Educational and Psychological
Measurement
2015, Vol. 75(6) 1063–1070
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013164414564052
epm.sagepub.com
SAGE

Intraclass Correlation Coefficients in Hierarchical Design Studies With Discrete Response Variables: A Note on a Direct Interval Estimation Procedure

Tenko Raykov[1] and George A. Marcoulides[2]

Abstract

A latent variable modeling procedure that can be used to evaluate intraclass correlation coefficients in two-level settings with discrete response variables is discussed. The approach is readily applied when the purpose is to furnish confidence intervals at prespecified confidence levels for these coefficients in setups with binary or ordinal outcome measures and nesting of subjects within higher order units. The method can aid educational and behavioral researchers in their study of sources of observed outcome variability and model choice considerations in multilevel settings, and is illustrated with empirical survey data.

Keywords

hierarchical data, interval estimation, intraclass correlation coefficient, multilevel modeling

Data are commonly termed multilevel, nested, or hierarchical when they are obtained, described, or organized through various levels of clustering or aggregation of units of analyses. For example, data about the attitudes and beliefs of university employees

[1] Michigan State University, East Lansing, MI, USA
[2] University of California, Santa Barbara, CA, USA

Corresponding Author:
Tenko Raykov, Measurement and Quantitative Methods, Michigan State University, 443A Erickson Hall, East Lansing, MI 48824, USA.
Email: raykov@msu.edu

can be hierarchically organized by department or by school. Such multilevel data may be considered the rule rather than the exception in the behavioral, educational, medical, psychological, organizational, and social disciplines. Nesting or clustering in studied populations in them is a phenomenon that tends to exist prior to an empirical investigation, and makes it quite likely that observed subjects' dependent variable scores at the time of data collection will be correlated within higher order units (e.g., departments, schools, colleges, teams, counselors, clinicians, etc.). These and related reasons have contributed over the past few decades to a great deal of interest in the use of multilevel modeling in the educational, behavioral, and social sciences. Given this enhanced attention to multilevel settings, having a readily available method for evaluating clustering effects in empirical studies becomes essential.

A popular measure for evaluating clustering effects is the intraclass correlation coefficient (ICC; e.g., Raudenbush & Bryk, 2002). While its point estimation can be straightforwardly accomplished with many widely circulated modeling software (e.g., HLM, Mplus, SPSS, Stata, or R), its interval estimation has received considerably less attention in the past but has been recently also made routinely available in some of them (e.g., Stata; see also Raykov, 2011). However, the underlying interval estimation procedures typically assume continuous response variables, while it is widely appreciated that outcome measures in many empirical studies may be binary, binary scored, or ordinal (such as true/false answers or Likert-type responses on individual items obtained via multicomponent measuring devices). In such situations, interval estimation of ICCs becomes particularly important. Unfortunately, to date, no routinely applicable procedure seems to be available for discrete or categorical outcomes that accounts adequately for the sampling distribution of the ICC.

The present note addresses this gap by discussing an interval estimation approach that can be used to obtain intervals of plausible population values of the ICCs in two-level settings with categorical response measures. The resulting confidence intervals can aid empirical researchers in studying observed variance decomposition and proportion between-group variance under these circumstances. In addition, the outlined procedure can be useful for making more informed decisions on model choice, in particular when considering whether to proceed with single- or two-level analysis, with the former being typically associated with simpler result interpretations. The method discussed below is developed within the framework of latent variable modeling (LVM) and is directly applicable and illustrated with the popular LVM software Mplus (Muthén & Muthén, 2014).

Background, Assumptions, and Notation

In the remainder of this note we consider a two-level setting, such as the one arising for instance when examined students are nested within schools, employees within departments, patients within clinicians, interviewees within interviewers, and so on. In this setup, denote by Y_{ij} the observed score on an outcome variable for the ith level-1 unit (e.g., student) in the jth level-2 unit (e.g., school; $i = 1, \ldots, n_j$, $j = 1, \ldots,$

J, with J being the number of sampled level-2 units and n_j the size of the sample of level-1 units within the jth level-2 unit). We assume in this discussion that Y_{ij} is a categorical response measure (e.g., a binary or binary scored item or a Likert-type item with more than 2 but relatively limited number of possible response options). The (fully) unconditional two-level model, which this discussion is based on, is defined as follows (e.g., Raudenbush & Bryk, 2002):

$$Y_{ij} = \beta_{0j} + r_{ij}, \tag{1}$$

and

$$\beta_{0j} = \gamma_{00} + u_{0j}, \tag{2}$$

where r_{ij} denotes the deviation of the individual score Y_{ij} from the mean β_{0j} in the jth level-2 unit and u_{0j} is the deviation of this mean from the response grand mean γ_{00}, with these two deviations assumed uncorrelated.

Equations (1) and (2) are typically used as the basis for defining the ICC with continuous outcome, in terms of the ratio of between-group to observed response variance (e.g., Rabe-Hesketh & Skrondal, 2012). When the outcome variable is categorical, however, while one could still work out this ratio as a potentially informative quantity associated with observed variance, interval estimation of the ICC cannot proceed along the same lines as in the continuous outcome case since the distribution of the response is not continuous. For this case, however, one can invoke the routinely made assumption in latent variable modeling of an underlying latent variable Y_{ij}^* ''behind'' the response, which variable is of actual interest but not directly observable or measurable (i = 1, . . ., n_j, j = 1, . . ., J; see also Snijders & Bosker, 2012). Specifically, if the outcome has k possible response options ($k > 1$), the existence of $k-1$ thresholds is also assumed, which are denoted $\tau_1, \tau_2, \ldots, \tau_{k-1}$, and their relationships to the underlying variable and response measure is as follows:

$$Y_{ij} = \begin{array}{ll} 0, & \text{if } -\infty < Y_{ij}^* \le \tau_1 \\ 1, & \text{if } \tau_1 < Y_{ij}^* \le \tau_2 \\ \vdots & \\ k-1 & \text{if } \tau_{k-1} < Y_{ij}^* < \infty \quad (i = 1, \ldots, n_j,\ j = 1, \ldots, J). \end{array} \tag{3}$$

The underlying variable Y_{ij}^* plays an important role in the rest of this note. A test of the assumption of existence of such variables, under certain conditions, is provided in Raykov and Marcoulides (2015).

Interval Estimation of Intraclass Correlation Coefficients With Discrete Outcomes

Intraclass Correlation Coefficients for a Categorical Outcome

When the outcome variable is categorical, Snijders and Bosker (2012) proposed to estimate the ICC, denoted below ρ, as the ratio

$$\rho = \tau^2/(\tau^2 + \pi^2/3), \tag{4}$$

where τ^2 is the between group variance and π denotes the popular constant 3.1416 . . . (see also Rabe-Hesketh & Skrondal, 2012).[1]Equation (4) is based on the logistic regression analysis approach that can be used to fit a two-level model with a discrete outcome (e.g., Raudenbush & Bryk, 2002), and the well-known fact that the standard logistic distribution of relevance has a variance of $\pi^2/3$. The ICC in Equation (4) is interpretable as the proportion of between group variance for the underlying latent variable Y^*_{ij} (Snijders & Bosker, 2012).

Interval Estimation of the Intraclass Correlation

While point estimation of the ICC in Equation (4) is straightforward using for instance LVM accounting for the discrete nature of the response variable, in particular employing the LVM software Mplus (e.g., see details in the appendix), its interval estimation with large samples needs to attend to the fact that the ICC cannot be negative or larger than 1. To account for this lower and upper bounds of the ICC, the approach followed in Raykov and Marcoulides (2011) can be used. Accordingly, after an initial logit transformation of the ICC, whose sampling distribution approximates better a normal distribution, a delta method-based standard error (SE) can be obtained for the logit of the ICC (e.g., Raykov & Marcoulides, 2004). With that SE, a symmetric confidence interval (CI) is furnished for that transformed ICC, by subtracting and adding 1.96 times the SE from the estimated logit of the ICC at a given confidence level (in this case, of 95%). Finally, using the logistic distribution that is the inverse of the logit function one obtains the endpoints of a large-sample CI of the ICC, which is of actual interest (see below).

Empirical Implementation of ICC Interval Estimation Procedure

In an empirical study, the large-sample SE for the logit of the ICC can be readily obtained using LVM and the popular software Mplus. The source code needed thereby is provided in the appendix. Once this SE is available, the CI for the ICC itself, which interval is of main interest, is furnished by employing the R-function ''ci.rel'' from Raykov and Marcoulides (2011, chap. 7). For completeness of this note, that R-function is also provided in the appendix, with a minor adaptation for the purposes of this discussion.

We demonstrate next the discussed ICC interval estimation procedure for discrete outcomes using empirical data.

Illustration on Empirical Data

In this section, we use data from the General Social Survey of 2006 (Hamilton, 2012). For our aims, we consider the binary responses to the question ''Should marijuana be legalized?'' from that survey as the outcome variable, and the census divisions as level-2 units where respondents can be considered nested within. If one is interested in finding out the intraclass correlation coefficient associated with this dichotomous item, as outlined in the preceding discussion one can fit the two-level model (1) and (2) accounting for the discrete nature of the response. To this end, as indicated earlier, we can use the LVM software Mplus (see also Raykov, 2011, for a latent variable modeling conceptualization of the currently considered setting, and the appendix to this note for the needed source code.)

Fitting this model to the data from the 652 respondents with answers on this question yields the between-group variance estimate $\hat{\tau}^2$= .204, with an SE of .147. With this estimate, based on Equation (4) (see also the appendix), the used software furnishes the estimated ICC value as

$$\hat{\rho} = .204/(.204 + 3.289) = .058, \tag{5}$$

which it also reports as being associated with an SE = .040. The R-function ‘‘ci.rel’’ from Raykov & Marcoulides (2011, chap. 7, see also the appendix) yields then with the estimate in Equation (5) and its SE of .40 the following 95% CI of the ICC for this survey question:

$$(.014,\ \ .205) \tag{6}$$

That is, the set of plausible population values for the percentage of between census division differences ranges (at the 95% confidence level) between 1.4% through 20.5%. Although the left end-point of the confidence interval (6) appears to be quite close to 0, its right end-point suggests that as high as 20% (rounded off) of the variance in the underlying propensity/inclination for agreement with marijuana legalization—the survey question of interest here—could be due to census division differences, that is, stem from potential geographic region differences. One could thus also conclude that it would be recommendable that subsequent analyses of the data on this survey question proceed within a two-level modeling framework.[2]

Conclusion

This note was concerned with confidence interval estimation of the popular ICC in two-level settings where the outcome variable is discrete, such as binary or binary scored answers or Likert-type item/question responses. The discussed procedure provides a large-sample CI for this coefficient, which yields a range of plausible values for the population proportion of between group outcome variance. This CI can therefore significantly assist educational, behavioral, and social scientists seeking to describe the sources of individual differences on a discrete response measure in a study where examined subjects are nested or clustered in higher order units. The method has also the potential to help scholars in the difficult and at times controversial model selection process in empirical research, especially when they consider choosing between single- and two-level modeling approaches.

The discussed approach has also several limitations. As indicated earlier, the underlying estimation procedure is best used with large samples both with respect to number of level-2 units as well as subjects within them (e.g., Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002; Raudenbush & Xiao-Feng, 2001). This follows from the fact that the outlined method instrumentally relies on maximum likelihood estimation (cf. the appendix) that is itself grounded in a large-sample statistical theory. We encourage future research for developing possible guidelines or procedures that may be followed in evaluating sample size requirements for the underlying asymptotic maximum likelihood estimation theory to obtain practical relevance in an empirical setting. Furthermore, we would like to point out that our aim in this note is not to suggest any minimal threshold for the ICC below which one could employ single-level models in a two-level setting. In our view such a threshold, if at all possible to arrive at, will typically be at least related if not mostly dependent on substantive considerations, prior research and accumulated knowledge in a subject matter domain of application, as well as on the particular research question(s) and study details and aims.

In conclusion, this note provides educational, behavioral, and social scientists with a readily used method for interval estimation of ICCs when the dependent variable cannot be considered approximately continuous but should instead be treated as discrete (categorical). The confidence intervals furnished by the discussed procedure are likely to be a helpful aid for (a) better understanding of the variation in clustered data with discrete responses and (b) potentially facilitating more informed conclusions about model choice in hierarchical design studies.

Appendix

Mplus Source Code for Point and Estimation of the Proportion of Second-Level Variation

```
TITLE:      EMPIRICAL EXAMPLE OF ICC INTERVAL ESTIMATION
            WITH DISCRETE RESPONSE.
DATA:       FILE = <NAME OF RAW DATA FILE >;
VARIABLE:   NAMES = <NAMES OF VARIABLES IN FILE >;
            USEV = MARIJUAN; ! SELECT RESPONSE VARIABLE OF INTEREST
            CLUSTER = CENDIV; ! NAME OF LEVEL-2 UNITS
            CATEGORICAL = MARIJUAN;
ANALYSIS:   TYPE = TWOLEVEL;
            ESTIMATOR = ML;
MODEL:      %WITHIN%
            %BETWEEN%
            MARIJUAN (BETW_VAR);
MODEL CONSTRAINT:
            NEW(ICC);
            ICC = BETW_VAR/(BETW_VAR + 3.2899); ! CONSTANT IS pi^2/3
OUTPUT:     CINTERVAL;
```

Note. After a title and indication of the location of the data to be analyzed, the VARIABLE command assigns names to the measures in the file. The following ANALYSIS command requests two-level modeling, while the subsequent MODEL command defines the two-level model of relevance (see Equations 1 and 2) and assigns a symbol/label to the between variance. The MODEL CONSTRAINT subsection first introduces a 'place holder' for the ICC parameter ρ (Equation 5), and then defines it formally. (For further details on the Mplus syntax, see Muthén & Muthén, 2012.)

R-Function for Interval Estimation of Intraclass Correlation With Discrete Outcome

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

Note. At the R prompt, paste this R-function, and call it subsequently by entering for ''icc_do'' and ''se'' correspondingly the estimate and standard error of the ICC with discrete outcome, which are obtained with the Mplus command file above in this appendix. (This function is an adaptation for the present purposes of the R-function ''ci.rel'' in Raykov & Marcoulides, 2011, chap. 7.)

Acknowledgments

We are grateful to K. Preacher for valuable discussions on multilevel modeling.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. In this article, all population (and sample estimates of) variances are assumed to be positive, which is at most a rather mild assumption in educational and behavioral studies. This ensures that the ICC in Equation (5) exists, like its logit transformation that is used in a later section.
2. Throughout this note, we purposely dispense with presentation of *p*-values associated with possible statistical tests. For the aims of the note, and in particular for evaluation of proportion between group variance, in our view confidence intervals contain a great deal of

relevant information about outcome and related variability. For this reason, we are exclusively concerned throughout with confidence intervals instead of *p*-values (e.g., Schmidt, 1996).

References

Hamilton, L. C. (2012). *Statistics with Stata: Version 12* (8th ed.). College Station, TX: Stata Corp.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Corp.

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

Raudenbush, S. W., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*, 387-401.

Raykov, T. (2011). Intra-class correlation coefficients in hierarchical designs: Evaluation using latent variable modeling. *Structural Equation Modeling*, *18*, 73-90.

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parametric functions in covariance structure models. *Structural Equation Modeling*, *11*, 659-675.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.

Raykov, T., & Marcoulides, G. A. (2015). On examining the underlying normal variable assumption in latent variable models with categorical indicators. *Structural Equation Modeling*, *22*, 581-587.

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

Snijders, T. A. B., & Bosker, R. (2012). *Multilevel analysis*. Thousand Oaks, CA: Sage.