The Variance of Intraclass Correlations In Three and Four Level Models

By

Larry V. Hedges

Northwestern University

Eric C. Hedberg

NORC

Arend M. Kuyper

Northwestern University

Citation: Hedges, L. V., Hedberg, E. C., & Kuyper, A. M. (2012). The variance of intraclass correlations in three-and four-level models. Educational and Psychological Measurement, 72(6), 893-909.

This paper is based in part on work supported by the US Institute of Education Sciences under grant # R305D11032 and National Science Foundation (NSF) under grants #0815295. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily represent the views of the IES or NSF.

The Variance of Intraclass Correlations in Three and Four Level Models

Intraclass correlations play important roles in genetics, epidemiology, psychometrics, and the design of social and educational experiments. In multistage sampling models and multilevel statistical models, they play a major role in quantifying the amount of clustering inherent in the data. The intraclass correlation was introduced by Fisher (1925) who offered an estimator of the intraclass correlation, derived its variance in balanced experiments, and discovered a variance stabilizing transformation. Estimators of the intraclass correlation from unbalanced designs and estimators based on slightly different principles have subsequently been introduced. See Donner (1986) for a survey of the literature on point and interval estimation of the intraclass correlation.

Intraclass correlations are important input parameters used for power computations and the computation of optimal sample allocations among levels in the design of multilevel randomized experiments (see Raudenbush, 1997; Kostantopoulos, 2009), and for computation of effect size estimates and their variances (see Hedges, 2007, 2010). Consequently there has been considerable interest in the estimation of intraclass correlations from sample surveys using multistage samples to estimate intraclass correlations (e.g., Hedges & Hedberg, 2007). Such studies typically fit unconditional multilevel models to the survey data to estimate variance components at each level of the sampling design. Other studies use datasets that were assembled in the course of carrying out randomized experiments (Bloom et al., 2007). While the surveys often have large total sample sizes, the number of sampled units at some levels (typically the higher levels) may not be large enough make negligible the sampling uncertainty of estimates of variance components and functions of variance components (such as intraclass correlations). Even experiments that are normatively large (that is large for experiments) typically involve much smaller sets of schools than national surveys (typically less than 100 schools). Consequently, for either type of study designed to provide reference values of intraclass correlations, it is important to provide some assessment of the uncertainty of the estimates. However because the sample sizes in surveys are actually large, estimates of sampling uncertainty based on large sample methods should be accurate enough to give this guidance.

Although intraclass correlation was originally introduced in relation to two level sampling models, the concept extends naturally to sampling models with three or more levels. The intraclass correlation concept in cases of three and four level sampling models is of great interest in the design and analysis of experiments in education (Hedges and Rhoads, 2010; Konstantopoulos, 2008ab, 2009). Three and four level experimental designs are of more than theoretical interest. A survey of recent educational experiments revealed that the most common designs actually involve four levels of sampling (Spybrook and Raudenbush, 2009). The literature on estimation of intraclass correlations, however, has largely been restricted to the case of two level models. An exception is a paper by Raykov (2010) that derived the large sample standard error for one of the intraclass correlations in the three level hierarchical linear model. We became interested in the problem of quantifying the uncertainty of estimates of intraclass correlations in three and four level sampling models as part of a project to develop estimates of three and four level intraclass correlations from survey data to provide guidance for planning randomized experiments in education.

The object of this paper is to provide simple estimates of the sampling variances and covariances of intraclass correlation estimates derived from multilevel statistical analyses of survey data. For example, suppose that we estimate the variance components associated with three level fully unconditional model (no covariates at any level of the model). Let σ_1^2, σ_2^2, and σ_3^2, be the variance components at the levels 1, 2, and 3 and let s_1^2, s_2^2, and s_3^2 be their maximum likelihood estimates. The large sample variances of the intraclass correlation estimates depend on the variances of the variance component estimates and on their covariances (when there are three or more levels).

Note that the *residuals* at different levels are usually assumed to be independent. However this does not mean that the *estimates* of residual variances from different levels are independent. In fact, in balanced designs where closed form expressions for the covariances are available, it is easy to see that the estimates are not independent (see Casella, McCollough, and Searle, 1992)

Statistical software for multilevel analyses (e.g., HLM) does not always provide estimates of the covariances of variance components at different levels, however. That is, while HLM and other programs provide the variances of s_1^2, s_2^2, and s_3^2, they do not provide the covariance of s_2^2 and s_3^2. Therefore we have emphasized providing expressions for the standard errors of the quantities of interest that do not require knowledge of the covariances between variance component estimates at different levels.

In realistic situations involving large scale survey data, the level 1 variance component is essentially known because there are so many level 1 units, so that $s_1^2 \approx \sigma_1^2$. That is, the variance v_1 of s_1^2 is essentially zero. Let v_j be the variance (squares of the standard errors of) the variance component estimates s_j^2 for $j > 1$. STATA, HLM, or SAS

Proc Mixed provide the standard errors of variance component estimates at level 2 or higher, so these (and therefore the corresponding v_j) are available even in unbalanced cases.

Two Level Model

In a two level model there are two variance components σ_1^2 and σ_2^2, which are estimated by s_1^2 and s_2^2, respectively, where $s_1^2 \approx \sigma_1^2$. Let v_1 and v_2 be the variances of s_1^2 and s_2^2. The condition $s_1^2 \approx \sigma_1^2$ implies that $v_1 = 0$. Let m denote the number of clusters (level 2 units) and n_i denote the number of level 2 units in the i^{th} level 2 unit. When the design is balanced, $n_1 = \cdots = n_m = n$. The interclass correlation in the two level model is

$$\rho = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

which is estimated by

$$r = \frac{s_2^2}{s_1^2 + s_2^2}.$$

Then the large sample variance of r, the estimate of ρ in a balanced design is

$$\frac{(1-\rho)^2 v_2}{\sigma_T^4} \qquad (1)$$

where v_2 is the variance of s_2^2, the sample estimate of the level 2 variance component and $\sigma_T^2 = \sigma_1^2 + \sigma_2^2$ is the total variance. The sample estimate of the variance of r is obtained by replacing all of the parameters in (1) by their samples estimates, that is the estimate of the variance of r is

$$\frac{(1-r)^2 v_2}{\left(s_1^2 + s_2^2\right)^2}.$$

The standard error of r is just the square root of its estimated variance.

Expression (1) does not appear to be similar to conventional expressions for estimates of the variance of intraclass correlations in balanced designs. For example, the expression given by Fisher (1925) for the large sample variance of the estimator of ρ is (in our notation)

$$\frac{2(1-\rho)^2\left[1+(n-1)\rho\right]^2}{n(n-1)(m-1)}. \qquad (2)$$

Donner and Koval (1980) gave an expression for the large sample variance of r computed from unbalanced designs as

$$\frac{2N(1-\rho)^2}{N\sum_{i=1}^{m}\left[1+(n_i-1)\rho^2\right]\Big/\left[1+(n_i-1)\rho\right]^2-\rho^2\left[\sum_{i=1}^{m}n_i(n_i-1)\Big/\left[1+(n_i-1)\rho\right]\right]^2}, \qquad (3)$$

where N is the total sample size. Note that (3) reduces to (2) when $n_1 = \cdots = n_m = n$.

Noting that the variance of s_2^2 in a balanced design is

$$\frac{2\left[\sigma_1^2+n\sigma_2^2\right]^2}{n^2(m-1)}=\frac{2\sigma_T^4\left[1+(n-1)\rho\right]^2}{n^2(m-1)},$$

we see that (1) differs from (2) only in that the implicit n^2 term in the denominator of (1) corresponds to an $n(n-1)$ term in the denominator of (2), that is they differ only in terms of order n^2, which implies that they are equivalent in large samples (that is, as $n \to \infty$).

Three Level Model

In the three level model, there are three variance components, one associated with each level of the model. Denote the number of level 3 units (clusters) by m, the number of level 2 units (subclusters) within the i^{th} level 3 unit by p_i, and the number of level 1 units (individuals) within the j^{th} level 2 unit within the i^{th} level 3 unit by n_{ij}. For example, if level 3 units are schools, level 2 units are classrooms, and level 1 units are individuals, then m denotes the number of schools, p_i is the number of classrooms in the i^{th} school, and n_{ij} is number of students within the j^{th} classroom of the i^{th} school. If the design is balanced, then $p_1 = \cdots = p_m$ and $n_{ij} = n$, $i = 1, \ldots, m; j = 1, \ldots, p$. Let σ_1^2, σ_2^2, and σ_3^2, be the level 1, 2, and 3 variance components, and let s_1^2, s_2^2, and s_3^2, be their maximum likelihood estimates. In realistic situations, the level 1 variance component is essentially known because there are so many level 1 units, so that $s_1^2 \approx \sigma_1^2$. Let v_2 and v_3 be the variances (squares of the standard errors of) the variance component estimates s_2^2 and s_3^2.

STATA provides the standard errors of variance component estimates, so these (and therefore v_2 and v_3) are available even in unbalanced cases. In the balanced case, it is easy to show that covariance between s_2^2 and s_3^2 is $-v_2/p$. Using this covariance, we can get the (large sample) variance of the intraclass correlation estimates.

Define the level 2 intraclass correlation as

$$\rho_2 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \frac{\sigma_2^2}{\sigma_T^2}$$

and its estimate as

$$r_2 = \frac{s_2^2}{s_1^2 + s_2^2 + s_3^2} = \frac{s_2^2}{s_T^2}.$$

Define the level 3 intraclass correlation as

$$\rho_3 = \frac{\sigma_3^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \frac{\sigma_3^2}{\sigma_T^2}$$

and its estimate as

$$r_3 = \frac{s_3^2}{s_1^2 + s_2^2 + s_3^2} = \frac{s_3^2}{s_T^2}.$$

The expressions for the large sample variances of intraclass correlations in the unbalanced three level model are quite complicated, even asymptotically. A quick approximation can be derived by assuming that the covariances of the maximum likelihood estimates of the variance components in the unbalanced case are the same as in the balanced case. In a balanced design, the number p of level 2 units per level 3 unit is a constant. In an unbalanced design, use an average value (such as the mean or harmonic mean) in place of p in the formulas below.

A straightforward derivation leads to the following variances assuming that the covariances between s_2^2 and s_3^2 are the same as in a balanced design. The variance of r_2 is

$$\frac{\left[p(1-\rho_2)^2 + 2\rho_2(1-\rho_2)\right]v_2}{p\sigma_T^4} + \frac{\rho_2^2 v_3}{\sigma_T^4}, \tag{4}$$

the variance of r_3 is

$$\frac{\left[p\rho_3^2 + 2\rho_3(1-\rho_3)\right]v_2}{p\sigma_T^4} + \frac{(1-\rho_3)^2 v_3}{\sigma_T^4} \tag{5}$$

and the covariance between r_2 and r_3 is

$$\frac{-\left[p\rho_3(1-\rho_2) + \rho_2\rho_3 + (1-\rho_2)(1-\rho_3)\right]v_2}{p\sigma_T^4} - \frac{\rho_2(1-\rho_3)v_3}{\sigma_T^4}. \tag{6}$$

Because you have the variance component estimates s_1^2, s_2^2, and s_3^2, you can compute r_2 r_3, and $s_T^2 = s_1^2 + s_2^2 + s_3^2$. Then, because we also have the variances v_2 and v_3 of the variance component estimates s_2^2, and s_3^2, we can compute estimates of the variances and covariance above by substituting the sample estimates r_2, r_3, and s_T^2 in place of ρ_2, ρ_3, and

σ_T^2 in (4), (5), and (6). The standard errors of r_2 and r_3 are just the square roots of their estimated variances.

When the design is unbalanced, the covariance between maximum likelihood estimates of the variance components s_2^2 and s_3^2 is more complex. Using results from Searle (1970), the covariance c_{23} between s_2^2 and s_3^2 is

$$c_{23} = -\frac{v_2 \sum_{i=1}^{m} \left(\sum_{j=1}^{p_i} n_{ij}^2 \Big/ \left(n_{ij}\sigma_2^2 + \sigma_1^2 \right)^2 \right) \Big/ \left(1 + a_i \sigma_3^2 \right)}{\sum_{i=1}^{m} a_i^2 \Big/ \left(1 + a_i \sigma_3^2 \right)} \tag{7}$$

where a_i is given by

$$a_i = \sum_{j=1}^{p_i} n_{ij} \Big/ \left(n_{ij}\sigma_2^2 + \sigma_1^2 \right).$$

Note that when the design is balanced so that $p_i = p$ and $n_{ij} = n$ for all i and j, then this covariance reduces to $-v_2/p$ as it should.

Using this covariance for unbalanced designs, the variance of r_2 is

$$\frac{\left(1-\rho_2\right)^2 v_2}{\sigma_T^4} + \frac{\rho_2^2 v_3}{\sigma_T^4} - \frac{2\rho_2\left(1-\rho_2\right)c_{23}}{\sigma_T^4} \tag{8}$$

and the variance of r_3 is

$$\frac{\rho_3^2 v_2}{\sigma_T^4} + \frac{\left(1-\rho_3\right)^2 v_3}{\sigma_T^4} - \frac{2\rho_3\left(1-\rho_3\right)c_{23}}{\sigma_T^4} \tag{9}$$

Calculations using (4) show that, unless there is extreme imbalance, the value of the variances of r_2 and r_3 computed from (8) and (9), respectively are remarkably similar to those obtained by using the mean of the p_i in (4) and (5) respectively.

Four Level Model

In the four level model, there are four variance components, one associated with each level of the model. Let σ_1^2, σ_2^2, σ_3^2, and σ_4^2, be the level 1, 2, 3, and 4 variance components, and let s_1^2, s_2^2, s_3^2, and s_4^2, be their maximum likelihood estimates. Denote the number of level 4 units (clusters) by m, the number of level 3 units (subclusters) within each level 4 unit by q, the number of level 2 units (sub-subclusters) within each level 3 unit by p, and the number of level 1 units (individuals) within each level 2 unit by n. For example, if level 4 units are school districts, the level 3 units are schools, the level 2 units are classrooms, and level 1 units are individuals, then m denotes the number of school districts, q is the number of schools per district, p is the number of classrooms per school, and n is number of students per classroom. Note that we assume the design is balanced. In realistic situations, the level 1 variance component is essentially known because there are so many level 1 units, so that $s_1^2 \approx \sigma_1^2$. Let v_2, v_3, and v_4 be the variances (squares of the standard errors of) the variance component estimates s_2^2, s_3^2, and s_4^2. Standard software provides the standard errors of variance component estimates, so these (and therefore v_2, v_3, and v_4) are available even in unbalanced cases. In the balanced case, it is easy to show that covariance between s_2^2 and s_3^2 is $-v_2/p$, where p is the number of level 2 units per level 3 unit, the covariance between s_3^2 and s_4^2 is

$$-\frac{v_3}{q}+\frac{v_2}{qp^2},$$

where q is the number of level 3 units per level 4 unit, and the covariance between s_2^2 and s_4^2 is zero.

Define the level 2 intraclass correlation as

$$\rho_2 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2} = \frac{\sigma_2^2}{\sigma_T^2}$$

and its estimate as

$$r_2 = \frac{s_2^2}{s_1^2 + s_2^2 + s_3^2 + s_4^2} = \frac{s_2^2}{s_T^2}.$$

Define the level 3 intraclass correlation as

$$\rho_3 = \frac{\sigma_3^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2} = \frac{\sigma_3^2}{\sigma_T^2}$$

and its estimate as

$$r_3 = \frac{s_3^2}{s_1^2 + s_2^2 + s_3^2 + s_4^2} = \frac{s_3^2}{s_T^2}.$$

Define the level 4 intraclass correlation as

$$\rho_4 = \frac{\sigma_4^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2} = \frac{\sigma_4^2}{\sigma_T^2}$$

and its estimate as

$$r_4 = \frac{s_4^2}{s_1^2 + s_2^2 + s_3^2 + s_4^2} = \frac{s_4^2}{s_T^2}.$$

Assuming that the covariance of the variance components is the same as in the balanced case, the large sample variance of the estimators r_2, r_3, and r_4 of the intraclass correlations can be obtained in terms of the variances v_2, v_3, and v_4.

The variance of r_2 is

$$\frac{\left[qp^2\left(1-\rho_2\right)^2 + 2qp\rho_2\left(1-\rho_2\right) + 2\rho_2^2\right]v_2}{qp^2\sigma_T^4} + \frac{\left(q-2\right)\rho_2^2 v_3}{q\sigma_T^4} + \frac{\rho_2^2 v_4}{\sigma_T^4}. \tag{10}$$

The variance of r_3 is

$$\frac{\left[qp^2\rho_3^2 + 2\left(qp-1\right)\rho_3\left(1-\rho_3\right)\right]v_2}{qp^2\sigma_T^4} + \frac{\left[q\left(1-\rho_3\right)^2 + 2\rho_3\left(1-\rho_3\right)\right]v_3}{q\sigma_T^4} + \frac{\rho_3^2 v_4}{\sigma_T^4}. \tag{11}$$

The variance of r_4 is

$$\frac{\left[qp(p-2)\rho_4^2 - 2\rho_4(1-\rho_4)\right]v_2}{qp^2\sigma_T^4} + \frac{\left[q\rho_4^2 + 2\rho_4(1-\rho_4)\right]v_3}{q\sigma_T^4} + \frac{(1-\rho_4)^2 v_4}{\sigma_T^4}. \quad (12)$$

The covariance between r_2 and r_3 is

$$\frac{\left[-qp^2\bar{\rho}_2\rho_3 - qp(\rho_2\rho_3 + \bar{\rho}_2\bar{\rho}_3) - \rho_2\bar{\rho}_3 + \rho_2\rho_3\right]v_2}{qp^2\sigma_T^4} - \frac{\left[q\rho_2\bar{\rho}_3 - \rho_2\bar{\rho}_3 + \rho_2\rho_3\right]v_3}{q\sigma_T^4} + \frac{\rho_2\rho_3 v_4}{\sigma_T^4}, \quad (13)$$

the covariance between r_2 and r_4 is

$$\frac{\left[-qp^2\bar{\rho}_2\rho_4 + qp(\bar{\rho}_2\rho_4 - \rho_2\rho_4) + \rho_2\rho_4 - \rho_2\bar{\rho}_4\right]v_2}{qp^2\sigma_T^4} + \frac{\left[q\rho_2\rho_4 - \rho_2\rho_4 + \rho_2\bar{\rho}_4\right]v_3}{q\sigma_T^4} - \frac{\rho_2\bar{\rho}_4 v_4}{\sigma_T^4}, \quad (14)$$

and the covariance between r_3 and r_4 is

$$\frac{\left[qp^2\rho_3\rho_4 + qp(\bar{\rho}_3\rho_4 - \rho_3\rho_4) + \rho_3\rho_4 - \bar{\rho}_3\bar{\rho}_4\right]v_2}{qp^2\sigma_T^4} - \frac{\left[q\bar{\rho}_3\rho_4 + \rho_3\rho_4 + \bar{\rho}_3\bar{\rho}_4\right]v_3}{q\sigma_T^4} - \frac{\rho_3\bar{\rho}_4 v_4}{\sigma_T^4}, \quad (15)$$

where $\bar{\rho}_2 = 1-\rho_2$, $\bar{\rho}_3 = 1-\rho_3$, and $\bar{\rho}_4 = 1-\rho_4$.

As in the three level case, we can compute estimates of the variances and covariances above by substituting the sample estimates r_2, r_3, r_4, and s_T^2 in place of ρ_2, ρ_3, ρ_4, and σ_T^2 in (10) through (15). The standard errors of r_2, r_3, and r_4 are just the square roots of their estimated variances.

Examples

For the examples presented in this paper we utilize a sample of the reading results from the 2009-2010 Kentucky Core Content Test, which is the state's test given to students in the spring of the academic year (Kentucky Department of Education, 2012). Our data includes reading scores from 46,849 fifth graders. These test scores are spread across 173 districts, 715 schools, and 2,142 teachers. The harmonic mean of the number of schools per district was 1.81 (arithmetic mean was 4.13), the harmonic mean number of teachers per district was 4.12 (arithmetic mean was 12.38), and the harmonic mean number of teachers per school was 2.04 (arithmetic mean was 3.00). The arithmetic average of 22 students per teacher (harmonic mean was about 8 students per teacher).

While the official evaluation of students includes both open ended responses as well as multiple choice responses, we focused only on the multiple choice portion of the test. This part of the test includes 39 items that are scored as correct or incorrect, with the final score representing the number of correct responses. In our data, the test scores ranged from 0 to 33, with a mean of 27.01 and a variance of 27.51.

The results of our example analyses are presented in Table 1. For the two level example, we fit a mixed effect model where we nested students (level 1) in schools (level 2). For the three level model, we fit a mixed model where we nested students (level 1) in teachers (level 2), and teachers in schools (level 3). In the four level model, our mixed effect model nests students (level 1) in teachers (level 2), teachers in schools (level 3), and schools in districts (level 4). In order to keep our example numbers tractable, we did not standardize our outcome, nor did we include any covariates in our models.

Two Level Models

Fitting a two level model to the Kentucky dataset, with students nested within schools, the total variance is estimated to be s_T^2 = 27.650, the level 2 variance component is s_2^2 = 2.409, and the variance of the level 2 variance component is v_2 = 0.024. With these parameters, we calculate the estimate of the school level intraclass correlation to be r = 0.087. We then calculate the variance of r as

$$\mathrm{var}(r) = \frac{(1-0.087)^2\, 0.024}{27.650^2} = \frac{0.020}{764.523} = 0.00003,$$

which implies a standard error of $\sqrt{0.00003} = 0.005$. Donner's formula gave approximately the same result for the standard error, also rounding to 0.005.

Three Level Models

We return to the Kentucky data for our example of a three level fully unconditional model, where students are the level 1 units, teachers define the second level, and schools define the third level. The outcome variable is again the raw reading scores. The parameter estimates from the model are $s_2^2 = 3.190$, $s_3^2 = 1.557$, $s_T^2 = 28.582$, $v_2 = 0.040$, $v_3 = 0.030$, $r_2 = 0.112$, and $r_3 = 0.054$. The harmonic mean of the number of teachers per school is $p = 2.042$. Using the formulas for the balanced case with $p = 2.042$, we calculate the variance of r_2 as

$$\begin{aligned}\mathrm{Var}(r_2) &= \frac{\left[2.042(1-0.112)^2 + 2(0.112)(1-0.112)\right]0.040}{2.042(28.582^2)} + \frac{0.112^2(0.030)}{28.582^2} \\ &= \frac{0.072}{1668.173} + \frac{0.0004}{816.931} = 0.00004 + 0.0000005 = 0.0000405\end{aligned}$$

which implies a standard error of $\sqrt{0.0000405} = 0.006$.

Using the formulas for the balanced case with $p = 2.042$, we calculate the variance of r_3 as

$$\begin{aligned}\mathrm{Var}(r_3) &= \frac{\left[(2.042)(0.054^2) + 2(0.054)(1-0.054)\right]0.040}{2.042(28.582^2)} + \frac{(1-0.054)^2\,0.030}{28.582^2} \\ &= \frac{0.004}{1668.173} + \frac{0.027}{816.931} = 0.000002 + 0.00003 = 0.000032\end{aligned}$$

which implies a standard error of $\sqrt{0.000032} = 0.006$.

Four Level Models

Using the Kentucky data for our four level model example, we fit a fully unconditional mixed model with teachers defining the second level, schools defining the third level, and school districts defining the fourth level. We again used the raw reading scores. The parameters from the model are $s_2^2 = 3.160$, $s_3^2 = 0.957$, $s_4^2 = 0.314$, $s_T^2 = 28.265$, $v_2 = 0.039$, $v_3 = 0.021$, $v_4 = 0.008$, $r_2 = 0.112$, $r_3 = 0.034$, and $r_4 = 0.011$. The harmonic mean of the number of teachers per school is $p = 2.042$, and the harmonic mean for the number of schools per district is $q = 1.818$. Using the formulas for the balanced case, we calculate the variance of r_2 as

$$\begin{aligned}\operatorname{Var}(r_2) &= \frac{\left[1.818\left(2.042^2\right)(1-0.112)^2+2(1.818)(2.042)0.112(1-0.112)+2\left(0.112^2\right)\right]0.039}{1.818\left(2.042^2\right)28.265^2}\\ &+\frac{(1.818-2)\left(0.112^2\right)0.021}{1.818\left(28.265^2\right)}+\frac{\left(0.112^2\right)0.008}{28.265^2}\\ &=\frac{0.263}{6056.244}+\frac{-0.00005}{1452.419}+\frac{0.0001}{798.910}=0.00004+(-0.00000003)+0.0000001\\ &=0.00004007\end{aligned}$$

which implies a standard error of $\sqrt{0.00004007} = 0.006$.

The variance of r_3 is calculated as

$$\begin{aligned}\operatorname{Var}(r_3) &= \frac{\left[1.818\left(2.042^2\right)0.034^2+2(1.818(2.042)-1)0.034(1-0.034)\right]0.039}{1.818\left(2.042^2\right)28.265^2}\\ &+\frac{\left[1.818(1-0.034)^2+2(0.034)(1-0.034)\right]0.021}{1.818\left(28.265^2\right)}+\frac{0.034^2(0.008)}{28.265^2}\\ &=\frac{0.007}{6056.244}+\frac{0.037}{1452.419}+\frac{0.00001}{798.910}=0.000001+0.00003+0.00000001\end{aligned}$$

which implies a standard error of $\sqrt{0.00003101} = 0.006$.

The variance of r_4 is calculated as

$$\text{Var}(r_4) = \frac{\left[1.818(2.042)(2.042-2)\left(0.011^2\right)-2(0.011)(1-0.011)\right]0.039}{1.818\left(2.042^2\right)28.265^2}$$

$$+\frac{\left[1.818\left(0.011^2\right)+2(0.011)(1-0.011)\right]0.021}{1.818\left(28.265^2\right)}+\frac{(1-0.011)^2\,0.008}{28.265^2}$$

$$=\frac{-0.001}{6056.244}+\frac{0.0005}{1452.419}+\frac{0.008}{798.910}=-0.0000002+0.0000003+0.00001$$

which implies a standard error of $\sqrt{0.0000101} = 0.003$.

ICCVAR Software for Stata

We have developed software to perform these calculations in Stata. Once the user downloads and installs the ICCVAR.ado program, performing these calculations is quite simple. The first step is to run the mixed model as you normally do in Stata. Then, immediately after the mixed model is estimated, simply type in "iccvar" into the command prompt to estimate the ICCs and variances, just as you would any other Stata "post-estimation" command. The program then uses the variance components and variances stored by the estimation command to perform the calculations. The ICCVAR program automatically detects the number of levels used in the previous mixed model. When a three level model is specified, you can enter an optional "unb" command to use the unbalanced formulas. Figure 1 provides an example session using the program. This program also stores the matrix of intraclass correlations and the variance-covariance matrix in the results memory. This SATA code can be downloaded from http://www.ipr.northwestern.edu/qcenter/iccvar.html .

Conclusions

We have provided formulas for the large sample variances and covariances of estimates of intraclass correlations in three and four level models. These expressions are suitable for computing standard errors for estimates of three and four level intraclass correlations when the sampling designs are balanced or nearly so. These should be useful in providing some estimates of sampling uncertainty for analyses designed to yield reference values of intraclass correlations based on large scale data collections.

Figure 1: Example Stata session using ICCVAR

```
. quietly : xtmixed reading || schno :, var

. iccvar

Intraclass Correlation Estimates


--------------------------------------------------------------------
             |      ICC      Std. Err.      [95% Conf. Interval]
-------------+------------------------------------------------------
       schno |   0.08714      0.00507        0.07720       0.09707
--------------------------------------------------------------------

. quietly : xtmixed reading || schno : || readingid :, var

. iccvar

Intraclass Correlation Estimates

Harmonic Mean of Level 2 Units per Level 3 Unit  =       2.042

--------------------------------------------------------------------
             |      ICC      Std. Err.      [95% Conf. Interval]
-------------+------------------------------------------------------
       schno |   0.05447      0.00593        0.04284       0.06610
   readingid |   0.11162      0.00662        0.09864       0.12460
--------------------------------------------------------------------

. iccvar, unb

Intraclass Correlation Estimates

Harmonic Mean of Level 2 Units per Level 3 Unit  =       2.042

--------------------------------------------------------------------
             |      ICC      Std. Err.      [95% Conf. Interval]
-------------+------------------------------------------------------
       schno |   0.05447      0.00586        0.04298       0.06596
   readingid |   0.11162      0.00650        0.09888       0.12436
--------------------------------------------------------------------

. quietly : xtmixed reading || leaid : || schno : || readingid :, var

. iccvar

Intraclass Correlation Estimates

Harmonic Mean of Level 2 Units per Level 3 Unit  =       2.042
Harmonic Mean of Level 3 Units per Level 4 Unit  =       1.818

--------------------------------------------------------------------
             |      ICC      Std. Err.      [95% Conf. Interval]
-------------+------------------------------------------------------
       leaid |   0.01111      0.00312        0.00500       0.01722
       schno |   0.03385      0.00512        0.02381       0.04389
   readingid |   0.11179      0.00658        0.09889       0.12470
--------------------------------------------------------------------
```

Table 1: Example Results

	Two Level Model		Three Level Model		Four Level Model	
	Estimate	Variance	Estimate	Variance	Estimate	Variance
Fixed Effect						
Intercept	26.931	0.004	26.665	0.005	26.872	0.007
Random Effects						
District	--	--	--	--	0.314	0.008
School	2.409	0.024	1.557	0.030	0.957	0.021
Teacher	--	--	3.190	0.040	3.160	0.039
Student	25.241	0.000[a]	23.835	0.000[a]	23.834	0.000[a]
Total Variance	27.650		28.582		28.265	
ICCS						
District	--	--	--	--	0.011	0.0000101
School	0.087	0.00003	0.054	0.000032	0.034	0.00003101
Teacher	--	--	0.112	0.0000405	0.112	0.00004007

Notes: N= 46,849 students in 2,142 teachers in 715 schools in 173 districts. The harmonic mean of teachers per school is 2.042, the harmonic mean number of schools per district is 1.818, a: we assume this variance to be 0.000

References

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision: Empirical guidelines for studies that randomize schools to measure the impacts of educational interventions. *Educational Evaluation and Policy Analysis*, *29(1)*, 30-59.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review, 54*, 67-82.

Donner, A. & Koval, J. J. (1980). The large sample variance of the intraclass correlation. *Biometrika, 67*,719-722.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Hedges, L. V. & Rhoads, C. (2010). *Statistical power analysis in educational research*. Washington, DC: Institute of Education Science, National Center for Educational Research.

Hedges, L.V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, *32(2),* 151-179.

Hedges, L.V. & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29(1)*, 60-87.

Kentucky Department of Education. 2012. Kentucky Core Content Test. Retrieved from http://www.education.ky.gov/kde/administrative+resources/testing+and+reporting+/district+support/kentucky+core+content+test/ on Jan 1 2012.

Konstantopoulos, S. (2008a). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness, 1,* 66-88.

Konstantopoulos, S. (2008b). The power of the test for treatment effects in three-level block randomized designs. *Journal of Research on Educational Effectiveness, 1,* 265-288.

Konstantopoulos, S. (2009). Incorporating Cost in Power Analysis for Three-Level Cluster Randomized Designs. *Evaluation Review, 33,* 335-357.

Rao, C. R. (1973). *Linear statistical inference*. New York: Wiley.

Raykov, T. (2010). Proportion of third level variation in multi-level studies: A note on an interval estimation procedure. *British Journal of Mathematical and Statistical Psychology, 63*, 417-426.

Raudenbush, S. W. (1997). Statistical Analysis and Optimal Design for Cluster Randomized Experiments. *Psychological Methods*, *2*, 173-185.

Searle, S. R. (1970). Large sample variances of maximum likelihood estimators of variance components. *Biometrics, 26*, 505-524.

Searle, S. R. (1971). *Linear models*. New York: John Wiley.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: John Wiley.

Spybrook, J., & Raudenbush, S.W. (2009). An examination of the precision and technical accuracy of the first wave of group randomized trials funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis*, *31(3)*.

Appendix

This appendix describes the multilevel models on which the results of this paper are based and the method used to derive those results.

Two Design

Suppose that *m* clusters of size *n*. Let Y_{ij} be the j^{th} observation in the i^{th} cluster. Then the level 1 (individual level) model is

$$Y_{ij} = \beta_{0i} + \varepsilon_{ij} \quad , i = 1, \ldots, m; j = 1, \ldots n$$

where β_{0i} is mean of the i^{th} cluster and the ε_{ij} are independently normally distributed with mean 0 and variance σ_1^2. The level 2 (cluster level) model is

$$\beta_{0i} = \gamma_{00} + \eta_{0i}, i = 1, ..., m,$$

where γ_{00} is the grand mean and the η_{0i} are independently normally distributed with mean 0 and variance σ_2^2.

We use the delta method to obtain the results of the paper (see, e.g., Rao, 1973, pp 385-387), which implies that if a random variable *Y* is a function f(*T*) of another random variable *T* which has variance *v*, then the variance of *Y* in large samples is $(\partial f(\theta)/\partial T)^2 v$, where $\theta = E(T)$. Starting with s_2^2, we note that *r* is a function of s_2^2 which is random and $s_1^2 \approx \sigma_1^2$, which is a constant. Therefore $r = f(s_2^2) = s_2^2/(s_1^2 + s_2^2)$, and $\partial f/\partial s_2^2 = (1 - r)/s_2^2$, which implies (1).

Three Level Design

Suppose that *m* clusters, each with *p* subclusters of size *n*. Let Y_{ijk} be the k^{th} observation in j^{th} subcluster of the i^{th} cluster. Then the level 1 (individual level) model is

$$Y_{ijk} = \beta_{0ij} + \varepsilon_{ijk} \;, i = 1, \ldots, m; j = 1, \ldots p;\; k = 1, \ldots n$$

where β_{0ij} is mean of the j^{th} subcluster in the i^{th} cluster, and the ε_{ijk} are independently normally distributed with mean 0 and variance σ_1^2. The level 2 (subcluster level) model is

$$\beta_{0ij} = \gamma_{00i} + \eta_{0ij}, j = 1, ..., p; i = 1, ..., m,$$

where γ_{00} is the mean of the i^{th} cluster and the η_{0ij} are independently normally distributed with mean 0 and variance σ_2^2. The level 3 (cluster level) model is

$$\gamma_{00i} = \pi_{00} + \xi_{0i}, i = 1, ..., m,$$

where π_{00} is the grand mean, and the ξ_{0i} are independently normally distributed with mean 0 and variance σ_3^2.

We use the multivariate delta method to obtain the results of the paper (see, e.g., Rao, 1973, pp 388-389). This method uses the fact that if a 1 x p vector of random variables $\mathbf{T} = (T_1, \ldots, T_p)$ has p x p covariance matrix $\mathbf{\Sigma}$, and if $\mathbf{Y} = \mathrm{f}(\mathbf{T}) = (\mathrm{f}_1(\mathbf{T}), \ldots, \mathrm{f}_q(\mathbf{T}))$ is a 1 x q vector of differentiable functions of $\mathbf{T}$, then the covariance matrix of $\mathbf{Y}$ in large samples is given by $\mathbf{A\Sigma A}'$, where $\mathbf{A}$ is a q x p matrix in which the element of the i^{th} row and j^{th} column is $(\partial \mathrm{f}_i(\boldsymbol{\theta})/\partial T_j)$, where $\boldsymbol{\theta} = \mathrm{E}(\mathbf{T})$. Apply this result with $q = 2$ and $p = 2$. Define $\mathbf{T} = (s_2^2, s_3^2)$, note that s_1^2 is a constant, $r_2 = \mathrm{f}_1(\mathbf{T}) = s_2^2/(s_1^2 + s_2^2 + s_3^2)$, and $r_3 = \mathrm{f}_2(\mathbf{T}) = s_3^2/(s_1^2 + s_2^2 + s_3^2)$. Noting that $(\partial \mathrm{f}_i(\boldsymbol{\theta})/\partial T_j) = (1 - \rho_i)/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)$ when $i = j$ and $(\partial \mathrm{f}_i(\boldsymbol{\theta})/\partial T_j) = -\rho_i/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)$ when $i \neq j$ yields (4), (5), and (6).

Four Level Design

Suppose that m clusters, each with q subclusters in each cluster, and p sub-subclusters within each subcluster of size n. Let Y_{ijkl} be the l^{th} observation in k^{th} sub-subcluster of the j^{th} subcluster of the i^{th} cluster. Then the level 1 (individual level) model is

$$Y_{ijkl} = \beta_{0ijk} + \varepsilon_{ijkl} \qquad , i = 1, \ldots, m; j = 1, \ldots p; \ k = 1, \ldots q; l = 1, \ldots, n$$

where β_{0ijk} is mean of the k^{th} sub-subcluster of the j^{th} subcluster in the i^{th} cluster, and the ε_{ijkl} are independently normally distributed with mean 0 and variance σ_1^2. The level 2 (sub-subcluster level) model is

$$\beta_{0ijk} = \gamma_{0ij} + \eta_{0ijk}, \ i = 1, ..., m; j = 1, ..., p; k = 1 , \ldots, q,$$

where γ_{0ij} is the mean of the j^{th} subcluster in the i^{th} cluster and the η_{0ijk} are independently normally distributed with mean 0 and variance σ_2^2. The level 3 (subcluster level) model is

$$\gamma_{0ij} = \pi_{0i} + \xi_{0ij}, j = 1, ..., p, i = 1, ..., m,$$

where π_{0i} is the mean of the i^{th} cluster and the ξ_{0ij} are independently normally distributed with mean 0 and variance σ_3^2. The level 4 (cluster level) model is

$$\pi_{0i} = \theta_0 + \zeta_{0i}, i = 1, ..., m,$$

where θ_0 is the grand mean, and the ζ_{0i} are independently normally distributed with mean 0 and variance σ_4^2.

Use the multivariate delta method again with $q = 3$ and $p = 3$. Define the 1 x 3 vector $\mathbf{T} = (s_2^2, s_3^2, s_4^2)$ note that s_1^2 is a constant, $r_2 = \mathrm{f}_1(\mathbf{T}) = s_2^2/(s_1^2 + s_2^2 + s_3^2 + s_4^2)$, and $r_3 = \mathrm{f}_2(\mathbf{T}) = s_3^2/(s_1^2 + s_2^2 + s_3^2 + s_4^2)$, and $r_4 = \mathrm{f}_3(\mathbf{T}) = s_4^2/(s_1^2 + s_2^2 + s_3^2 + s_4^2)$. Noting that $(\partial \mathrm{f}_i(\boldsymbol{\theta})/\partial T_j) = (1 - \rho_i)/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2)$ when $i = j$ and $(\partial \mathrm{f}_i(\boldsymbol{\theta})/\partial T_j) = -\rho_i/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2)$ when $i \neq j$ yields (10) through (15).