FI = HLML03bb

MORE HLM MODELS

by
Ralph B. Taylor
breck@rbtaylor.net

We explore here more varieties of "reduced" HLM models.

ANOVA

This assumes we already understand the oneway ANOVA HLM model, and what it tells us.

The ANOVA model was covered in the last file HLML02

It is extremely important to conduct the oneway ANOVA as a data exploration step. It tells you about important features of your data you would not otherwise learn about.

- percent between and within variance on the outcome, and related intraclass correlation
- overall model lack of fit (expressed as deviance) when you have no predictors
- reliability of group outcome means

RANDOM EFFECTS

The one-way ANOVA model is a random effects model because "the group effects are construed as random" (B&R p. 17).

MEANS AS OUTCOMES REGRESSION (MAOR)

The next variety of model is MAOR (B&R 18, 64-66; for a detailed example see pp. 86-92)

In this model:

- you enter no Level 1 predictors
- you enter Level 2 predictors to predict estimated group outcome scores (B0Js)

<u>Advantages</u>

Why is this approach better than a standard "ecological correlation" analysis of group-level means as the outcome?

1. Precision weighting means small groups do not have undue influence

B&R (p. 92) suggest that if you have groups with varying sample sizes, it is possible that a L-2 unit with small n can become an outlier in your regression with strong leverage; in other words, it can have an undue influence on the regression estimates. HLM with its precision weighting protects against this.

2. No matter how strongly cases are clustered within units, you get proper estimates of standard errors, and therefore proper statistical tests

In addition, HLM takes into account within-group similarities, and adjusts your standard errors of fixed effects accordingly, resulting in smaller standard errors (p. 92). HLM provides "unbiased and efficient estimates of the fixed effects ... and provides proper standard error estimates, regardless of the degree of within unit clustering, that are more closely approximated by L-2 analysis" (B&R p. 94).

3. Better estimates of variance explained

When you are making judgments about the importance of a L-2 predictor, HLM is better than standard ecological correlation because
a) percent outcome explained is not influenced by the degree of clustering on Y within each L-2 unit.
b) percent outcome explained is not influenced by the unreliability of group outcome means: Y.j
c) You get the clearest evidence of how important a L-2 predictor is because you are focusing on variance in G00; i.e., T00. It is "true" outcome variance you are trying to explain with your L-2 predictors

<u>Equations</u>

In this submodel we have the following L-1 model (B&R, Eq. 2.6):

YIJ=B0J+RIJ Eq. 3.1

This is the same L-1 equation as the ANOVA model

We have the following L-2 model (B&R, Eq. 2.11):

B0J=G00 + G01WJ + U0J Eq. 3.2

Of course, you can have as many L-2 predictors as you wish:

B0J=G00 + G01W1J + G02W2J + U0J Eq. 3.3

and so on

In the case of one predictor this leads to the following combined model (B&R Eq. 2.12):

YIJ = G00 + G01WJ + U0J + RIJ Eq. 3.4

Notice the following features about these equations

- The L-2 equation differs from a standard ecological analysis (noted above) because outcome is "true" parameter variance

- In the combined model, again, error is decomposed into separate parts: the unique effects of each L-2 unit, summed across units, on "true" outcome means **after controlling for L-2 fixed predictors**, (U0J) and the unique within-group variation (RIJ).

- **Note espec. how U0J has changed**. Whereas in the ANOVA model it represented simply the unique variation across L-2 units associated with true outcome means, now it represents unique variation across L-2 units associated with true outcome means **after controlling for other L-2 predictors**.

Hypothesis test of U0J

This last point gives us a clue about when to stop adding L-2 predictors.

Recall that in the ANOVA model we can test the hypothesis that the variance of the U0Js (T00); i.e., the L-2 differences in true means around the true grand mean, are significantly different from 0.

An important question becomes: **after** I have added a L-2 predictor, does the variance of U0j, i.e., T00, remain significantly different from 0?

In short, we have a hypothesis test that the residual T00 remains significantly different from 0.

If at any point, after adding X many L-2 predictors, we fail to reject the null hypothesis that the variance of T00 = 0, **we should stop adding L-2 predictors**

COMMENT ON DATA TO FOLLOW
The example below comes from a 1981 survey of 1,622 residents spread across 66 Baltimore city neighborhoods; respondents were household heads of spouses of household heads. BIGFEAR was a multi-item fear of crime index where higher scores indicated greater fear of crime. PENDROB was the neighborhoods percentile score (1 - 100) with a higher score indicating being higher a higher crime neighborhood. Those in the 100^{th} percentile were the most crime ridden. P80BLACK refers to the percent of the neighborhood population that was African-American, based on census rather than survey data , in 1980.

An example using BIGFEAR as an outcome

Comparing results of One-Way ANOVA vs. MAOR

Parameter	ANOVA	MAOR W=PENDROB	MAOR W1=P80BLACK	MAOR W1=PENDROB W2=P80BLACK
GOO	1.732	1.510	1.556	1.48
T00 (Var(U0J))	.062	.036	.032	.027
DEVIANCE	3245	3233	3228	3233
G01		.005	.004	.003
G02		--	--	.003
Reliability	.788	.687	.66	.618

Notice

What happens to grand mean true score (G00) and its interpretation?
Between variance and its interpretation
#What happens to reliability and why?

To develop a PRE measure of variance explained (B&R, p. 65) we proceed as follows:

$$\text{Proportion variance explained in B0J} = \frac{\text{T00 Random ANOVA} - \text{T00 MAOR}}{\text{T00 Random ANOVA}}$$

So, comparing the first and second columns we would say

PROP. EXPLAINED = [(.062 - .036) / .062] = .419

REMEMBER this is not just between-group variance we are explaining. Rather we are explaining variance in between group **true scores** on the outcome.