -38.32 0.000 -0.444 -0.400

 other | -0.194 0.017 -11.49 0.000 -0.227 -0.161

 notenglish | -0.244 0.009 -26.01 0.000 -0.262 -0.226

 sen | 0.267 0.011 23.42 0.000 0.245 0.290

 fsm | 0.377 0.008 44.91 0.000 0.360 0.393

 \_cons | 2.126 0.021 103.07 0.000 2.086 2.167

-------------+----------------------------------------------------------------

 /lnalpha | -0.246 0.006 -38.41 0.000 -0.259 -0.234

-------------+----------------------------------------------------------------

school |

 var(\_cons)| 0.103 0.007 0.089 0.118

------------------------------------------------------------------------------

LR test vs. nbinomial model: chibar2(01) = 5652.76 Prob >= chibar2 = 0.0000

.

. \* Deviance

. scalar define deviance = -2\*e(ll)

. scalar list deviance

 deviance = 415437.59

.

. \* Linear predictor

. predict xb, xb

.

. \* Cluster variance

. scalar define sigma2u = \_b[/var(\_cons[school])]

. scalar list sigma2u

 sigma2u = .10259547

.

. \* Overdispersion parameter

. scalar define alpha = exp(\_b[/lnalpha])

. scalar list alpha

 alpha = .78186163

.

. \* Marginal expectation

. generate expectation = exp(xb + sigma2u/2)

.

. \* Marginal variance

. generate variance = expectation + expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

.

. \* Marginal variance: Level-2 component

. generate variance2 = expectation^2\*(exp(sigma2u) - 1)

.

. \* Marginal variance: Level-1 component

. generate variance1 = expectation + expectation^2\*exp(sigma2u)\*alpha

.

. \* Level-2 VPC

. generate vpc2 = variance2/(variance2 + variance1)

.

. \* Level-1 VPC

. generate vpc1 = variance1/(variance2 + variance1)

.

. \* Summarize marginal statistics

. summarize expectation variance variance2 variance1 vpc2 vpc1

 Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

 expectation | 66,955 8.502207 2.885131 3.258426 21.21049

 variance | 66,955 87.04828 61.42828 13.60374 459.5689

 variance2 | 66,955 8.709502 6.498338 1.14713 48.60693

 variance1 | 66,955 78.33878 54.9303 12.45661 410.9619

 vpc2 | 66,955 .0979414 .0035452 .0843246 .1057664

-------------+---------------------------------------------------------

 vpc1 | 66,955 .9020586 .0035452 .8942336 .9156754

.

. \* Figure 3: Line plot of Level-2 VPC against the marginal expectation

. line vpc2 expectation, sort

.

. \* Figure 3: Spikeplot of marginal expectation

. spikeplot expectation

.

. \* Predict cluster random intercept effects

. predict u, reffects

(calculating posterior means of random effects)

(using 7 quadrature points)

.

. \* Keep the cluster identifier and the predicted cluster effect

. keep school u

.

. \* Collapse the data down to one observation per cluster

. duplicates drop

Duplicates in terms of all variables

(66,521 observations deleted)

.

. \* Save the predicted cluster effects

. save "model4.dta", replace

file model4.dta saved

.

. \* Load model 2 predicted cluster random effects

. use "model2.dta", clear

. rename u model2u

.

. \* Merge in model 4 predicted cluster random effects

. merge 1:1 school using "model4"

(label urn already defined)

 Result # of obs.

 -----------------------------------------

 not matched 0

 matched 434 (\_merge==3)

 -----------------------------------------

. rename u model4u

.

. \* Figure 4: Scatterplot of model 4 vs. model 2 predicted cluster random effects

. scatter model4u model2u

. correlate model4u model2u

(obs=434)

 | model4u model2u

-------------+------------------

 model4u | 1.0000

 model2u | 0.9184 1.0000

.

. \* Rank the model 2 predicted cluster random effects

. sort model2u

. generate model2urank = \_n

.

. \* Rank the model 4 predicted cluster random effects

. sort model4u

. generate model4urank = \_n

.

. \* Figure 4: Scatterplot of ranks of model 4 vs. model 2 predicted effects

. scatter model4urank model2urank

. correlate model4urank model2urank

(obs=434)

 | model4~k model2~k

-------------+------------------

 model4urank | 1.0000

 model2urank | 0.9091 1.0000

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \* Table 2: Model 5: Two-level random-coefficient negative binomial model

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \* Clear memory

. clear all

.

. \* Load the data

. use "absence.dta", clear

.

. \* Fit model

. menbreg y quintile2 quintile3 quintile4 quintile5 spring winter autumn ///

> female mixed asian black other notenglish sen fsm ///

> || school: fsm, covariance(unstructured) ///

> startvalues(iv)

Fitting fixed-effects model:

Iteration 0: log likelihood = -213128.96

Iteration 1: log likelihood = -210651.83

Iteration 2: log likelihood = -210545.23

Iteration 3: log likelihood = -210545.17

Iteration 4: log likelihood = -210545.17

Refining starting values:

Grid node 0: log likelihood = -208637.83

Fitting full model:

Iteration 0: log likelihood = -208637.83 (not concave)

Iteration 1: log likelihood = -208461.37 (not concave)

Iteration 2: log likelihood = -208287.8 (not concave)

Iteration 3: log likelihood = -208213.99

Iteration 4: log likelihood = -208042.22

Iteration 5: log likelihood = -207760.47

Iteration 6: log likelihood = -207661.3

Iteration 7: log likelihood = -207635.69

Iteration 8: log likelihood = -207633.84

Iteration 9: log likelihood = -207633.82

Iteration 10: log likelihood = -207633.82

Mixed-effects nbinomial regression Number of obs = 66,955

Overdispersion: mean

Group variable: school Number of groups = 434

 Obs per group:

 min = 31

 avg = 154.3

 max = 315

Integration method: mvaghermite Integration pts. = 7

 Wald chi2(15) = 5401.93

Log likelihood = -207633.82 Prob > chi2 = 0.0000

-------------------------------------------------------------------------------

 y | Coef. Std. Err. z P>|z| [95% Conf. Interval]

--------------+----------------------------------------------------------------

 quintile2 | -0.048 0.012 -4.09 0.000 -0.072 -0.025

 quintile3 | -0.116 0.012 -9.52 0.000 -0.140 -0.092

 quintile4 | -0.219 0.012 -18.49 0.000 -0.242 -0.196

 quintile5 | -0.326 0.014 -23.25 0.000 -0.353 -0.299

 spring | 0.026 0.011 2.47 0.014 0.005 0.047

 winter | 0.078 0.011 7.37 0.000 0.057 0.099

 autumn | 0.112 0.011 10.64 0.000 0.091 0.133

 female | 0.122 0.009 13.86 0.000 0.105 0.139

 mixed | -0.074 0.014 -5.36 0.000 -0.101 -0.047

 asian | -0.198 0.013 -15.49 0.000 -0.223 -0.173

 black | -0.421 0.011 -38.14 0.000 -0.443 -0.400

 other | -0.195 0.017 -11.56 0.000 -0.228 -0.162

 notenglish | -0.242 0.009 -25.81 0.000 -0.261 -0.224

 sen | 0.267 0.011 23.35 0.000 0.244 0.289

 fsm | 0.372 0.013 29.56 0.000 0.347 0.397

 \_cons | 2.126 0.021 99.26 0.000 2.084 2.168

--------------+----------------------------------------------------------------

 /lnalpha | -0.255 0.006 -39.51 0.000 -0.267 -0.242

--------------+----------------------------------------------------------------

school |

 var(fsm)| 0.035 0.005 0.027 0.046

 var(\_cons)| 0.116 0.009 0.100 0.135

--------------+----------------------------------------------------------------

school |

cov(\_cons,fsm)| -0.027 0.005 -5.23 0.000 -0.037 -0.017

-------------------------------------------------------------------------------

LR test vs. nbinomial model: chi2(3) = 5822.71 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

.

. \* Deviance

. scalar define deviance = -2\*e(ll)

. scalar list deviance

 deviance = 415267.64

.

. \* Linear predictor

. predict xb, xb

.

. \* Cluster intercept variance

. scalar define sigma2u0 = \_b[/var(\_cons[school])]

. scalar list sigma2u0

 sigma2u0 = .11603906

.

. \* Cluster slope variance

. scalar define sigma2u15 = \_b[/var(fsm[school])]

. scalar list sigma2u15

 sigma2u15 = .03503611

.

. \* Cluster intercept-slope covariance

. scalar define sigmau015 = \_b[/cov(fsm[school],\_cons[school])]

. scalar list sigmau015

 sigmau015 = -.02662019

.

. \* Overdispersion parameter

. scalar define alpha = exp(\_b[/lnalpha])

. scalar list alpha

 alpha = .77526043

.

. \* Cluster-level variance function

. generate zomegauz = sigma2u0 + 2\*sigmau015\*fsm + sigma2u15\*fsm^2

.

. \* Marginal expectation

. generate expectation = exp(xb + zomegauz/2)

.

. \* Marginal variance

. generate variance = expectation + expectation^2\*(exp(zomegauz)\*(1 + alpha) - 1)

.

. \* Marginal variance: Level-2 component

. generate variance2 = expectation^2\*(exp(zomegauz) - 1)

.

. \* Marginal variance: Level-1 component

. generate variance1 = expectation + expectation^2\*exp(zomegauz)\*alpha

.

. \* Level-2 VPC

. generate vpc2 = variance2/(variance2 + variance1)

.

. \* Level-1 VPC

. generate vpc1 = variance1/(variance2 + variance1)

.

. \* Summarize marginal statistics

. summarize expectation variance variance2 variance1 vpc2 vpc1

 Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

 expectation | 66,955 8.524475 2.850676 3.302331 21.06587

 variance | 66,955 87.20277 59.4723 14.13889 446.0752

 variance2 | 66,955 9.036841 6.074755 1.341797 45.61101

 variance1 | 66,955 78.16592 53.43194 12.79709 400.4642

 vpc2 | 66,955 .1038891 .0063813 .0879661 .1158658

-------------+---------------------------------------------------------

 vpc1 | 66,955 .8961109 .0063813 .8841342 .9120339

.

. \* Figure 5: Line plot of Level-2 VPC against the marginal expectation by FSM status

. line vpc2 expectation, sort by(fsm)

.

. \* Figure 5: Spikeplot of marginal expectation by FSM status

. spikeplot expectation, by(fsm)

.

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. exit

# S10. R code for student absenteeism application

In this section we present R code to replicate the results for models 1-5 presented in Tables 1 and 2 of the paper. The code is annotated to help the reader.

## 10.1 R code

################################################################################

# Variance partitioning in multilevel count models

################################################################################

# R script file to replicate results presented in the article

################################################################################

################################################################################

# Basics

################################################################################

# Load library

#library(lme4)

library(haven)

library(ggplot2)

library(glmmTMB)

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Figure 1: Spikeplot of days absent

ggplot(data = absence, mapping = aes(x = y)) + geom\_bar()

# Figure 2: Bar chart of district means

ggplot(data = absence, mapping = aes(x = reorder(district, y), y = y)) +

 geom\_bar(stat = "summary", fun.y = "mean")

# Figure 2: Bar chart of school means

ggplot(data = absence, mapping = aes(x = reorder(school, y), y = y)) +

 geom\_bar(stat = "summary", fun.y = "mean")

# Table S7.1 Covariate distributions, including mean days absent

by(absence$y, absence$quintile, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$season, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$female, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$ethnicity, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$notenglish, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$sen, function(x) c(N = length(x), mean = mean(x)))

by(absence$y, absence$fsm, function(x) c(N = length(x), mean = mean(x)))

################################################################################

# Table 1: Model 1: Two-level variance-components Poisson model

################################################################################

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Fit model

fm1 <- glmmTMB(y ~ 1 + (1|school), data = absence, family = poisson)

summary(fm1)

# Intercept

str(summary(fm1))

beta0 <- summary(fm1)$coefficients$cond[1,1]

beta0

# Cluster variance

str(summary(fm1))

sigma2u <- summary(fm1)$varcor$cond$school[1,1]

sigma2u

# Marginal expectation

expectation <- exp(beta0 + sigma2u/2)

expectation

# Marginal variance

variance <- expectation + expectation^2\*(exp(sigma2u) - 1)

variance

# Marginal variance: Level-2 component

variance2 <- expectation^2\*(exp(sigma2u) - 1)

variance2

# Marginal variance: Level-1 component

variance1 <- expectation

variance1

# Level-2 VPC

vpc2 <- variance2/(variance2 + variance1)

vpc2

# Level-1 VPC

vpc1 <- variance1/(variance2 + variance1)

vpc1

################################################################################

# Table 1: Model 2: Two-level variance-components negative binomial model

################################################################################

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Fit model

fm2 <- glmmTMB(y ~ 1 + (1|school), data = absence, family = nbinom2)

summary(fm2)

# Intercept

str(summary(fm2))

beta0 <- summary(fm2)$coefficients$cond[1,1]

beta0

# Cluster variance

str(summary(fm2))

sigma2u <- summary(fm2)$varcor$cond$school[1,1]

sigma2u

# Overdispersion parameter

str(summary(fm2))

alpha <- 1/(summary(fm2)$sigma)

alpha

# Marginal expectation

expectation <- exp(beta0 + sigma2u/2)

expectation

# Marginal variance

variance <- expectation + expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

variance

# Marginal variance: Level-2 component

variance2 <- expectation^2\*(exp(sigma2u) - 1)

variance2

# Marginal variance: Level-1 component

variance1 <- expectation + expectation^2\*exp(sigma2u)\*alpha

variance1

# Level-2 VPC

vpc2 <- variance2/(variance2 + variance1)

vpc2

# Level-1 VPC

vpc1 <- variance1/(variance2 + variance1)

vpc1

# Predict cluster random intercept effects

fm2u <- ranef(fm2)

fm2u

################################################################################

# Table 1: Model 3: Three-level variance-components negative binomial model

################################################################################

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Fit model

fm3 <- glmmTMB(y ~ 1 + (1|district) + (1|school),

 data = absence, family = nbinom2)

summary(fm3)

# Intercept

str(summary(fm3))

beta0 <- summary(fm3)$coefficients$cond[1,1]

beta0

# Supercluster variance

str(summary(fm3))

sigma2v <- summary(fm3)$varcor$cond$district[1,1]

sigma2v

# Cluster variance

str(summary(fm3))

sigma2u <- summary(fm3)$varcor$cond$school[1,1]

sigma2u

# Overdispersion parameter

str(summary(fm3))

alpha <- 1/(summary(fm3)$sigma)

alpha

# Marginal expectation

expectation <- exp(beta0 + sigma2v/2 + sigma2u/2)

expectation

# Marginal variance

variance <- expectation +

 expectation^2\*(exp(sigma2v + sigma2u)\*(1 + alpha) - 1)

variance

# Marginal variance: Level-3 component

variance3 < (expectation^2\*(exp(sigma2v) - 1)

variance3

# Marginal variance: Level-2 component

variance2 <- expectation^2\*exp(sigma2v)\*(exp(sigma2u) - 1)

variance2

# Marginal variance: Level-1 component

variance1 <- expectation + expectation^2\*exp(sigma2v + sigma2u)\*alpha

variance1

# Level-3 VPC

vpc3 <- variance3/(variance3 + variance2 + variance1)

vpc3

# Level-2 VPC

vpc2 <- variance2/(variance3 + variance2 + variance1)

vpc2

# Level-1 VPC

vpc1 <- variance1/(variance3 + variance2 + variance1)

vpc1

################################################################################

# Table 2: Model 4: Two-level random-intercept negative binomial model

################################################################################

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Fit model

fm4 <- glmmTMB(y ~ 1 + fsm + (1|school), data = absence, family = nbinom2)

summary(fm4)

# Linear predictor

absence$xb <- predict(fm4)

head(absence)

# Cluster variance

str(summary(fm4))

sigma2u <- summary(fm4)$varcor$cond$school[1,1]

sigma2u

# Overdispersion parameter

str(summary(fm4))

alpha <- 1/(summary(fm4)$sigma)

alpha

# Marginal expectation

absence$expectation <- exp(absence$xb + sigma2u/2)

head(absence)

# Marginal variance

absence$variance <- absence$expectation +

 absence$expectation^2\*(exp(sigma2u)\*(1 + alpha) - 1)

head(absence)

# Marginal variance: Level-2 component

absence$variance2 <- absence$expectation^2\*(exp(sigma2u) - 1)

head(absence)

# Marginal variance: Level-1 component

absence$variance1 <- absence$expectation +

 absence$expectation^2\*exp(sigma2u)\*alpha

head(absence)

# Level-2 VPC

absence$vpc2 <- absence$variance2/(absence$variance2 + absence$variance1)

head(absence)

# Level-1 VPC

absence$vpc1 <- absence$variance1/(absence$variance2 + absence$variance1)

head(absence)

# Summarize marginal statistics

colnames(absence)

sapply(absence[7:12], mean)

# Figure 3: Line plot of Level-2 VPC against the marginal expectation

ggplot(data = absence, mapping = aes(x = expectation, y = vpc2)) + geom\_line()

# Figure 3: Spikeplot of marginal expectation

ggplot(data = absence, mapping = aes(x = expectation)) +

 geom\_histogram(binwidth=1)

# Predict cluster random intercept effects

fm4u <- ranef(fm4)

str(fm4u)

head(fm4u$cond$school)

# Figure 4: Scatterplot of model 4 vs. model 2 predicted cluster random effects

fm4vsfm2 <- cbind(fm2u$cond$school,fm4u$cond$school)

colnames(fm4vsfm2)

colnames(fm4vsfm2) <- c("fm2u", "fm4u")

colnames(fm4vsfm2)

head(fm4vsfm2)

ggplot(data = fm4vsfm2, mapping = aes(x = fm2u, y = fm4u)) + geom\_point()

cor(fm4vsfm2)

# Rank the model 2 predicted cluster random effects

fm4vsfm2$fm2urank <- rank(fm4vsfm2$fm2u)

head(fm4vsfm2)

# Rank the model 4 predicted cluster random effects

fm4vsfm2$fm4urank <- rank(fm4vsfm2$fm4u)

head(fm4vsfm2)

# Figure 4: Scatterplot of ranks of model 4 vs. model 2 predicted effects

ggplot(data = fm4vsfm2, mapping = aes(x = fm2urank, y = fm4urank)) +

 geom\_point()

colnames(fm4vsfm2)

cor(fm4vsfm2[,3:4])

################################################################################

# Table 2: Model 5: Two-level random-coefficient negative binomial model

################################################################################

# Load the data

absence <- read\_dta("absence.dta")

head(absence)

# Fit model

fm5 <- glmmTMB(y ~ 1 + fsm + (1 + fsm |school), data = absence, family = nbinom2)

summary(fm5)

# Linear predictor

absence$xb <- predict(fm5)

head(absence)

# Cluster intercept variance

str(summary(fm5))

sigma2u0 <- summary(fm5)$varcor$cond$school[1,1]

sigma2u0

# Cluster slope variance

str(summary(fm5))

sigma2u15 <- summary(fm5)$varcor$cond$school[2,2]

sigma2u15

# Cluster intercept-slope covariance

str(summary(fm5))

sigmau015 <- summary(fm5)$varcor$cond$school[1,2]

sigmau015

# Overdispersion parameter

str(summary(fm5))

alpha <- 1/(summary(fm5)$sigma)

alpha

# Cluster-level variance function

absence$zomegauz = sigma2u0 + 2\*sigmau015\*absence$fsm + sigma2u15\*absence$fsm^2

head(absence)

# Marginal expectation

absence$expectation = exp(absence$xb + absence$zomegauz/2)

head(absence)

# Marginal variance

absence$variance = absence$expectation +

 absence$expectation^2\*(exp(absence$zomegauz)\*(1 + alpha) - 1)

head(absence)

# Marginal variance: Level-2 component

absence$variance2 = absence$expectation^2\*(exp(absence$zomegauz) - 1)

head(absence)

# Marginal variance: Level-1 component

absence$variance1 = absence$expectation +

 absence$expectation^2\*exp(absence$zomegauz)\*alpha

head(absence)

# Level-2 VPC

absence$vpc2 = absence$variance2/(absence$variance2 + absence$variance1)

head(absence)

# Level-1 VPC

absence$vpc1 = absence$variance1/(absence$variance2 + absence$variance1)

head(absence)

# Summarize marginal statistics

colnames(absence)

sapply(absence[8:13], mean)

# Figure 5: Line plot of Level-2 VPC against the marginal expectation by FSM status

ggplot(data = absence, mapping = aes(x = expectation, y = vpc2)) +

 geom\_line(aes(group = fsm))

# Figure 5: Spikeplot of marginal expectation by FSM status

ggplot(data = absence, mapping = aes(x = expectation)) +

 geom\_histogram(binwidth=1)

################################################################################