

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

For use with either S-PLUS or R

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 dataset that will be used in this analysis is called “fixrand.sdd”.

Running the null model

```
> fixrand.null<-lme(mathach~1, data=fixrand, random=~1|school)
> summary(fixrand.null)
Linear mixed-effects model fit by REML
Data: fixrand
      AIC      BIC    logLik
696.5485 704.3338 -345.2742
```

Random effects:

```
Formula: ~ 1 | school
(Intercept) Residual
StdDev:    2.709817 7.441326
```

Fixed effects: mathach ~ 1

```
Value Std.Error DF t-value p-value
(Intercept) 77.54 1.13492 90 68.32198 <.0001
```

Standardized Within-Group Residuals:

```
Min      Q1      Med      Q3      Max
-2.696532 -0.4780901 -0.01380274 0.5041729 2.668927
```

Number of Observations: 100

Number of Groups: 10

Next we will add a fixed effect that has hours of homework that a student has completed each week.

```
> fixrand.one<-lme(mathach~hrs, data=fixrand, random=~1|school)
> summary(fixrand.one)
Linear mixed-effects model fit by REML
Data: fixrand
      AIC      BIC    logLik
699.3993 709.7392 -345.6997
```

Random effects:

```
Formula: ~ 1 | school
(Intercept) Residual
StdDev:    2.698394 7.48272
```

Fixed effects: mathach ~ hrs

```
Value Std.Error DF t-value p-value
(Intercept) 77.44182 1.631677 89 47.46148 <.0001
hrs 0.02182 0.260515 89 0.08375 0.9334
```

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.

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.671276	-0.477946	-0.01511379	0.5090011	2.670207

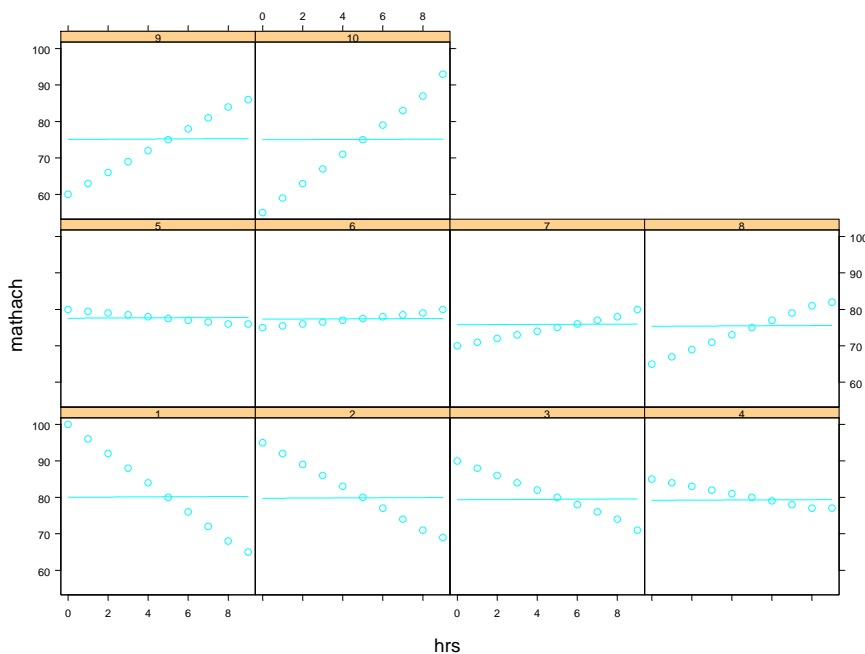
Number of Observations: 100

Number of Groups: 10

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.

We can see the plots of the predicted values of the fixed effects of the “hrs” slopes with the command:

```
plot(augPred(fixrand.one))
```



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.

```

> fixrand.two<-lme(mathach~hrs, data=fixrand, random=~hrs|school)
> summary(fixrand.two)
Linear mixed-effects model fit by REML
Data: fixrand
      AIC      BIC   logLik
190.1714 205.6812 -89.0857

Random effects:
Formula: ~ hrs | school
Structure: General positive-definite
           StdDev   Corr
(Intercept) 15.1245957 (Inter
           hrs  2.5884575 -0.998
Residual    0.3067424

Fixed effects: mathach ~ hrs
           Value Std.Error DF  t-value p-value
(Intercept) 77.44182  4.783157 89 16.19052  <.0001
           hrs  0.02182  0.818612 89  0.02665  0.9788

Standardized Within-Group Residuals:
           Min      Q1      Med      Q3      Max
-2.184688 -0.4472466 -0.06318906 0.3361167 4.334256

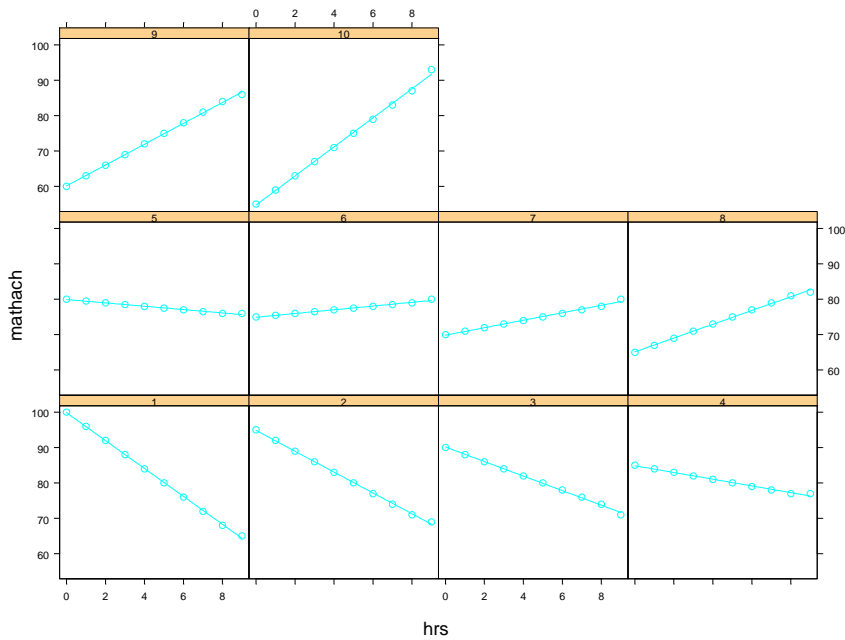
Number of Observations: 100
Number of Groups: 10

```

Notice that the fixed effect for “hrs” is not statistically significant if we look just at the statistical significance of the p-value. 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 “fixrand.two” than in “fixrand.one”.

The predicted regression lines would then be plotted with:

```
plot(augPred(fixrand.two))
```



This graph makes the dramatic point that this is the best model for interpreting our data. Although our 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.

```
> VarCorr(fixrand.one)
school = pdSymm(~ 1)
      Variance StdDev
(Intercept)  7.281332 2.698394
Residual    55.991094 7.482720

> VarCorr(fixrand.two)
school = pdSymm(~ hrs)
      Variance StdDev Corr
(Intercept) 228.75339486 15.1245957 (Intr)
      hrs    6.70011203  2.5884575 -0.998
Residual    0.09409091  0.3067424
```

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 use the “freeredc” variable in this analysis. Using the “*” command between “hrs” and “freeredc” will make the model fit both the two main effects and the interaction effect. To do this, we would run the following code:

```
> fixrand.three<-lme(mathach~hrs*freeredc, data=fixrand, random=~hrs|school)
> summary(fixrand.three)
Linear mixed-effects model fit by REML
Data: fixrand
      AIC      BIC    logLik
166.5459 187.0606 -75.27293

Random effects:
Formula: ~ hrs | school
Structure: General positive-definite
           StdDev  Corr
(Intercept) 1.8141855 (Inter
           hrs 0.3922948 -0.89
Residual 0.3067424

Fixed effects: mathach ~ hrs * freeredc
           Value Std.Error DF   t-value p-value
(Intercept)  77.44182  0.5765217  88   134.3259 <.0001
           hrs    0.02182  0.1245134  88     0.1752  0.8613
           freeredc  1.63695  0.0661969   8    24.7285 <.0001
hrs:freeredc  -0.27907  0.0142968  88   -19.5196 <.0001

Standardized Within-Group Residuals:
           Min           Q1           Med           Q3           Max
-2.164161 -0.4224458 -0.06901339  0.3547929  4.313095

Number of Observations: 100
Number of Groups: 10
```

Let’s now look at the Empirical Bayesian estimated coefficients for the first random effects model (fixrand.two) and the random effects model with the second-level effect and the cross-level interaction (fixrand.three).

```

> coef(fixrand.two)
  (Intercept)      hrs
1    99.85330 -3.9437244
2    94.85199 -2.9451175
3    90.14236 -2.0550744
4    84.84935 -0.9479036
5    79.92789 -0.4720571
6    74.92657  0.5265498
7    69.85927  1.0566783
8    65.14964  1.9467214
9    60.14833  2.9453283
10   54.70947  4.1067810

> coef(fixrand.three)
  (Intercept)      hrs freeredc hrs:freeredc
1    79.37716 -0.45360068 1.636951  -0.2790675
2    79.28388 -0.29163643 1.636951  -0.2790675
3    77.86077  0.03864902 1.636951  -0.2790675
4    74.23814  0.86241874 1.636951  -0.2790675
5    75.85423  0.22296432 1.636951  -0.2790675
6    77.38069  0.10821969 1.636951  -0.2790675
7    77.23210 -0.20033199 1.636951  -0.2790675
8    75.80899  0.12995346 1.636951  -0.2790675
9    78.95517 -0.26150005 1.636951  -0.2790675
10   78.42705  0.06304574 1.636951  -0.2790675

```

It is easy to see just by “eyeballing” the data that the “fixrand.three” 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.

```

> VarCorr(fixrand.two)
school = pdSymm(~ hrs)
      Variance      StdDev      Corr
(Intercept) 228.75339486 15.1245957 (Intr)
      hrs      6.70011203  2.5884575 -0.998
Residual    0.09409091  0.3067424

> VarCorr(fixrand.three)
school = pdSymm(~ hrs)
      Variance      StdDev      Corr
(Intercept)  3.29126895  1.8141855 (Intr)
      hrs    0.15389525  0.3922948 -0.89
Residual    0.09409091  0.3067424

```

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”.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.