

Hlml09

RESIDUAL FILES IN HLM

by

Ralph B. Taylor

breck@rbtaylor.net

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

HOW TO CREATE

You create a residual file by issuing the command for one. This is under the basic command specifications. If you are doing a 3 level model, you can request both 2 level and 3 level residual files. Since you are just doing a level 2 model, you can only ask for a level 2 residual file.

From the menu it looks like it can add in additional variables, beyond those in the model, but in HLM4.02a this does not always work.

Notice when the residual menu comes up you can tell it the type of file you want to create. Change SYSTAT - the default option - to SPSS. Also you will need to enter a name for the file - it is by default set to RESFIL.CMD for the two level file. You want to call it BLABLARE.SPS because what you are doing is not creating a residual file per se; rather you are creating an SPSS command file, with the data embedded within it, that will be used to create the residual file.

RESFIL.SPS

This file really has **three** parts:

- * information about what variables are where, so that SPSS or can read this information and create a system file;
- * some predicted values.
- * actual residuals from the analysis you have run. See. pp. 30 on in the manual.

WHAT IS IN IT

The file starts with a bunch of file definition information, then goes on to include:

a list of variables

format statements for variables

recodes for missing values.

then the actual data in the residual file.

There is a row of data for each Level 2 unit.

CREATING THE SYSTEM FILE

Start up SPSS. Bring BLABLARE.SPS into a syntax box. At end of command be sure to give it the appropriate name for the data file it will save. Highlight all the syntax and run it. You should have the new data file in your data window.

Hlml09

THE VARIABLES IN THE RESIDUAL FILE

ID\$ ID number for the L-2 unit
NJ number of L-1 cases in corresponding L-2 unit
CHIPCT "expected values of the order statistics for a sample size J selected from a population that is distributed Chi square(v)" where v=number of random effects per unit (manual, p. 32)

My best guess is that these are expected values for lack of fit, for the overall model, for each Jth unit.

MDIST These are Mahalanobis distances (D squared) ("standardized squared distance from the center of a v-dimensional distribution, where v is the number of random effects per unit. Essentially, MDIST provides a single, standardized summary measure of the distance of a unit's EB [empirical Bayes] estimates, B*, from its 'fitted value' $G_{j0} + G_{js}W_{sj}$ " (manual, p. 32)

These are standardized measures of lack of fit. They might be of inherent interest in and of themselves.

THESE NEXT THREE MEASURES ARE MEASURES OF TOTAL VARIANCE (LNTOTVAR) AND RESIDUAL VARIATION

LNTOTVAR "natural log of total standard deviation within each L-2 unit" (manual, 32). This is the total standard deviation on **observed Y**.

NOTE: these can be negative, suggesting standard deviations of less than one.

OLSRVAR "natural log of residual standard deviation within each unit based on its least squares regression." (manual, p. 33) In other words, the standard deviation of raw (unstandardized) OLS residuals in each L-2 unit, logged. If you did OLS, this is your measure of residual variance.

MDRSVAR "natural logarithm of the residual standard deviation from the final fitted fixed effects model." (manual, p. 33) In other words, the log of the standard deviation of the residuals after HLM has fit its best model.

THESE NEXT TWO ITEMS GIVE YOU THE VALUE FOR EACH L-2 UNIT OF ITS RESIDUAL WHEN TRYING TO PREDICT INTERCEPTS:

HOW MUCH THE FITTED VALUE FOR THE INTERCEPT DISAGREED WITH THE ACTUAL. These are different ways of getting at the U0Js

EBINTRCP Empirical Bayes best estimate of the residual L-2 value - these are your estimates of U0J:

U_{0j} .

Note these have an average of zero, as residuals must axiomatically. Note the standard deviation of EBINTRCP is probably smaller than the standard deviation of OLINTRCP; the estimates are somewhat shrunken.

H1m109

OLSINTRCP OLS estimates of L-2 residual value: U0J. Note these have an average of zero, as they must.

IF IN YOUR RUN YOU ALLOWED SLOPES TO VARY ACROSS YOUR LEVEL 2 UNITS YOU WILL ALSO GET THINGS THAT TELL YOU ABOUT YOUR U1JS - FOR EXAMPLE

EBZAGE Empirical Bayes estimate of the discrepancy between the average slope for z scored age, and the slope in each L-2 unit. This is:

U*1J

Will be dramatically shrunken when compared to OLS estimate of same which is:

OLZAGE OLS estimate of same: U1J