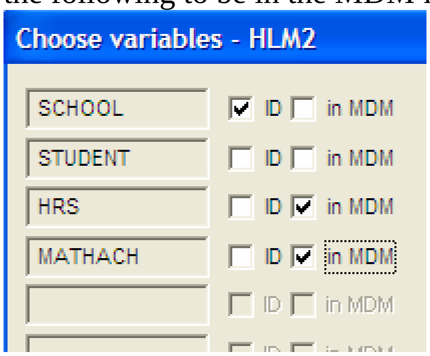
# Dataset 2 – Fixed versus Random Coefficients and Level-2 Predictors.

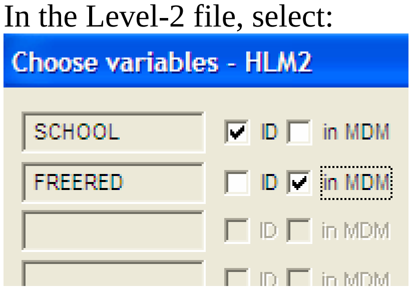
For use with HLM

The design of this dataset is to show how adding a random coefficient to a model dramatically increases data fit even though the coefficients to the slope coefficient are not statistically significant. A level-2 predictor is also added to show the mediation of this effect. The level-1 and level-2 datasets that will be used in this analysis are called "fixrand_L1.sav" and "fixrand_L2.sav" respectively.

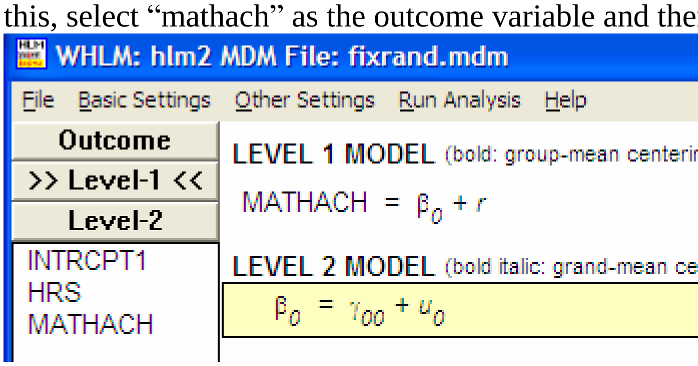
First, we will setup the .MDM file so that we can run the null model. In the Level-1 file, select the following to be in the MDM file:




In the Level-2 file, select:




We will first fit the null/unconditional model in which only the intercept is estimated. To do this, select "mathach" as the outcome variable and then run this model:

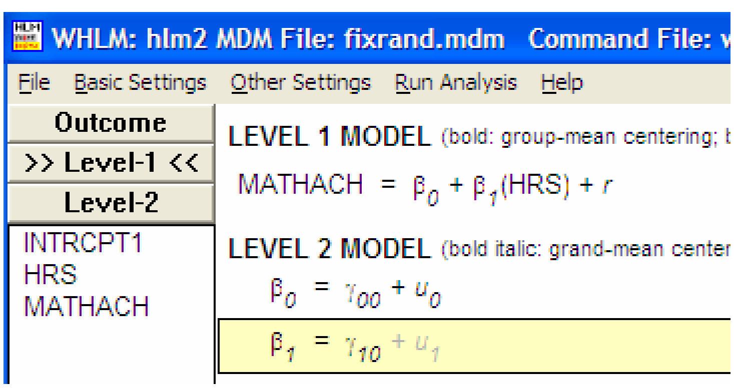



In viewing the output we can see the following (abridged);

Roberts, J. K. (2005, April). Datasets illustrating specific strengths of hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

       Y = B0 + R

Level-2 Model
       B0 = G00 + U0

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                        Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          77.540000   1.134920    68.322         9    0.000
 ----------------------------------------------------------------------------

Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        2.70982       7.34311     9      20.93499    0.013
  level-1,       R         7.44133      55.37333
 -----------------------------------------------------------------------------

 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 688.710612
 Number of estimated parameters = 2
```

Next we will add a fixed effect that has hours of homework that a student has completed each week. We do this by specifying “hrs” as an uncentered level-1 predictor. We do not center hours because it has a meaningful zero (no hours of homework completed each week). Specifying this will produce the following:

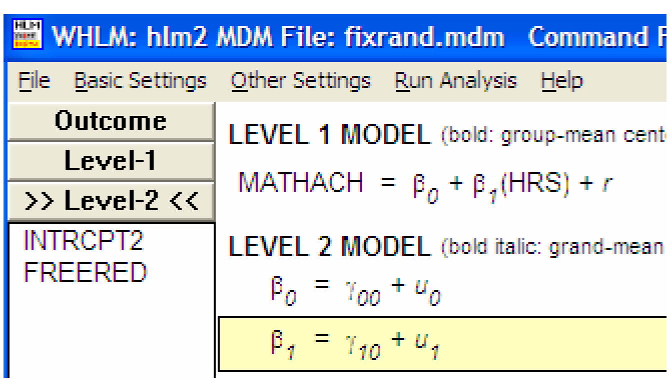


,

with results (abridged):

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

        Y = B0 + B1*(HRS) + R

Level-2 Model
        B0 = G00 + U0
        B1 = G10

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          77.441818   1.631933    47.454         9    0.000
 For      HRS slope, B1
    INTRCPT2, G10           0.021818   0.260511     0.084        98    0.934
 ----------------------------------------------------------------------------

Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        2.69958       7.28772     9      20.70465    0.014
  level-1,       R         7.48260      55.98937
 -----------------------------------------------------------------------------


 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 691.399309
 Number of estimated parameters = 2
```

These results would tell us that on average if a student moves from doing no homework each week to doing one hour of homework each week, their score on the math achievement variable would increase by 0.02 points. This would not seem to be a meaningful outcome, but we may recognize that further analysis is warranted. Because we forced a fixed slope with no random effects, the predicted regression lines are the same for each group. Because of the increase in AIC and BIC, we might be inclined to say that the predictor “hrs” is not a meaningful addition to the model. However if we add the random coefficient for “hrs” we can see that the model fit increases by a dramatic amount. We can do this simply by clicking on the random component associated with the “hrs” variable ($u_1$) and running the analysis again.




Roberts, J. K. (2005, April). Datasets illustrating specific strengths of hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

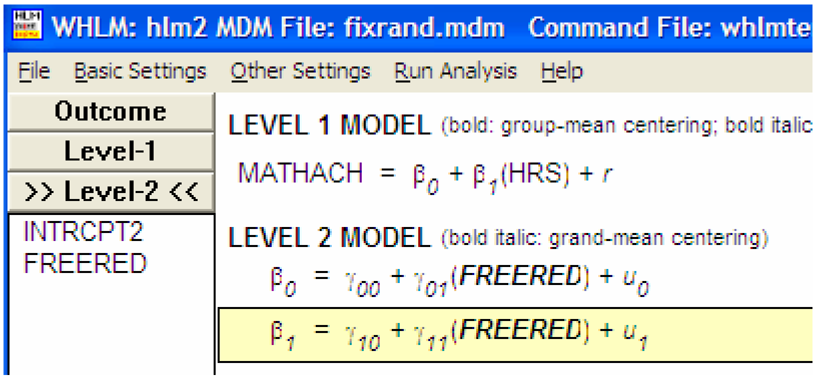
Running this model will produce the following results:

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

       Y = B0 + B1*(HRS) + R

Level-2 Model
       B0 = G00 + U0
       B1 = G10 + U1

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
   INTRCPT2, G00         77.441818   4.783157    16.191         9    0.000
 For      HRS slope, B1
   INTRCPT2, G10          0.021818   0.818612     0.027         9    0.980
 ----------------------------------------------------------------------------

Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0       15.12460      228.75339     9   63348.04068    0.000
      HRS slope, U1        2.58846        6.70011     9   52881.62319    0.000
  level-1,       R         0.30674        0.09409
 -----------------------------------------------------------------------------


 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 178.171395
 Number of estimated parameters = 4
```

Notice that the fixed effect for “hrs” is not statistically significant if we consider just at the statistical significance of the p-value. Outside of just looking at the amount of reduction in the Deviance statistic (simply the Chi-square), is to consider the AIC (Akaike Information Criteria) and the BIC (Bayesian Information Criteria). The AIC is computed by $AIC = -2 * \ell + 2K$ where $K$ is the number of parameters estimated and $-2 * \ell$ is an estimate of the Chi-Square. The BIC is computed by $BIC = -2 * \ell + K * Ln(N)$ where $Ln(N)$ is the number of people in the study. What is dramatic is to see the amount of reduction in the AIC and BIC which reduced from 699.40 to 190.17 and from 709.74 to 205.68 respectively. This means that we are fitting a much better model in the model with random slopes than in model with only fixed effects for the slopes.

Although the variance for the intercept has actually increased, the random component of the “hrs” variable has effectively reduced our level-1 variance (labeled “Residual”) to almost nil.

Now we can turn our attention to finding a variable to include in our model that explains the additional variance in the intercept and slope coefficients. To do this, we are going to add both the fixed effect and cross-level interaction for a school-level variable that tells the percentage of students on free/reduced lunch at that school. Since "hrs" is a variable that has a meaningful zero, we are not going to center it. We are, however, going to center our free/reduced lunch variable and will therefore grand mean center the variable "freered". To do this, we would run the following model:




Doing this will produce the following results (abridged):

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

        Y = B0 + B1*(HRS) + R

Level-2 Model
        B0 = G00 + G01*(FREERED) + U0
        B1 = G10 + G11*(FREERED) + U1

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          77.441818   0.576522   134.326         8    0.000
     FREERED, G01           1.636951   0.066197    24.728         8    0.000
 For      HRS slope, B1
    INTRCPT2, G10           0.021818   0.124513     0.175         8    0.866
     FREERED, G11          -0.279068   0.014297   -19.520         8    0.000
 ----------------------------------------------------------------------------

Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,        U0       1.81419       3.29127     8     818.05552    0.000
      HRS slope,  U1       0.39229       0.15390     8    1087.49710    0.000
  level-1,        R        0.30674       0.09409
 -----------------------------------------------------------------------------


 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 150.545856
 Number of estimated parameters = 4
```

Roberts, J. K. (2005, April). Datasets illustrating specific strengths of hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Let's now look at the Empirical Bayesian estimated coefficients for the first random effects model and the random effects model with the second-level effect and the cross-level interaction.

Fixed values for the model with just "hrs" having both fixed and random effects.

| | ecintrcp | echrs |
|---|---|---|
| 1 | 99.853 | -3.944 |
| 2 | 94.852 | -2.945 |
| 3 | 90.142 | -2.055 |
| 4 | 84.849 | -.948 |
| 5 | 79.928 | -.472 |
| 6 | 74.927 | .527 |
| 7 | 69.859 | 1.057 |
| 8 | 65.150 | 1.947 |
| 9 | 60.148 | 2.945 |
| 10 | 54.709 | 4.107 |

In order to obtain the values for the Bayesian estimated effects in the model with Free/Reduced Lunch added, we must do a little calculation as HLM does not readily give them to us. The first two columns in the below screen shot from the SPSS residuals file are variables that are generated by HLM. We can then find the actual Bayesian estimate by adding 77.441 (the grand intercept from the output above) to each of the values in the "ebintrcp" column. These are reflected in the computed column "int_new". Likewise, we can find the estimates for the slope of "hrs" by taking our variable "ebhrs" and adding the grand slope estimate (0.022). These corrected values are reflected in the computed column "hrs_new".

| | ebintrcp | ebhrs | int_new | hrs_new |
|---|---|---|---|---|
| 1 | 1.935 | -.475 | 79.38 | -.45 |
| 2 | 1.842 | -.313 | 79.28 | -.29 |
| 3 | .419 | .017 | 77.86 | .04 |
| 4 | -3.204 | .841 | 74.24 | .86 |
| 5 | -1.588 | .201 | 75.85 | .22 |
| 6 | -.061 | .086 | 77.38 | .11 |
| 7 | -.210 | -.222 | 77.23 | -.20 |
| 8 | -1.633 | .108 | 75.81 | .13 |
| 9 | 1.513 | -.283 | 78.95 | -.26 |
| 10 | .985 | .041 | 78.43 | .06 |

It is easy to see just by "eyeballing" the data that the final model reduces the variance in the intercept and the "hrs" variable by a considerable amount. We can also investigate this effect by looking at the variance components for both models.

As can be seen here, the residual variance associated with the intercept reduced from 228.75 to 3.29 and the residual variance associated with the "hrs" slope reduced from 6.70 to 0.15.

We now have our final model in which the second-level variable is able to mediate much of the variance produced by the level-1 variable "hrs".