Estimating explained variance in mixed-effects models using R

Hugo Quené
Utrecht institute of Linguistics OTS, Utrecht University

28 June 2012

Here is a more detailed explanation on how to estimate the proportional reduction of prediction error, which is the mixed-effects equivalence of R^2. This example is based on Snijders & Bosker (1999), Chapter 7.

During today's course session we obtained the following code and output:

```
require(languageR)
require(lme4)
arnhold <- read.table(
   file=url("http://www.hugoquene.nl/bimm2012/arnhold2.txt"),
   header=T)

arnholdm.m00 <- lmer(HdistM1end*1000~1+(1|subject)+(1|item),
   data=arnhold, REML=F)  # empty model

arnholdm.m00
#Linear mixed model fit by maximum likelihood
#Formula: HdistM1end * 1000 ~ 1 + (1 | subject) + (1 | item)
#   Data: arnhold
#   AIC   BIC logLik deviance REMLdev
# 28601 28624 -14296    28593   28588
#Random effects:
# Groups   Name        Variance Std.Dev.
# item     (Intercept)   104.00  10.198
# subject  (Intercept)   220.14  14.837
# Residual             10148.14 100.738
#Number of obs: 2367, groups: item, 24; subject, 17
#
```

```
#Fixed effects:
#            Estimate Std. Error t value
#(Intercept)  -15.512      4.652  -3.335

# add quantity as predictor, will explain between-item var
arnholdm.m01 <- update(arnholdm.m00, .~.+quantity)

anova(arnholdm.m00,arnholdm.m01)
#             Df   AIC   BIC logLik  Chisq Chi Df Pr(>Chisq)
#arnholdm.m00  4 28601 28624 -14296
#arnholdm.m01  5 28599 28627 -14294 4.2344      1    0.03961 *

arnholdm.m01
#Linear mixed model fit by maximum likelihood
#Formula: HdistM1end * 1000 ~ (1 | subject) + (1 | item) + quantity
#   Data: arnhold
#   AIC   BIC logLik deviance REMLdev
# 28599 28627 -14294    28589   28578
#Random effects:
# Groups   Name        Variance  Std.Dev.
#item     (Intercept)    72.419   8.5099
#subject  (Intercept)   219.070  14.8010
# Residual             10146.757 100.7311
#Number of obs: 2367, groups: item, 24; subject, 17
#
#Fixed effects:
#            Estimate Std. Error t value
#(Intercept)  -21.355      5.275  -4.049
#quantitytwo   11.566      5.419   2.134
```

prediction error at level two

First, we should realize that the predictor named *quantity* is a between-items predictor, i.e. it captures linguistic properties of the items, and not of the subjects nor of the individual responses. Hence we will inspect the reduction of prediction error for item means — and not the reduction of prediction error for individual responses nor for subject means. In terms of Snijders & Bosker (1999), this amounts to reduction of variance at level two (level one being individual responses).

"The level-two explained proportion of variance is now defined as the proportional reduction in mean square prediction error for $\overline{Y}_{.j}$", i.e. for item means (Snijders & Bosker 1999, p.103). It is estimated as the proportional reduction in the value of $\hat{\sigma}^2/n + \hat{\tau}^2$, "where n is an representative value for the group size", i.e. for the number of responses per item in this example. The quantity $\hat{\sigma}^2$ is the estimated residual variance (for model `arnholdm.m00`, this is 10148.14); the quantity $\hat{\tau}^2$ is the estimated variance between units (i.e. between items, for model `arnholdm.m00`, this is 104.00).

As representative value, we use the harmonic mean of n, here rounded to 97:

```
harmonic.mean <- function( x, na.rm=T )
    { return( 1/mean(1/x,na.rm=na.rm)) }
table(arnhold$item) -> table1
harmonic.mean(table1)
# [1] 96.80596
```

For model `arnholdm.m00`, we obtain $\hat{\sigma}^2/n + \hat{\tau}^2 = 10148.14/97 + 104.00 = 208.62$. For model `arnholdm.m01`, we obtain $\hat{\sigma}^2/n + \hat{\tau}^2 = 10146.757/97 + 72.419 = 177.02$.

The proportional reduction in mean square prediction error at level 2 (items), R_2^2, is then estimated as $1 - (177.02/208.62) = 0.151$ or 15%.

prediction error at level one

For completeness, we might also estimate the proportional reduction in mean square prediction error at level one, i.e. for the individual responses. It is defined as the proportional reduction in the value of $\hat{\sigma}^2 + \hat{\tau}^2$ (without division by n), with quantities defined as above.

For model `arnholdm.m00`, we obtain $\hat{\sigma}^2 + \hat{\tau}^2 = 10148.14 + 104.00 = 10252.14$. For model `arnholdm.m01`, we obtain $\hat{\sigma}^2 + \hat{\tau}^2 = 10146.757 + 72.419 = 10219.18$.

The proportional reduction in mean square prediction error at level 1 (residuals), R_1^2, is then estimated as $1 - (10219.18/10252.14) = 0.003$. Obviously adding a between-item predictor does not reduce the estimated within-item variance.

Reference

Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling.* London: Sage. ISBN 0-7619-5890-8.