University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Interval Estimation of Proportion of Second-level Variance in Multi-level Modeling

Steven Svoboda
University of Nebraska-Lincoln, ssvoboda@huskers.unl.edu

Interval Estimation of Proportion of Second-level Variance in Multi-level Modeling

Steven Svoboda

Department of Education Psychology
University of Nebraska-Lincoln

Abstract

Physical, behavioral and psychological research questions often relate to hierarchical data systems. Examples of hierarchical data systems include repeated measures of students nested within classrooms, nested within schools and employees nested within supervisors, nested within organizations. Applied researchers studying hierarchical data structures should have an estimate of the intraclass correlation coefficient (ICC) for every nested level in their analyses because ignoring even relatively small amounts of interdependence is known to inflate Type I error rate in single-level models. Traditionally, researchers rely upon the ICC as a point estimate of the amount of interdependency in their data. Recent methods utilizing an interval estimation of the amount of interdependency based the proportion of second-level variance between groups have been developed that avoid relying solely upon point estimates. The likelihood of committing a Type I error when using the interval estimation of the proportion of second-level variance remains unknown. The current project addressed this deficiency in knowledge utilizing simulated data to assess the accuracy of a 95% confidence interval estimation of the proportion of second-level variance (CI-PSLV). Standard errors tended to decrease as sample size increased, and the CI-PSLV captured the second level ICC in 95% of replications.

The recognition of hierarchical data structures that account for the dependence of observations and corresponding methods to analyze them have received considerable amounts of attention in the past few decades (Raykov, 2010). Traditionally, the intraclass correlation coefficient (ICC), or the proportion of variance in an outcome variable that is between groups or contained at the higher levels of the nested data structure, is used to determine whether or not accounting for the hierarchical nesting of participants is necessary (Raudenbush & Bryk, 2002). Unfortunately, widely agreed upon guidelines for interpreting the magnitude of the ICC do not exist (Raykov, 2010). Recent methods have been developed to interpret the proportion of variance at higher levels between groups using a confidence interval estimation procedure rather than relying solely upon point estimates such as the ICC because the magnitude of the ICC is difficult to interpret (Raykov et al., 2016). However, there remains a critical need to investigate the accuracy of these methods and to develop guidelines that may be followed when evaluating sample size requirements for the underlying asymptotic maximum likelihood estimation theory to obtain practical relevance when using an interval estimation procedure (Raykov et al., 2016). The purpose of the current study was to establish a 95% confidence interval estimation procedure for the proportion of second-level variance (CI-PSLV) as a valuable tool for applied researchers to consider when deciding whether or not to account for hierarchical data structures in their samples. Specifically, the current study provided a demonstration of the CI-PSLV and explored its accuracy in various sample sizes under various degrees of second level dependence.

Motivating Context

Physical, behavioral and psychological research questions often relate to hierarchical data systems (Mass & Hox, 2004). Examples of hierarchical data systems include repeated measures of students nested within classrooms, nested within schools and employees nested within

supervisors, nested within organizations. Hierarchical data structures exist in nature whether psychologists and behavioral scientists recognize their existence and account for the nesting of their subjects within higher order units in applied research. Obviously, observations may not be independent in these data structures. Failure to account for hierarchical data structures likely violates the assumption that errors are independent of each other and identically distributed. This violation would result in biased standard errors associated with the regression coefficients, which in turn, leads to an increased Type I error rate and erroneous interpretations of statistical tests (Mass & Hox, 2004; Raudenbush & Bryk, 2002). Standard (single-level) models are not appropriate in these hierarchical data systems because individual observations at the lowest level are not independent (Mass & Hox, 2004). Multi-level modeling (MLM), also known as hierarchical linear modeling, techniques avoid having to meet the independence of observations assumption of single-level regression models by accounting for the interdependence of level one observations due to hierarchical nesting structures (Raykov, 2010). However, researchers must adhere to financial budgets and time limits that may render multi-level models infeasible due to the costs associated with collecting additional data from higher nested levels. Because of these limitations, it is imperative that applied researchers consider the proportion of variance in an outcome associated with a given level when deciding how to analyze their hierarchical data.

Intraclass Correlation Coefficient

The intraclass correlation coefficient (ICC) is commonly used to provide researchers with an estimate of the amount of interdependence due to hierarchical nesting structures. The ICC informs researchers' decisions when choosing between MLM techniques and single-level modeling (Mass & Hox, 2005; Raudenbush & Bryk, 2002). For two-level models the ICC is calculated by:

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}},$$

where σ^2 is the level one variance and τ_{00} is the level two variance.

Applied researchers should have an estimate of the ICC for every nested level in their analyses because ignoring even relatively small amounts of interdependence (ICC values as small as .005) is known to inflate Type I error rate in single-level models (Mass & Hox, 2005; Raudenbush & Bryk, 2002). However, ICCs greater than .005 are commonly observed in hierarchical data structures. For example, the ICCs in a recent study of school climate on students' academic achievement range from .04 to .08 (Maxwell et al., 2017). ICCs as great as .522 have been observed in a study of players from track and field clubs (Swierzy et al., 2018). These examples of hierarchical data structures contained 2 levels of nesting. The formula for calculating the ICC is easily adapted for additional levels (more than two) by adding a term for the variance associated with each higher level to the denominator and inserting variance of the level of interest into the numerator and provides a point estimate of the proportion of variance in an outcome variable associated with a given level.

The Problem with Point Estimates

Point estimates, such as the ICC, have been criticized for being too dependent on the characteristics of a single, usually small, sample (Schmidt, 1996). This is especially problematic for multi-level modeling techniques that require relatively large sample sizes. As a result, the ICC may be significantly different from zero simply because of the large sample sizes commonly collected from hierarchical data structures. Furthermore, no informative guidelines exist for interpreting the magnitude of the ICC and the definition of a "meaningful" ICC depends on the

context of the research (Raykov et al., 2016). Some nationally recognized scientific organizations, such as the American Psychological Association, encourage researchers to report confidence intervals for each statistic because of these known problems with point estimates and null hypothesis significance tests in general.

Interval Estimation of the Proportion of Second-level Variance

Recent methods have been developed utilizing an interval estimation procedure to estimate the proportion of variance in an outcome variable attributed to higher levels of nesting that avoid the problem of relying on a null hypothesis significance test for the ICC. One such method is the CI-PSLV (Raykov et al., 2016). To determine the CI-PSLV, first an unconditional model is fit to the data to furnish an estimate of the standard error associated with the PSLV using the delta method (Raykov & Marcoulides, 2004). Second, the PSLV is calculated using the following formula:

$$PSLV = \frac{\tau_\pi}{\sigma^2 + \tau_\pi + \tau_\beta} \, ,$$

where σ^2 is the level one variance, τ_π is the level two variance, and τ_β is the level three variance. Finally, a 95% confidence interval is obtained based on the estimate of the PSLV and its standard error (Raykov et al., 2016). Please refer to Appendix A for an R-function that computes the endpoints of the CI-PSLV.

Rather than relying solely upon a point estimate, the CI-PSLV provides a range of plausible values for the PSLV in a population under consideration and is an informative supplement to the ICC (Raykov et al., 2016). The current study seeks to establish the accuracy of a 95% confidence interval of the PSLV in various sample sizes and under various degrees of second level dependence.

Method

The current study examines the accuracy of the CI-PSLV using simulated data. Data were generated in R 3.4.3 (R Core Team, 2017) using the following three-level model:

Level 1: $\quad Y_{ijk} = \beta_{0jk} + e_{ijk}$

Level 2: $\quad \beta_{0jk} = \gamma_{00k} + r_{0jk}$

Level 3: $\quad \gamma_{00k} = \delta_{000} + u_{00k}$

Composite: $\quad Y_{ijk} = \delta_{000} + u_{00k} + r_{0jk} + e_{ijk}$,

where Y_{ijk} is the score of an individual i in second level j within the third level k on the dependent variable in the simulated data. δ_{000} is the fixed effect and the remaining terms represent random effects at levels three, two and one respectively in the composite model. The variance of Y_{ijk} is normally distributed.

Simulation Conditions

The following three conditions, as specified in Mass and Hox (2005), varied in the simulation: level two ICC (ICC = .1, .2 & .3), number of clusters in level two (L2NC = 30, 50 & 100), and number of clusters in level three (L3NC = 30, 50 & 100). The variance of Y_{ijk} was fixed to one in level one and .1 in level three across conditions, whereas the variance of Y_{ijk} at level two differed between conditions to vary the level two ICC. Sample size varied between conditions based on the total number of clusters in levels two (L2NC) and three (L3NC). Level one group size was fixed at 30 across conditions because thirty level-one units is a reasonable number to expect in educational settings (Mass & Hox, 2005). Level two ICC, L2NC and L3NC varied across 27 conditions (3 x 3 x 3).

Procedure

One thousand datasets were generated for each combination of conditions. An unconditional, three-level model was fit to each simulated dataset to demonstrate the accuracy of the CI-PSLV associated with a predicted outcome variable Y_{ijk} using M*plus*. A model constraint was used to estimate the PSLV its standard error in all datasets. Please refer to Appendix B for M*plus* source code for the estimation of PSLV and its standard error from the data generated with R. Estimates of the PSLV and their associated standard errors were saved for each dataset within each condition and were imported into R to calculate their 95% confidence intervals.

The accuracy of the CI-PSLV was assessed by comparing average standard errors of estimated PSLV, the proportion of CI-PSLV that include the actual ICC, and average width of CI-PSLV between simulated conditions. Accuracy of the CI-PSLV was assessed by the proportion of confidence intervals that included the population's level two ICC within a given simulation condition.

Results

Results from the simulation provide evidence in favor of the utility of the CI-PSLV under the conditions studied and are provided in Table 1. As can be seen from Table 1, the population level two ICC fell within the CI-PSLV in about 95% of repeated samples. Standard errors associated with the estimate of PSLV and the average width of confidence intervals tended to decrease as sample size (L2NC & L3NC) increased and were slightly larger in conditions with larger ICC conditions. The CI-PSLV seemed to capture the actual second level ICC and none of the confidence intervals included zero. Lower standard errors of the estimate of PSLV were associated with a greater number of level three clusters relative to level two clusters in conditions with equal, overall sample size (L3NC=100 & L2NC = 50 compared to L3NC=50 & L2NC=100;

L3NC=100 & L2NC = 30 compared to L3NC=30 & L2NC=100; and L3NC=50 & L2NC = 30 compared to L3NC=30 & L2NC=100).

Discussion

The current study provides evidence of the applied utility of the CI-PSLV. The CI-PSLV appears to capture the actual second level ICC as long as sample size requirements for multi-level modeling are met to begin with as described in Mass and Hox (2005). None of the confidence intervals in the current study included zero. Future work needs to be conducted to determine the usefulness of the CI-PSLV for detecting a second level ICC close to zero (smaller than .1 which is the lower bound of the current study). Similarly, research needs to be conducted to address the robustness of the CI-PSLV to violations of the assumption that outcome measures (Y_{ijk}) are normally distributed.

Multi-level modeling is appropriate in a variety of fields given the inherent, hierarchical nature of data they utilize. Similarly, the application of an interval estimate of the PSLV is appropriate anytime three-level data structures are encountered, regardless of the specific research area. The results of the current study provide evidence of the accuracy of the CI-PSLV, which avoids many of the known problems associated with relying solely on point estimates for null hypothesis significance tests of the traditional ICC. This does not suggest the CI-PSLV should replace the traditional ICC, but rather, it should be reported as another piece of evidence in conjunction with the ICC (Raykov et al., 2016). Once its efficiency is established, the CI-PSLV procedure will provide a range of plausible estimates for the amount of interdependency of scores due to hierarchical data structures and that should be reported in addition to the traditional ICC.

References

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58,* 127-137.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92.

Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2017). The impact of school climate and school identification on academic achievement: Multilevel Modeling with student and teacher data. *Frontiers in Psychology, 8*, 1-21.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Application and data analysis methods* (2nd ed). Thousand Oaks, CA: Sage.

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.

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling, 11*, 621-637.

Raykov, T., Patelis, T., Marcoulides, G. A., & Lee, C. (2016). Examining intermediate omitted levels in hierarchical designs via latent variable modeling. *Structural Equation Modeling, 23*, 111-115.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115-129.

Swierzy, P., Wicker, P., & Breuer, C. (2018). Usefulness of multilevel modeling in sport management research: The case of voluntary roles in nonprofit sports clubs.

Measurement in Physical Education and Exercise Science, 22, 1-12.

Table 1.

Average Standard Errors of the PSLV Estimate, Number of Intervals Capturing ICC & CI Width Across Simulation Conditions for Data Generated Using R

		Level 2 ICC								
		.3			.2			.1		
Number of Clusters										
L3NC	L2NC	S.E.	#CI	Width	S.E	#CI	Width	S.E	#CI	Width
100	100	.0043	954	.0169	.0033	957	.0131	.0020	942	.0080
	50	.0055	957	.0216	.0043	952	.0168	.0026	954	.0104
	30	.0068	950	.0268	.0053	954	.0209	.0034	952	.0129
50	100	.0061	951	.0239	.0047	951	.0186	.0029	956	.0113
	50	.0078	951	.0305	.0061	951	.0238	.0037	949	.0146
	30	.0097	957	.0379	.0075	954	.0295	.0047	951	.0183
30	100	.0079	956	.0309	.0061	960	.0240	.0037	962	.0146
	50	.0101	959	.0394	.0078	955	.0307	.0048	956	.0189
	30	.0125	940	.0489	.0097	940	.0381	.0060	940	.0237

Appendix A

R-Function for Interval Estimation of Proportion of Second-Level Variance from Raykov et al. (2016)

```
ci.pslv = function(pslv, se){
 l = log(pslv /(1 - pslv))
 sel = se/(pslv*(1 - pslv))
 ci_l_lo = l-1.96*sel
 ci_l_up = l+1.96*sel
 ci_pslv_lo = 1/(1+exp(-ci_l_lo))
 ci_pslv_up = 1/(1+exp(-ci_l_up))
 ci = c(ci_pslv_lo, ci_pslv_up)
 ci
}
```

Appendix B
M*plus* Source Code for Estimation of the Proportion of Second-Level Variance adapted from
Raykov et al. (2016)

TITLE: Interval Estimation of PSLV from Simulated Data

```
DATA:
FILE IS Mpluslist.txt;
TYPE=MONTECARLO;

VARIABLE:
NAMES ARE L1ID L2ID L3ID repID u00k r0jk eijk y;
USEVARIABLE ARE y;
CLUSTER ARE L3ID L2ID;

ANALYSIS:
TYPE = THREELEVEL;
ESTIMATOR IS ML;

MODEL:
%WITHIN%

y* (P1);
%BETWEEN L2ID%

y* (P2);
%BETWEEN L3ID%

y* (P3);

MODEL CONSTRAINT:
NEW(PSLV);
PSLV=P2/(P1+P2+P3);

OUTPUT:

STDYX;

SAVEDATA: RESULTS = results.dat;
```