Example R² Computation

In multilevel models, there are no universally agreed upon measures of multiple R^2 (total variance accounted for in the outcome), although several have been proposed (for reviews, see LaHuis, Hartman, Hakoyama, & Clark, 2014; Rights & Sterba, 2019; Roberts, Monaco, Stovall, & Foster, 2011). Because the R^2 values are not strictly defined in this circumstance, they are often considered "pseudo- R^2" values and I would recommend reporting them as "approximate variance accounted for." Below I illustrate a couple of the possibilities using the Bryk & Raudenbush's HSB data wihth `mathach` as the outcome variable. Following recommendations of the developers of these indices, the estimates below were obtained using full maximum likelihood (FML) rather than the default restricted maximum likelihood estimation (more on this distinction in a subsequent handout). To simplify calculations, the variance of the SES slope was not estimated. Snijders and Bosker (2012; see also LaHuis et al., 2012) recommend against estimating variance of slopes when computing R-squared estimates, but Hox (2010, p. 76) and Rights and Sterba (2019, 2020) have approaches for models with random slopes. The values form some R^2 values can sometimes be negative, although this is not supposed to be theoretically possible. Negative values are likely to only occur when there is a very small proportion of variance accounted for by the predictors and probably could just be reported as 0. Software often has not output R^2 values in the past, so they have tended to go unreported in the literature. The computations, however, are relatively simple; and having even an approximate variance accounted for is valuable. R^2 values can be obtained in R, and beginning in Version 29, SPSS reports an R^2 value but does not provide any documentation on which one or how it is computed. HLM does not report any R-square measures.

Computations

I've used the results from the HSB examples with math achievement and SES that I illustrated in class (see results from the handouts "Intercept Only Model Example (Random Effects ANOVA): SPSS, R, and HLM" and "ANCOVA Example (One Level-1 Predictor Assuming Homogeneous Slopes): SPSS, R, and HLM"). The "null model" (or intercept only) results come from the first handout and the "full model" results come from the second handout which included SES.

The measure suggested by Snijders and Bosker (1999, pp. 102-103) is one option. This approach distinguishes proportion of variance accounted for in the individual-level outcome Y_{ij} by the level-l predictors from the variance accounted for in the group-mean level outcome by the level-2 predictors.

Variance accounted for in Y_{ij} by level-1 predictors:

$$
\begin{aligned}
R_1^2 &= 1 - \frac{\hat{\sigma}^2(full) + \hat{\tau}_0^2(full)}{\hat{\sigma}^2(null) + \hat{\tau}_0^2(null)} \\
&= 1 - \frac{37.03 + 4.77}{39.15 + 8.61} \\
&= 1 - \frac{41.80}{47.76} = 1 - .88 = .12
\end{aligned}
$$

Where the *full* refers to the model tested and *null* refers to the model without predictors, or the empty model. σ^2 is the within-group variance and τ_0^2 is the between group (or intercept) variance.

We do not have any level-2 predictors yet, but the variance accounted for in $\bar{Y}_{.j}$ is calculated as:

$$
\begin{aligned}
R_2^2 &= 1 - \frac{\hat{\sigma}^2(full)/B + \hat{\tau}_0^2(full)}{\hat{\sigma}^2(null)/B + \hat{\tau}_0^2(null)} \\
&= 1 - \frac{37.03/45 + 4.77}{39.15/45 + 8.61} \\
&= 1 - \frac{5.59}{9.48} \\
&= 1 - .59 = .41
\end{aligned}
$$

B is the average cluster size in the notation used by Roberts and colleagues. I cheated on the computation for *B*, because I simply took the arithmetic average by dividing the total sample size, 7185, by the number of groups, 160. This approach may be vulnerable to the influence of groups with very large or very small sample sizes. The harmonic mean of group sample sizes (i.e., average n_j) is recommended as a more accurate computation of *B*.

Xu (2003) proposed an overall measure of variance accounted (r^2 or Ω_0^2 "omega-squared") for that does not require specific reference to level-1 or level-2 predictors or outcomes. I have yet to see this measure reported much in the literature to date, but Xu's simulation work suggests that it performs well. Only the within-group variance is used in this measure, which I obtain from the output above.

$$\begin{aligned} r^2 &= 1 - \frac{\sigma^2}{\sigma_0^2} \\ &= 1 - \frac{37.03}{39.15} \\ &= .054 \end{aligned}$$

Xu uses σ^2 for the full model residual variance and σ_0^2 for the null model residual variance. Clearly the proportion of variance accounted for differs substantially from these two different approaches, so the definition used has important implications for the conclusions one might draw.

Lahuis and colleagues (2014) review another total variance measure proposed by Nakagawa and Schielzeth (2013) that uses the variance of predicted values of Y_{ij}, given below as $\mathrm{var}\left(\hat{Y}_{ij}\right)$.[1] The predicted scores must be saved from the model and their variance calculated separately. In the equation below, τ_{00} is the intercept variance and σ^2 is the within group variance. This R^2 worked well in the Lahuis simulation, but I suspect will be less widely used until it is programmed into software packages.

$$R^2\left(MVP\right) = \frac{\mathrm{var}\left(\hat{Y}_{ij}\right)}{\mathrm{var}\left(\hat{Y}_{ij}\right) + \tau_{00} + \sigma^2}$$

The simulation study by LaHuis and colleagues (2014) suggested that all of the measures they examined worked well, except the variance accounted for measure at level-2, R_2^2, given above (the Xu measure for total variance account for given above was not examined in their study).

Rights and Sterba (2019) propose a modified approach that avoids negative values for R-square. Their general framework divides potential variance accounted for up into several sources, fixed effects, random slope, and mean variation across groups (τ_0^2). Instead of two models, using null and full model residuals as other measure do, they just derive values from the full model only, allowing for any number of level-1 or level-2 predictors and random effects.

$$R^2 = \frac{explained\, variance\, from\, full\, model}{outcome\, variance\, from\, full\, model}$$

Outcome variance differs depending on which source is used and does not have a very simple expression (see Rights & Sterba, 2019, for details).

SPSS

There is no indication in the SPSS documentation (e.g., command syntax documentation or Advanced Statistics manual) about which pseudo-R^2 measures these are. The terms "marginal" and "conditional" refer to R-square measures in which the random effect is not included or included in the model, respectively (Orelien & Edwards, 2008). Note that the values do match the Rights and Sterba values obtained below with the `r2mlm` package.

[1] The `MuMIn` package with `r.squaredGLMM` function in R will compute the Nakagawa and Schielzeth (2013) measure.

Coefficients of Determination

Pseudo-R Square Measures	Marginal	.077
	Conditional	.182

R code for computing the Xu pseudo-R-square measure

```
> #get empty (or null) model using ML rather than REML
> library(nlme)
> modeln <- lme(mathach ~ 1, random = ~ 1|schoolid, data = mydata, method="ML")
> summary(modeln)
        [output omitted]

> #get full model using ML rather than REML
> modelf <- lme(mathach ~ ses, random = ~ 1|schoolid, data = mydata, method="ML")
> summary(modelf)
        [output omitted]

> #compute Xu (2003) r-square manually
> 1-(var(residuals(modelf))/var(residuals(modeln)))
[1] 0.05285959

> #Rights & Sterba (2019) R-square measures (random slopes allowed)
> #in the output, sources of variance are f=level-1 and level-2 predictors combined
> #(f1=level-1 predictors, f2=level-2 predictors),
> #v=level-1 predictors via random slope, m=cluster specific means via random intercept
> #fv and fvm are from multiple sources.
> library(r2mlm)
> r2mlm(modelf)
$Decompositions
                    total
fixed          0.07680054
slope variation 0.00000000
mean variation  0.10453867
sigma2          0.81866079

$R2s
         total
f   0.07680054
v   0.00000000
m   0.10453867
fv  0.07680054
fvm 0.18133921
```

For both the `f` (fixed effects only) and the `fv` values (fixed plus random slope) R^2 = .077, and these seem to make the most sense to me in the context of how R-squared is defined elsewhere. As there is no random slope in this model, these two values are the same.

Comments

Reporting of these values is by no means universal at this point. One reason is that there has been disagreement about the best approach, because there is no simple parallel to the R^2 obtained with standard OLS regression. In some instances, these proportion of variance measures can be negative. The issue of what to do with slope variance has also been a hindrance. More simulation work and consensus are likely needed and implementation in software packages is likely necessary before reporting of R^2 becomes more widespread. However, I believe it is better to use some metric of variance accounted for than none at all. Multilevel models have been quite negligent in providing magnitude of effect information, including computing and reporting standardized coefficients, to date.

References

Hox, J. J. (2010). *Quantitative methodology series. Multilevel analysis: Techniques and applications (2nd ed.)*. New York, NY, US: Routledge/Taylor & Francis

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*(4), 433-451.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142. doi:10.1111/j.2041-210x.2012.00261.x

Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R2 statistics. *Computational Statistics & Data Analysis, 52*(4), 1896-1907.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309–338

Rights, J. D., & Sterba, S. K. (2020). New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivariate Behavioral Research, 55*(4), 568-599.

Roberts, K.J., Monaco, J.P., Stovall, H., & Foster, V. (2011). Explained variance in multilevel models (pp.219-230). In J.J. Hox & J.K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis.* New York: Routledge.

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

Snijders, T.A.B., & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd Edition).* London: Sage. I