

## NOTES ON HLM TERMINOLOGY

by

Ralph B. Taylor

[breck@rbtaylor.net](mailto:breck@rbtaylor.net)

All materials copyright (c) 1998-2002 by Ralph B. Taylor

### LEVEL 1

Refers to the model describing units within a grouping: e.g., students in a school, residents in a neighborhood, officers in a precinct.

When modeling changes over time, however, LEVEL 1 refers to time periods over which an individual or unit is observed. For those of you who know ANOVA, it is the factor on which you have repeated measures where the independent variable refers only to different time points.

### LEVEL 2

This has a specific meaning within HLM, referring to the variation across units. This model can take several different forms, ranging from a simple one-way ANOVA describing how much variation in the outcome is between units, to much more complex models.

#### LEVEL 1 MODEL

$$Y_{IJ} = B_{0J} + B_{1J}X_{1I} + R_{IJ}$$

Eq. 1.1

Model extends to multiple predictors

### NOTATION IN LEVEL 1 MODEL

$Y_{IJ}$

Score of individual I in group J on the outcome

$B_{0J}$

The mean score of each Jth unit (group) on the outcome (Y) variable when all X variables = 0. In B&R (p. 31):

$$B_{0J} = E(Y_{IJ} | X_{IJ}=0)$$

Eq. 1.2

This meaning changes if we center predictors: see below.

This is analogous to but different from the intercept in a standard OLS

regression

$$Y = A + BX + E \quad \text{Eq. 1.3}$$

where

$$\begin{aligned} A &= E(Y_{IJ}|X=0) \\ &= Y_{\text{predicted}IJ} - B_1(X_1) \end{aligned}$$

It is different in two ways. First, you have an intercept for each group in the analysis. It is your best estimate of the group mean when all predictors score 0. (The meaning of scoring 0 varies depending upon whether you center, and if so how; see below). This may take on a very particular meaning when you have one or more dummy variable predictors. I.e., it may refer to the reference string. Again, however, this changes if you center predictors.

Second, the theoretical interpretation of  $B_{0J}$  depends upon how you have done the centering of predictors. See B&R p. 31. There are two ways to center or adjust the predictors that are of interest to us here.

## TANGENT ON CENTERING

More specifically:

IMAGINE A DATA SET LIKE THIS: We have 25 residents in each of 66 neighborhoods.  $X$  = age in years and  $Y$  = fear. In OLS regression, using surveys from all residents in all different neighborhoods,

$A$  = predicted fear score of an individual who is 0 years of age

Imagine in your OLS regression that you had centered age around the mean age of the sample. Say this grand mean ( $X_{..}$ ) = 42.

Now your OLS regression equation is:

$$Y_{\text{predicted}} = A + B(X - X_{..}) \quad \text{Eq. 1.4}$$

Now your intercept represents something different: the predicted fear score for an average-aged respondent of 42 years. Because:

$$Y_{\text{predicted}} = A + B(42 - 42) \quad \text{Eq. 1.5}$$

This is called **grand mean centering**.

But with HLM we have an intercept for each group. So with grand mean centering your group intercept (B&R p. 26) is the **adjusted mean for group J**:

$$B_{0J} = U_{yj} + B_{1J}(X_{.j} - X_{..}) \quad \text{Eq. 1.6}$$

where

$U_{yj}$  represents how different group  $J$  is from the grand mean on the  $Y$  outcome variable. It is the unique increment on  $Y$  associated with group  $J$ .

B0J = expected outcome for a person whose X score is equivalent the grand mean (X..) on the predictor. "This is the standard location for XIJ in the classical ANCOVA model" (p. 26).

The mean for group J has been adjusted by the mean for the entire sample; in other words, it is expressed as a deviation from the mean for the entire sample

Why would you want to do this grand mean centering? First, it does make the intercept somewhat more meaningful. A now equals the average predicted fear score for a 42 year old, your average aged respondent. In HLM B0J represents the average fear of a respondent in group J whose age is equivalent to the age for the overall sample.

Secondly, you may be interested in the contrasts between individuals and the "average" person in the sample for some theoretical purposes. Third, it may decrease multicollinearity between Level 1 predictors and Level 2 predictors, a topic we will talk about in a few weeks. Multicollinearity causes estimation problems in HLM just as it does in OLS. You also have the problem of collinearity between slopes (B1s) and intercepts (B0s) that can be affected by centering.

The variance of B0J represents the variance in Level 2 unit adjusted means on the outcome.

A second approach is called **group mean centering**.

Here, you focus on the contrast between each individual and the mean of his/her group.

Your predictor is not X but rather:

$$X - X.J \qquad \qquad \qquad \text{Eq. 1.7}$$

In ANOVA terminology you are pooling the within group variance on the predictor, and "throwing away" the between group variance on the predictor. Cronbach (1976) discusses the reasons why you might want to do this.

The variance of B0J represents the variance in Level-2 unit mean scores on the outcome.

For example, in a schooling situation, where your Level 1 units are students, and your Level 2 units are schools, you might be interested in focusing solely on how far above or below the school average a student is on predictor X, and how that affects his/her scores on an outcome. In our fear-age example, you may have theoretical reasons for contrasting residents with their "average" neighbor. In a relatively elderly neighborhood, it will be hard for respondents to have an extremely high score on a group-mean centered age variable. But in a relatively young neighborhood, an elderly person would have an extreme score on a group-mean centered age variable.

R's age	Neighborhood Average in Sample	Group Mean Centered X
80	45	35

80

30

50

In short, in the second case the respondent is more "deviant" in the sense that he/she is more distant from the typical respondent. An older person sticks out a lot more in a student ghetto than in a neighborhood of older families.

Note here that you have transformed your predictor theoretically. It now serves to contrast the individual and the group. It is no longer a measure of the individual. It reflects context.

Shinn calls these frog pond effects.

If you do OLS regression, with a clustered sample, and only include group-mean centered predictors, your results will reflect only pooled within-group variance in the outcome variable. The between group variance in the outcome variable is not touched.

The HLM program will give you opportunities to center your predictors. It is important to think carefully about what you are doing when you answer this question. This is more of a theoretical question than it is an operational one.

## PROGRAM HINT

So if you center your predictor (or predictors) by the group mean, what does B0J represent?

Each group's intercept now represents **the unadjusted mean on the outcome for group J**.

$$B0J = U_{yj}$$

Eq. 1.8

IN SUM: you can interpret B0J only if you know what X=0 means for all your predictors.

B1J

Each B1J represents the slope of your outcome on your predictor FOR EACH OF THE J GROUPS.

HLM runs an OLS regression for each group, estimating the regression equation.

NOTE: even though HLM calls these "betas" we interpret them like we would interpret B or unstandardized weights in a regression. It tells you for every UNIT (not standard deviation) increase on that X variable, IN THAT GROUP, how many units do scores on the outcome increase or decrease?

You need to be conscious of the choice of "location" for each Level-1 predictor, because this influences, as discussed above, the interpretation of

the intercept, AND because it also influences the interpretation of the slope. (B&R, p. 28). This is especially true for dummy variables.

## DUMMY

Say you have 1 dummy variable for elderly ( 1 = age > 64) and nonelderly (0 = <= 64).

If you do NOT center that predictor, each  $B_{1j}$  is telling you about the difference between elderly and nonelderly in each  $J$  group, on the outcome.

Note now that when  $X=0$ , that is you are talking about nonelderly, the expected value for  $Y$  is  $B_{0j}$ .

Suppose you grand mean center your age dummy? If  $R$  is elderly  $X = (1 - X_{..}) =$  the proportion of NONelderly in the sample.

If  $R$  is NOT elderly  $X = (0 - X_{..}) =$  minus the proportion of elderly in the sample.

What happens if you use group mean centering?

Now  $X = X - X_{.j}$

For elderly (raw value of 1) it reflects the proportion of NONelderly in the group or neighborhood.

For nonelderly it reflects minus the proportion of elderly in the group or neighborhood.

REMEMBER WHEN YOU HAVE MULTIPLE DUMMIES, you interpret the intercept as reflecting scores for those who are 0 on all dummies.

So if you have two predictors, and they are both dummies

X	0	1
Sex	Female	Male
Age	Non-Elderly	Elderly

and you have not centered them, the intercept refers to the average outcome score, in the  $J$ th group, for non-elderly women.

Notice how theory comes into play here. If you think that non-elderly males are the least afraid, you might set up your dummy coding so that the reference string corresponds to the group you expect to be lowest scoring, e.g.,

X	0	1
Sex	Male	Female
Age	Non-Elderly	Elderly

We can treat  $B_{ij}$  a number of different ways at Level-2; we will get to this.

## RIJ

This reflects the within-unit error. HLM assumes (R&B: 12)

```
# RIJ is normally distributed within each unit
# The variance of RIJ is relatively homogeneous across units
```

We will see that this gets 'taken apart' when we look at the Level 2 and combined models

## VARIANCE OF PARAMETERS

$\text{VAR}(B0J) = T00$

The variance of the intercepts is captured in T00. It is the dispersion of the group means.

HLM tests to see if this variance is significantly different from zero. WHAT DOES IT MEAN IF THE VARIANCE OF THE B0Js IS ZERO OR NOT SIGNIFICANTLY DIFFERENT FROM ZERO?

It means that there are no sizable between-unit differences on outcome scores.

This is a population variance.

$\text{VAR}(B1J) = T11$

The variance of the parameters capturing unit slopes is represented in T11. You can constrain these slopes to all be the same, in which case there would be no variance in the slopes. This is what you do in standard OLS. Or you can allow them to vary, which is what we usually do in HLM.

This is a population variance

$\text{COV}(B0J, B1J) = T01$

It is not unusual to have slopes and intercepts, across the J many groups, vary together. This is captured as a covariance T01. It also gets reported in the output as a correlation.

This is a population variance

Can we think about situations where this might happen?

If the correlation becomes too high, that creates a problem for HLM, and it will have difficulty estimating. We will talk about how to proceed when this happens.

$\text{VAR}(RIJ) = \sigma \text{ squared}$

This is the variance of the within group variability.

## THINKING ABOUT OUTCOME VARIANCE

Remember back to our plain outcome variable,  $Y_{IJ}$ ?

$$\text{VAR}(Y_{IJ}) = T_{00} + \sigma^2$$

Which leads us to the very important topic of **variance decomposition**.