Reliability of Coefficients

Intercepts and slopes in each group can be thought of as estimates of a corresponding value from a population, where the group can be conceptualized as a sample. Larger samples should be better estimates of their population values. For example, we generally expect a mean, randomly drawn from a population, to be a closer estimate to the population value if there is a larger sample size. For the multilevel model, the level-1 intercept, β_{0j} in the intercept-only model is a mean estimate of its corresponding value in a population. With a smaller sample or with higher degree of variability of scores within the group, we should have less confidence in the precision of the estimate. For multilevel regression, this degree of precision is quantified by the reliability of the coefficient.

The estimate of the reliability of a coefficient for a particular group, λ_j, is fairly simple to compute.

$$\lambda_j = \frac{\tau_0^2}{\tau_0^2 + \sigma^2 / n_j}$$

τ_0^2 is the intercept variance between groups, σ^2 is the withing group variance, and n_j is the number of individuals in the group. Values for λ_j, can range between 0, representing no reliability, and 1.0, representing perfect reliability.[1] You will notice that the equation is quite similar to the ICC equation (in fact, this is sometimes referred to as ICC2 and is *not unrelated* to measurement reliability; Bliese, 1988; Shrout & Fleiss, 1979). The second term in the denominator is the within-group variance, σ^2, divided by the group size, n_j. Both lower within-group variability or larger group size with make this term smaller and lead to higher reliability.

Reliability of the slope coefficient can also be computed in a similar fashion.

$$\lambda_j = \frac{\tau_1^2}{\tau_1^2 + \sigma^2 / n_j}$$

These equations represent the reliability estimate for a coefficient in a particular group, j, but an average reliability can be estimated by averaging all the individual group estimates of the coefficients (Raudenbush & Bryk, 2002; Eq 4.7).[2] This average coefficient reliability is useful sometimes for diagnosing problems with a model (e.g., convergence issues, warning messages with profile likelihood confidence intervals) as well as for estimating statistical power. Generally, intercepts will have higher reliability than slopes. This makes sense if you think about the fact that smaller samples are needed for a good estimate of the mean than the sample required for a good estimate of the slope.

Empirical Bayes or "Shrunken" Estimates

If we were to compute a regression line for every group, these coefficient values could be called ordinary least squares estimates (OLS). We can also obtain a predicted value for the intercept or slope value for a particular group using the model estimates, γ_{00} or γ_{01} and their associated estimates of between-group variance. These model-based estimates are sometimes called empirical Bayes estimates, because they are derived from a weighted average of the full sample estimate and the group estimate. They are thus "posterior" estimates, in which we borrow some other information from the rest of the groups. For example, for the posterior intercept estimate for a particular group, $\hat{\beta}_{0j}^{EB}$, information about the intercept from all the groups combined, γ_{00}, is used. The two values, group and sample, are combined, weighted by the reliability of the coefficient.

$$\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j}^{OLS} + \left(1 - \lambda_j\right)\hat{\gamma}_{00}$$

[1] There are not really any cutoffs that have been recommended for "acceptable" or "good" reliability in the multilevel context.
[2] Alternatively, a harmonic mean for the sample sizes of n_j can be computed (see examples below), which can be inserted into these equations for a single average estimate of reliability.

If the reliability is low, the posterior or empirical Bayes estimate will be more heavily weighted by the average intercept for the sample γ_{00}.[3] If the reliability is high, the posterior estimate will be more a function of the group estimate. In the extremes of reliability, if λ_j is 0, then $\hat{\beta}_{0j}^{EB} = \hat{\gamma}_{00}$ and if λ_j is 1, then $\hat{\beta}_{0j}^{EB} = \beta_{0j}^{OLS}$. Whenever, the reliability is less than perfect, the posterior estimate will be closer to the estimate from the full sample, and this leads to less variability across groups. They are, thus, said to be "shrunken" estimates. Plots of the model-based EB estimates will show less variability of intercepts or slopes (as we have seen in the handout "Plotting Within-Group Regression Lines: SPSS, R, and HLM") than of the OLS estimates.

Obtaining Coefficient Reliability Values

I used the results from the HSB example with random slopes (see the handout "Random Slopes Example: SPSS, R, and HLM") with math achievement as the outcome and SES as the predictor. Intercept, slope, and within-group variance estimates were obtained from those analyses. HLM provides reliability values in the output by default, so no special effort is needed (note that the reliability values in HLM below do not match the hand computation with SPSS, manual computation with R, or the `mlmhelpr` results from R).

SPSS

SPSS mixed does not provide reliability estimates, so they would have to be computed by hand by using the above equations with an average group size for n_j. The harmonic mean of the sample size is recommended for the average n_j [i.e., a sum of the reciprocal (inverse) of n_j is computed for each group, those values are summed (which is $\sum\left(1/n_j\right)$, and then the total number of groups is divided by the sum].

Here is code for computing the harmonic mean in SPSS by first extracting group sizes.

```
AGGREGATE
  /BREAK schoolid
  /nj=n(mathach).

*list vars=nj.

*reshaped data to get one nj value per record.
CASESTOVARS
 /ID=schoolid
 /DROP size to mathach.

*list vars=all.

*compute harmonic mean.
Compute invnj=1/nj.

List vars=nj invnj.
AGGREGATE
  /ngroups=n(nj)
  /suminv=sum(invnj).

Compute harmmean=ngroups/suminv.

Print /harmmean.
Exe.
```

The harmonic mean is 41.06, which is then entered into the above equation for either the intercept or the slope.

$$\lambda_{intercept} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2 / n} = \frac{4.828}{4.828 + 36.830 / 41.06} = .84$$

[3] Instead of $\hat{\beta}_{0j}^{OLS}$ we could also write $\bar{Y}_{.j}$, the group mean, in the equation.

$$\lambda_{slope} = \frac{\tau_1^2}{\tau_1^2 + \sigma^2 / n} = \frac{.413}{.413 + 36.830 / 41.06} = .32$$

R

The R package `mlmhelpr` provides close to the same results as obtained with the hand computation for SPSS above (due to rounding in my computation and slightly different variance estimates in SPSS and R).

```
> #can obtain reliabilities for intercept and slope but don't match HLM
> library(mlmhelpr)

Attaching package: 'mlmhelpr'

The following object is masked from 'package:psych':

    reliability> reliability(model)
$schoolid
schoolid (Intercept)         schoolid ses
           0.8456629            0.3294565
```

A manual computation could also be used and produces the same results as `mlmhelpr`.

```
#compute reliabilities manually
#get lme4 objects
sigma <- attr(VarCorr(model),"sc")^2
tau00 <- attr(VarCorr(model)$schoolid, "stddev")["(Intercept)"]^2
tau11 <- attr(VarCorr(model)$schoolid, "stddev")["ses"]^2

#can compute reliability using each group lambda and then average (see Raudenbush & Bryk, 2002, EQ 4.7)
#(same result as harmonic mean approach below)
lambdaj = (tau00/(tau00+sigma/nj))
lambdaj
relint = mean(lambdaj)
relint

#alternatively, use the coefficients and the harmonic mean.
   #obtain harmonic mean of Ns
   #first obtain vector for a variable
library(dplyr)
by_group <- d %>% group_by(schoolid)
nj = group_size(by_group)
   #calculate harmonic mean
library(psych)
n = harmonic.mean(nj)
#n = mean(nj)
   #using harmonic mean in the equation
relint = (tau00/(tau00+sigma/n))
relint
relslope = (tau11/(tau11+sigma/n))
relslope
```

HLM

Run the same model as illustrated in the random slopes handout and the reliability values are printed in the output by default. Notice these values are lower than those obtained above, despite similar variance estimates. HLM must be doing something different in weighting the average of the group reliability values.

Level-1 Model

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Mixed Model

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10}*SES_{ij} + u_{0j} + u_{1j}*SES_{ij} + r_{ij}$$

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.797
SES,β_1	0.179

References

Bliese, P. D. (1998). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods, 1*(4), 355-373.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.