Multilevel Models: Longitudinal Data

Read in and Prepare Data for these Notes

```
## Read in the Minneapolis reading data
> mplsWide = read.csv(file = "~/Data/minneapolis.csv")
> mplsWide


   studentID read.5 read.6 read.7 read.8 atRisk female minority ell sped  att
1          1    172    185    179    194      1      1        1   0    0 0.94
2          2    200    210    209     NA      1      1        1   0    0 0.91
3          3    191    199    203    215      1      0        1   0    0 0.97
4          4    200    195    194     NA      1      1        1   0    0 0.88
5          5    207    213    212    213      1      1        1   0    0 0.85
6          6    191    189    206    195      1      0        1   0    0 0.90
7          7    199    208    213    218      1      0        1   1    0 0.97
8          8    191    194    194     NA      1      1        1   1    1 0.97
9          9    149    154    174    177      1      1        1   0    1 0.97
10        10    200    212    213     NA      1      1        1   0    0 0.96
11        11    218    231    233    239      1      1        0   0    0 0.98
12        12    228    232    248    246      1      1        0   0    1 0.96
13        13    228    236    228    239      0      1        0   0    0 0.99
14        14    199    210    225    235      0      0        1   1    0 0.99
15        15    218    223    236     NA      0      1        0   0    0 1.00
16        16    228    226    234    227      0      0        0   0    0 0.97
17        17    201    210    208    219      0      0        0   0    0 0.98
18        18    218    220    217    221      0      0        0   0    0 1.00
19        19    215    216    221     NA      0      1        0   0    0 0.96
20        20    204    215    219    214      0      1        1   0    0 0.95
21        21    237    241    243     NA      0      0        0   0    0 0.98
22        22    219    233    236     NA      0      1        1   0    0 0.96
```

The wide format is good for computing summaries and variance-covariances or correlations between the repeated measures

```
> summary(mplsWide[2:5])

     read.5           read.6           read.7           read.8     
 Min.   :149.0   Min.   :154.0   Min.   :174.0   Min.   :177.0  
 1st Qu.:199.0   1st Qu.:201.2   1st Qu.:206.5   1st Qu.:213.2  
 Median :202.5   Median :212.5   Median :215.0   Median :218.5  
 Mean   :205.1   Mean   :211.5   Mean   :215.7   Mean   :218.0  
 3rd Qu.:218.0   3rd Qu.:225.2   3rd Qu.:231.8   3rd Qu.:233.0  
 Max.   :237.0   Max.   :241.0   Max.   :248.0   Max.   :246.0  
                                                 NA's   :8      
> sapply(mplsWide[2:5], mean, na.rm = TRUE)
> sapply(mplsWide[2:5], sd, na.rm = TRUE)

  read.5   read.6   read.7   read.8 
19.99356 20.06116 19.44562 19.37881 
```

The means of the repeated measures show an increasing mean reading score over grade level. The SDs suggest that each of the repeated measures have roughly the same degree of variation (equal variances?). Later time points show more missing data (common).

```
> var(mplsWide[2:5], use = "pairwise.complete.obs")

          read.5   read.6   read.7   read.8
read.5 399.7424 386.8874 355.3788 383.5385
read.6 386.8874 402.4502 361.1039 396.4615
read.7 355.3788 361.1039 378.1320 361.6923
read.8 383.5385 396.4615 361.6923 375.5385


> cor(mplsWide[2:5], use = "pairwise.complete.obs")

          read.5    read.6    read.7    read.8
read.5 1.0000000 0.9645804 0.9140704 0.8825860
read.6 0.9645804 1.0000000 0.9256661 0.9287729
read.7 0.9140704 0.9256661 1.0000000 0.9227732
read.8 0.8825860 0.9287729 0.9227732 1.0000000
```

The correlations show that the repeated measures are correlated. They also suggest that repeated measures that are further apart in time are less correlated (decay).

Long Format

```
## Reshape the data to the long format
> library(reshape2)

mpls = melt(
    mplsWide,
    id = c("studentID", "atRisk", "female", "minority", "ell", "sped", "att"),
    measure = c("read.5", "read.6", "read.7", "read.8")
    )

## The arrange() function from dplyr is used to order by studentID
> library(dplyr)
> mpls = mpls %>% arrange(studentID)

> head(mpls, 12)

   studentID atRisk female minority ell sped  att variable value
1          1      1      1        1   0    0 0.94   read.5   172
2          1      1      1        1   0    0 0.94   read.6   185
3          1      1      1        1   0    0 0.94   read.7   179
4          1      1      1        1   0    0 0.94   read.8   194
5          2      1      1        1   0    0 0.91   read.5   200
6          2      1      1        1   0    0 0.91   read.6   210
7          2      1      1        1   0    0 0.91   read.7   209
8          2      1      1        1   0    0 0.91   read.8    NA
9          3      1      0        1   0    0 0.97   read.5   191
10         3      1      0        1   0    0 0.97   read.6   199
11         3      1      0        1   0    0 0.97   read.7   203
12         3      1      0        1   0    0 0.97   read.8   215
```

Two Issues: Variable names and grade-level is a factor

```
## Rename the 8th and 9th columns
> names(mpls)[8] = "grade"
> names(mpls)[9] = "read"


## Change grade predictor to an integer
> mpls$grade = as.integer(mpls$grade)

  studentID atRisk female minority ell sped  att grade read
1         1      1      1        1   0    0 0.94     1  172
2         1      1      1        1   0    0 0.94     2  185
3         1      1      1        1   0    0 0.94     3  179
4         1      1      1        1   0    0 0.94     4  194
5         2      1      1        1   0    0 0.91     1  200
6         2      1      1        1   0    0 0.91     2  210

## Add 4 to get back to grade 5, 6, 7, and 8
> mpls$grade = mpls$grade + 4


  studentID atRisk female minority ell sped  att grade read
1         1      1      1        1   0    0 0.94     5  172
2         1      1      1        1   0    0 0.94     6  185
3         1      1      1        1   0    0 0.94     7  179
4         1      1      1        1   0    0 0.94     8  194
5         2      1      1        1   0    0 0.91     5  200
6         2      1      1        1   0    0 0.91     6  210
```

In Long Format, Rename Grade and Reading Score Variables

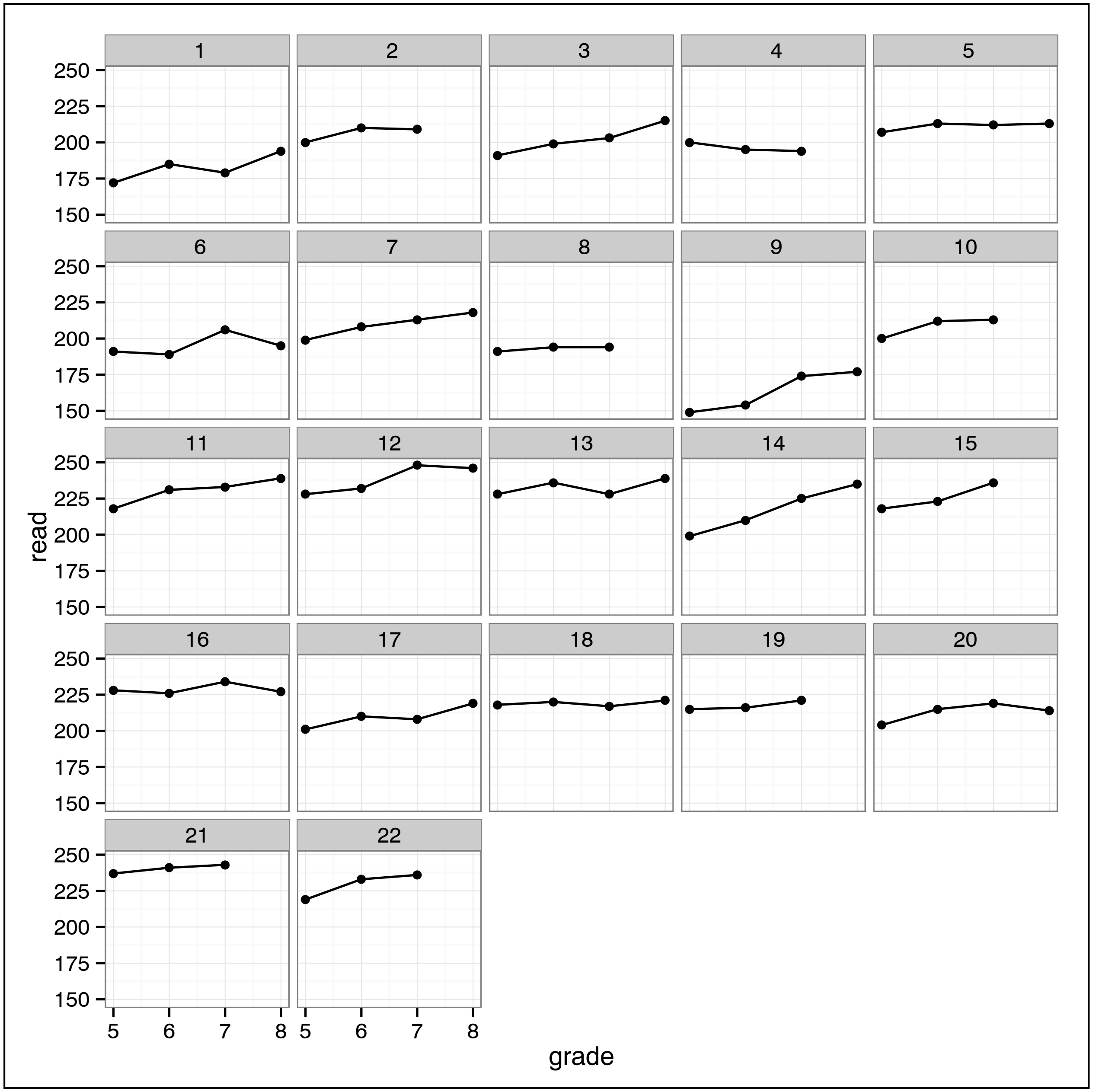
```
## Rename the 8th and 9th columns
> names(mpls)[8] = "grade"
> names(mpls)[9] = "read"


## Change grade predictor to an integer
> mpls$grade = as.integer(mpls$grade)
> mpls$grade = mpls$grade + 4


> mpls <- na.omit(mpls)

  subid risk gen eth ell sped  att grade read white
1     1    1   F Afr   0    N 0.94     5  172     0
2     1    1   F Afr   0    N 0.94     6  185     0
3     1    1   F Afr   0    N 0.94     7  179     0
4     1    1   F Afr   0    N 0.94     8  194     0
5     2    1   F Afr   0    N 0.91     5  200     0
6     2    1   F Afr   0    N 0.91     6  210     0

## Load libraries needed
> library(ggplot2)
> library(lme4)
```

The long format is good for plotting with ggplot and for carrying out the analysis with lmer

```
> head(mpls, 12)

   studentID atRisk female minority ell sped  att grade read
1          1      1      1        1   0    0 0.94     5  172
2          1      1      1        1   0    0 0.94     6  185
3          1      1      1        1   0    0 0.94     7  179
4          1      1      1        1   0    0 0.94     8  194
5          2      1      1        1   0    0 0.91     5  200
6          2      1      1        1   0    0 0.91     6  210
7          2      1      1        1   0    0 0.91     7  209
8          2      1      1        1   0    0 0.91     8   NA
9          3      1      0        1   0    0 0.97     5  191
10         3      1      0        1   0    0 0.97     6  199
11         3      1      0        1   0    0 0.97     7  203
12         3      1      0        1   0    0 0.97     8  215
```

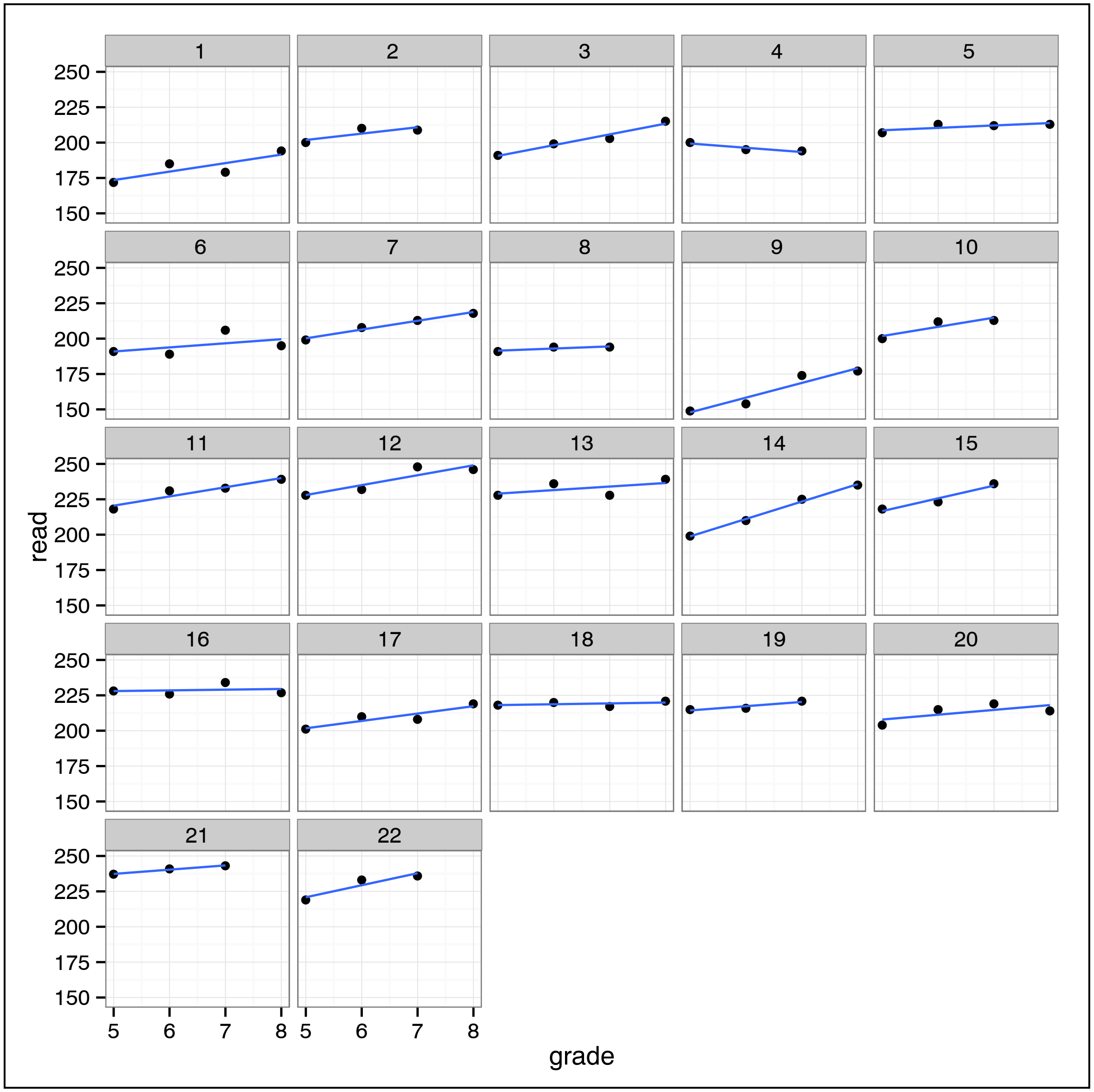
For these functions, there can only be one column for the outcome.

Spaghetti Plot of the Repeated Measures Grouped by ID



The change in reading scores seems like it could be modeled by a line.

Regression Lines Grouped by ID



Three features of these plots:

- Most seem approximately linear (but not always increasing over time)
- Some OLS trajectories fit very well (e.g., Subject #7)
- Other OLS trajectories show more scatter (e.g., Subject #13)

Unconditional Level-1 Model

Although we are interested in grade (time) predicting reading scores, we will first fit the unconditional varying intercepts model.

$$\text{Read}_{ij} = \beta_0^* + \epsilon_{ij}$$

$$\text{where} \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

$$\beta_0^* = \beta_{00} + \eta_0$$

$$\text{where} \quad \eta_0 \sim \mathcal{N}(0, \sigma_{\eta_0}^2)$$

The *i* subscript represents a particular time point, and the *j* subscript represents a particular person.

$$\text{Read}_{ij} = \beta_{00} + \eta_0 + \epsilon_{ij}$$

```
> library(lme4)
> model.a = lmer(read ~ 1  + (1 | studentID), data = mpls, REML = FALSE)
> summary(model.a)

Random effects:
 Groups    Name        Variance Std.Dev.
 studentID (Intercept) 319.3    17.870
 Residual               66.2     8.136
Number of obs: 80, groups:  studentID, 22

Fixed effects:
            Estimate Std. Error t value
(Intercept)  212.244      3.919   54.16
```

The mean reading score across students and grade levels is 212.2.

$$\rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2}$$

An estimated 83% of the total variation in reading scores is attributable to **differences between students**

$$\hat{\rho} = \frac{319.3}{319.3 + 66.2} = 0.83$$

An estimated 17% of the total variation in reading scores is attributable to differences **between time points** (within students)

Level-1 Model

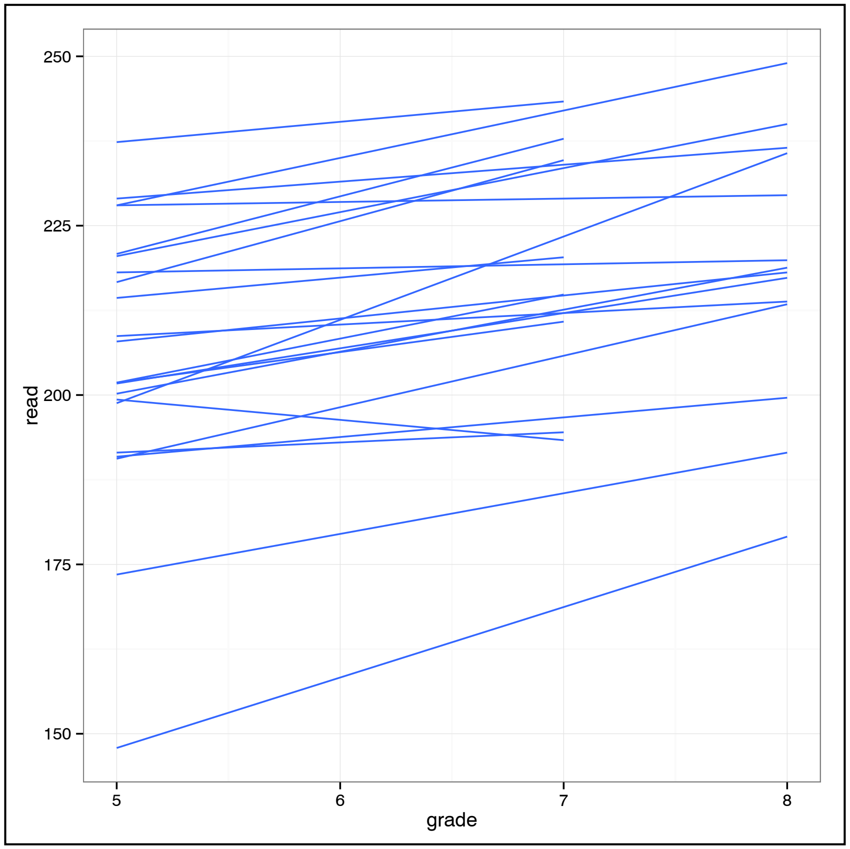
Having partitioned the total variation into within-students and between-students, we can now ask:
What role does time play?

$$\text{Read}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

$$\text{where} \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

We will use a line to model the level-1 functional form.

Specifying the Random Effects



There seems to be variation in both intercepts and slopes…

…we might want to allow both the intercepts and slopes to vary (include a random effect in each equation)

$$\beta_0^* = \beta_{00} + \eta_0$$
$$\beta_1^* = \beta_{01} + \eta_1$$

where

$$\begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\eta_0} & \sigma_{\eta_0,\eta_1} \\ \sigma_{\eta_1,\eta_0} & \sigma^2_{\eta_1} \end{bmatrix} \right)$$

The Level-2 residuals (the random effects) represent the deviations of each student's change trajectory around the average change trajectory.

$$\text{Read}_{ij} = \beta_{00} + \beta_{01}(\text{Grade}_{ij}) + \eta_0 + \eta_1(\text{Grade}_{ij}) + \epsilon_{ij}$$

```
> model.b = lmer(read ~ 1 + grade  + (1 + grade | studentID), data = mpls, REML = FALSE)
> summary(model.b)

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 740.469  27.212
           grade         6.966   2.639   -0.74
 Residual               18.315   4.280
Number of obs: 80, groups:  studentID, 22

Fixed effects:
            Estimate Std. Error t value
(Intercept) 181.3335     6.5614  27.636
grade         4.8823     0.7417   6.582

Correlation of Fixed Effects:
      (Intr)
grade -0.799
```

The mean reading score at grade = 0 is predicted to be 181.3.

Each grade-level difference is associated with a an average 4.9-point difference in reading score.

Grade level is an important predictor of the within-student variation (*Wald* $t = 6.58$).

Interpreting the Variance Components

```
Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 740.469  27.212
           grade         6.966   2.639   -0.74
 Residual               18.315   4.280
```

Level-1 (within-student)

There is still unexplained within-student residual variance (it is not 0)

Level-2 (between-students):

- There is between-student **residual variance in intercepts** (reading scores at grade 0)
- There is between-student **residual variance in rate of change** (slopes are different)
- Estimated residual correlation between intercepts and change is negative (students who are lower than average at grade 0 tend to grow faster than average, and vice versa)

Covariance = Correlation * SD_0 * SD_1

$$\sigma_{01} = r_{01} \cdot \sigma_0 \cdot \sigma_1$$

$$\begin{aligned}\sigma_{01} &= -0.361 \cdot 19.5 \cdot 2.6 \\ &= -18.6\end{aligned}$$

It is common to provide covariances (rather than correlations) when reporting results from LMER models

Covariances Rather than Correlations

```
Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 740.469  27.212
           grade         6.966   2.639   -0.74
 Residual               18.315   4.280
```

It is common to provide covariances (rather than correlations) when reporting results from LMER models

Covariance = Correlation * SD_0 * SD_1

$$\sigma_{\eta_0,\eta_1} = r_{\eta_0,\eta_1} \times \sigma_{\eta_0} \times \sigma_{\eta_1}$$

$$\begin{aligned}\sigma_{\eta_0,\eta_1} &= -0.74 \times 27.212 \times 2.639 \\ &= -53.14\end{aligned}$$

. . . Or

```
Get variance-covariance matrix of the random effects
> VarCorr(model.b)$studentID

            (Intercept)      grade
(Intercept)   740.46887 -53.403860
grade         -53.40386   6.966203
attr(,"stddev")
(Intercept)       grade
  27.211558    2.639357
attr(,"correlation")
            (Intercept)      grade
(Intercept)   1.0000000 -0.7435688
grade        -0.7435688  1.0000000
```

Centering a Level-1 Predictor for Better Interpretation

Centering the Time Predictor

To make the intercept (and the variance component associated with the intercept) more interpretable, it is a good idea to center the time predictor.

$$\text{Centered Grade}_{ij} = \text{Grade}_{ij} - 5$$

```
Get variance-covariance matrix of the random effects
> mpls$c.grade = mpls$grade - 5
> head(mpls)

  studentID atRisk female minority ell sped  att grade read c.grade
1         1      1      1        1   0    0 0.94     5  172       0
2         1      1      1        1   0    0 0.94     6  185       1
3         1      1      1        1   0    0 0.94     7  179       2
4         1      1      1        1   0    0 0.94     8  194       3
5         2      1      1        1   0    0 0.91     5  200       0
6         2      1      1        1   0    0 0.91     6  210       1
```

```
> model.b2 = lmer(read ~ 1 + c.grade  + (1 + c.grade | studentID), data = mpls,
    REML = FALSE)
> summary(model.b2)

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 380.586  19.509
           c.grade       6.966   2.639   -0.36
 Residual               18.315   4.280
Number of obs: 80, groups:  studentID, 22

Fixed effects:
            Estimate Std. Error t value
(Intercept) 205.7451     4.2322   48.61
c.grade       4.8823     0.7417    6.58

Correlation of Fixed Effects:
        (Intr)
c.grade -0.363
```

The mean reading score at grade 5 is predicted to be 205.7 (no more extrapolation).

Each grade-level difference is associated with a an average 4.9-point difference in reading score (doesn't change since we didn't change the scale).

```
Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 380.586  19.509
           c.grade       6.966   2.639   -0.36
 Residual               18.315   4.280
```

When we center predictors, the RE for the intercept and the correlation between the REs (that involve intercept) will change.

Level-1 (within-student)

There is still unexplained within-student residual variance (it is the same value as before)

Level-2 (between-students):

- The intercept is now grade 5….students' initial status…There is between-student **residual variance in initial status** (reading scores at grade 5)
- There is between-student **residual variance in rate-of-change** (slopes are different)
- Estimated residual correlation between initial status and rate-of-change is negative (students who are lower than average starting out at grade 5 tend to grow faster than average, and vice versa)

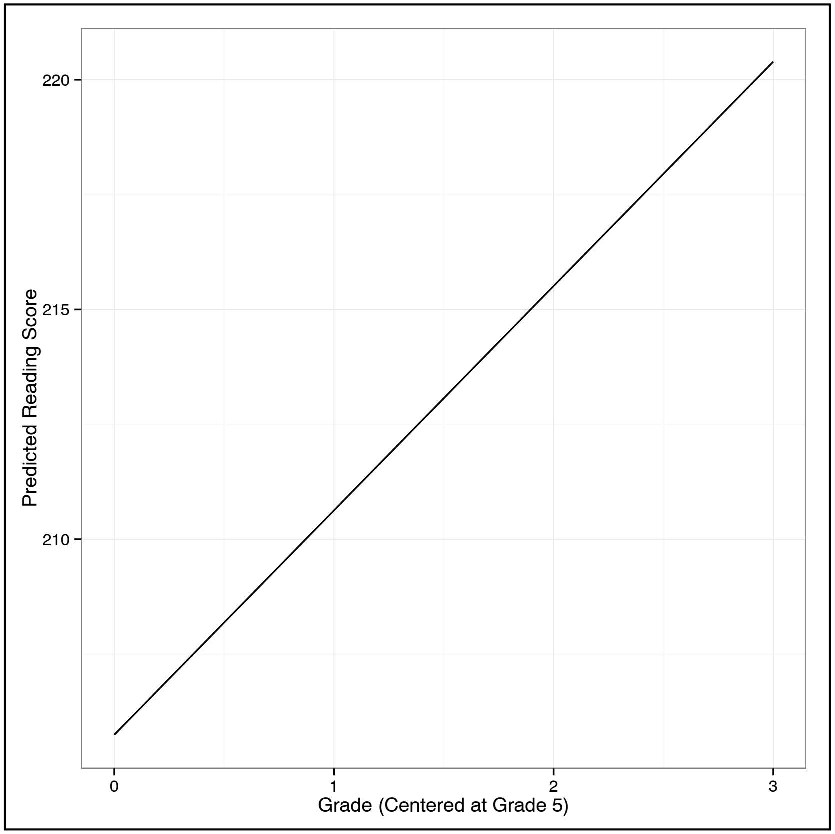
```
# Get variance-covariance matrix of the random effects
> VarCorr(model.b2)$studentID

            (Intercept)    c.grade
(Intercept)   380.58551 -18.572861
c.grade       -18.57286   6.966205
attr(,"stddev")
(Intercept)     c.grade
  19.508601    2.639357
attr(,"correlation")
            (Intercept)   c.grade
(Intercept)    1.000000 -0.360707
c.grade       -0.360707  1.000000
```

Because the correlation between the RE changed, we need to re-compute the covariance estimate.

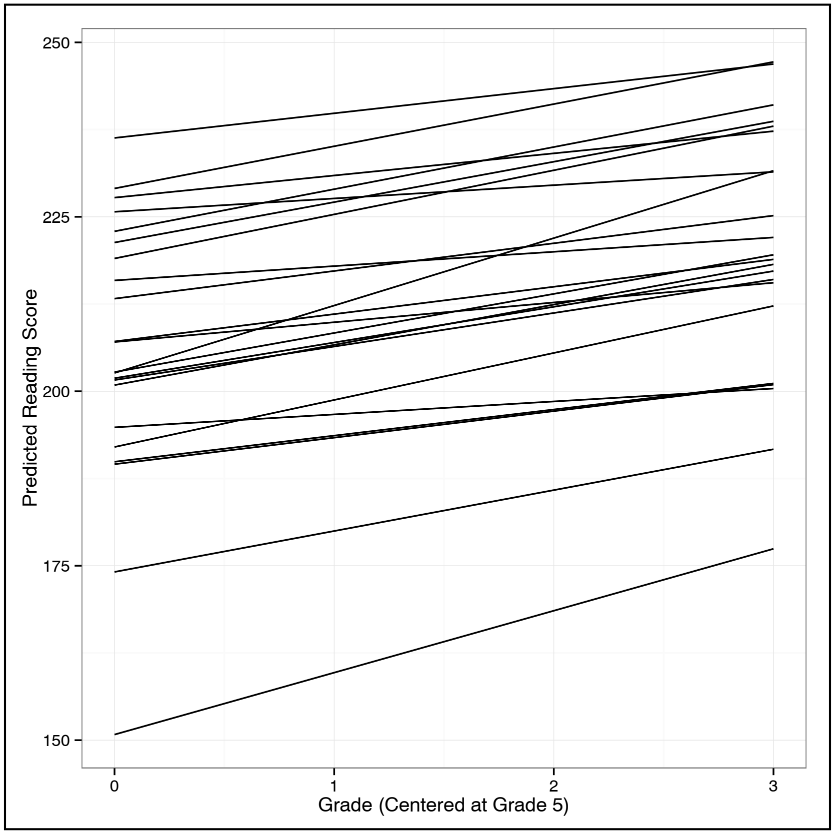
For all future models, we will continue to use the centered grade-level predictor so that we are comparing apples-to-apples.

Plotting Multilevel Models

There are two plots that you might want to consider: (1) a plot of the fixed-effects model; and (2) a plot of each student's model (fixed-effects + random effects).



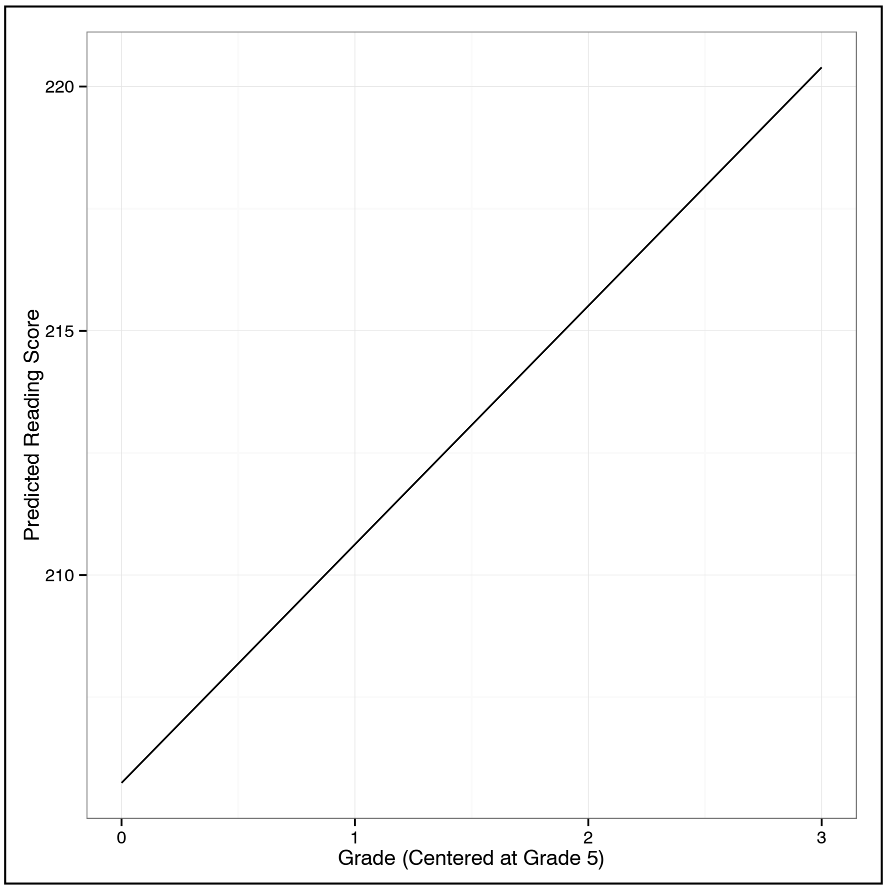
Fixed-effects model



Plot of Each Student's Model

Plot of the Fixed-Effects Model

$$\hat{\text{Read}} = 205.7 + 4.9(\text{Centered Grade})$$



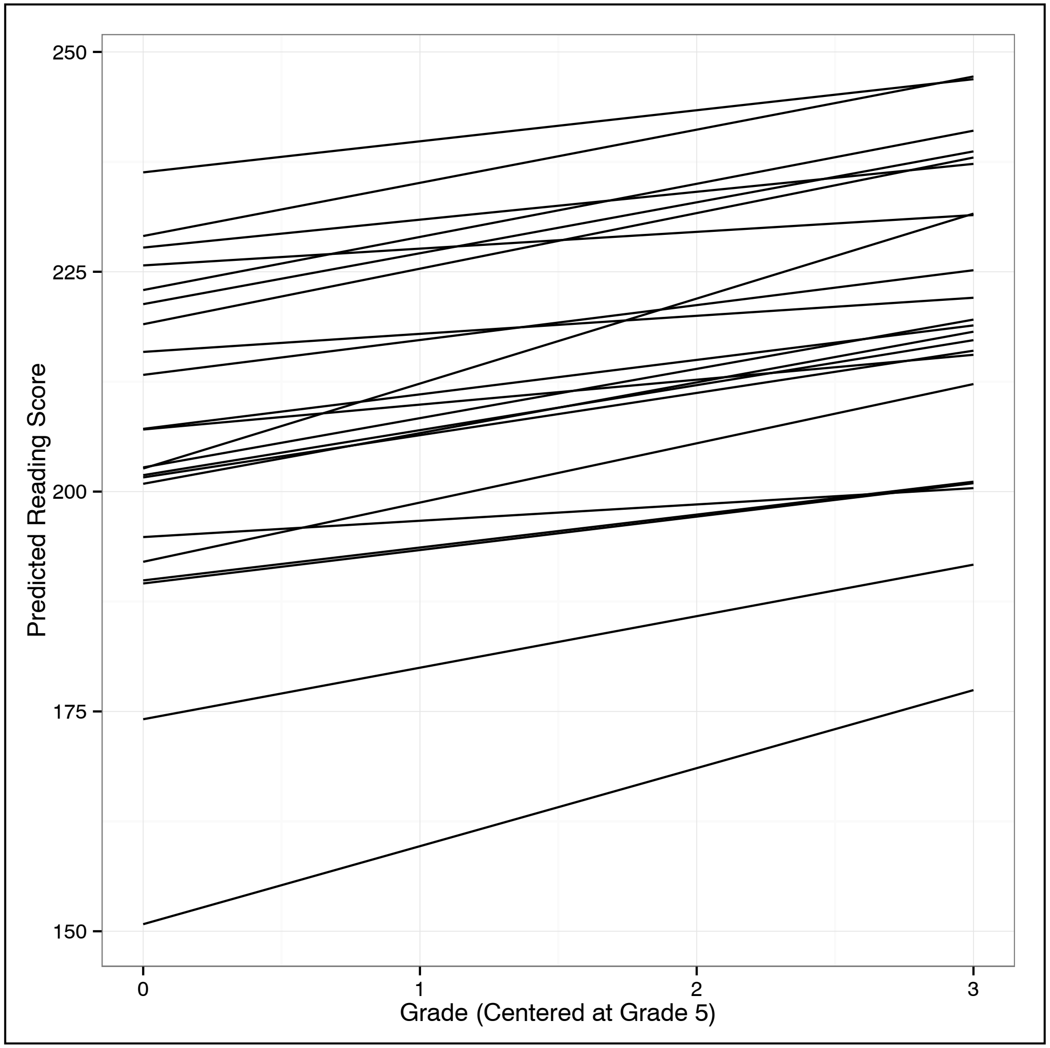
Typically we plot the fixed-effects portion of the model, since that part of the model informs us of the relationship between time and the outcome.

To plot the fixed-effects model, we (1) create a dataset with the appropriate predictors; (2) use the model to estimate the *Y*-hats; and (3) plot.

```
# What associated output do we get with an lmer() object?
> plotData = data.frame(
    c.grade = 0:3
    )

# The predict() function WILL NOT WORK to predict y-hats from the
# fixed-effects model. Use fixef() function instead.
> fixef(model.b2)

(Intercept)     c.grade
 205.745119    4.882322

# Compute y-hat values for the fixed-effects (average) model
> plotData$yhat = fixef(model.b2)["(Intercept)"] +
                  fixef(model.b2)["c.grade"] *  plotData$c.grade

# This is essentially: plotData$yhat = 205.745119 + 4.882322 * plotData$c.grade

> plotData

  c.grade     yhat
1       0 205.7451
2       1 210.6274
3       2 215.5098
4       3 220.3921
```

Now we can plot this using ggplot().

Plot of the Fitted Models for Each Subject



When we set up the data frame we use `expand.grid()` to include the level-2 grouping variable (`studentID`) and the predictors.

Then we can use `predict()` to get the y-hat values.

Another common plot provided includes a different line for each student. To plot this, we (1) create a dataset with the appropriate predictors **and** a student ID; (2) use the model to estimate the Y-hats; and (3) plot.

lmer() objects are a class of objects called S4 objects. They store output from the model in slots. To find out the names of the slots we use the `slotNames()` function.

```
# What associated output do we get with an lmer() object?
> slotNames(model.b2)

 [1] "resp"    "Gp"      "call"    "frame"   "flist"   "cnms"    "lower"   "theta"
 [9] "beta"    "u"       "devcomp" "pp"      "optinfo"
```

```
> model.b2@frame

   read c.grade studentID
1   172       0         1
2   185       1         1
3   179       2         1
4   194       3         1
5   200       0         2
6   210       1         2
```

The frame slot produces a matrix of the outcome, predictor(s), and level-2 ID for each student. This is what we need in Step 1.

```
# Set up the data for plotting
> plotData = model.b2@frame
```

```
# Use predict() to add the y-hat values
> plotData$yhat = predict(model.b2, newdata = plotData)
```

```
# Plot
> ggplot(data = plotData, aes(x = c.grade, y = yhat, group = studentID)) +
    geom_line()
```

The key in the plot is to use group= so that ggplot draws a different line for each student.

Graphically Examining the Level-1 Residuals to Evaluate Predictors

Each Subject's Fitted Model

Consider Student #1's fitted models

Model A

$$\hat{Read} = [212.2 - 28.3]$$

$$\hat{Read} = 183.9$$

This student has a lower average reading score across grades than the average student by 28.3 points.

Model B (Centered)

$$\hat{Read} = [205.7 - 31.6] + [4.9 + 0.98]\,(\text{Centered Grade})$$

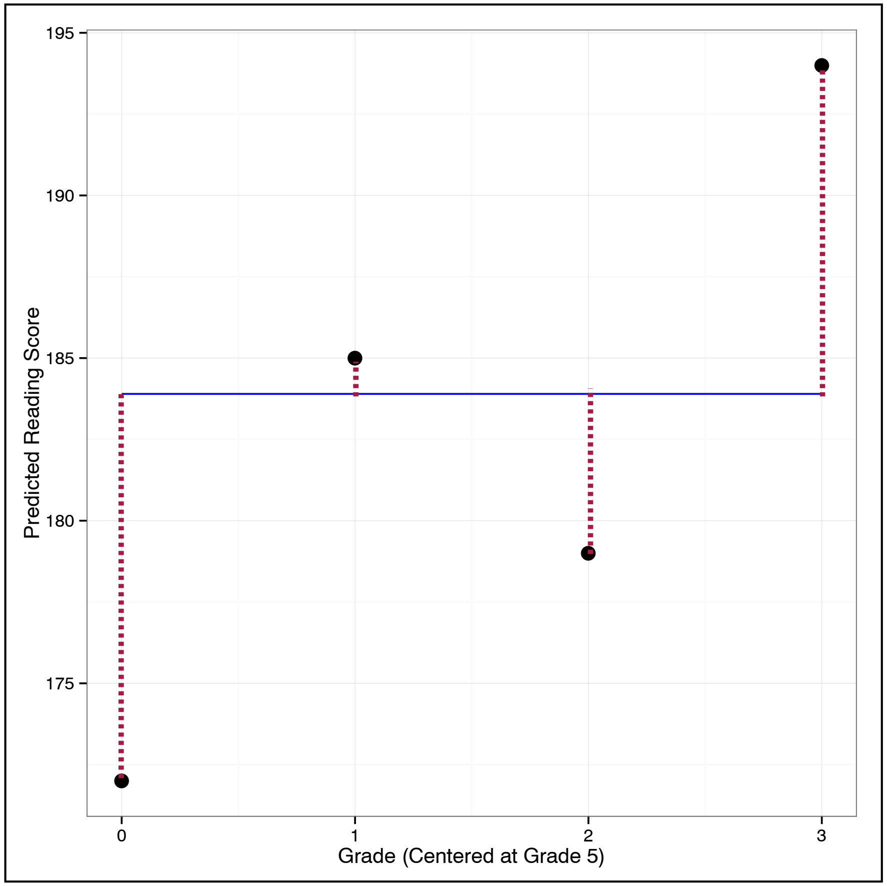
$$\hat{Read} = 174.1 + 5.88(\text{Centered Grade})$$

This student has a lower initial status (grade 5 reading score) than the average student by 31.6 points. S/he also has a higher rate-of-change than the average student by 0.98 points-per-grade.

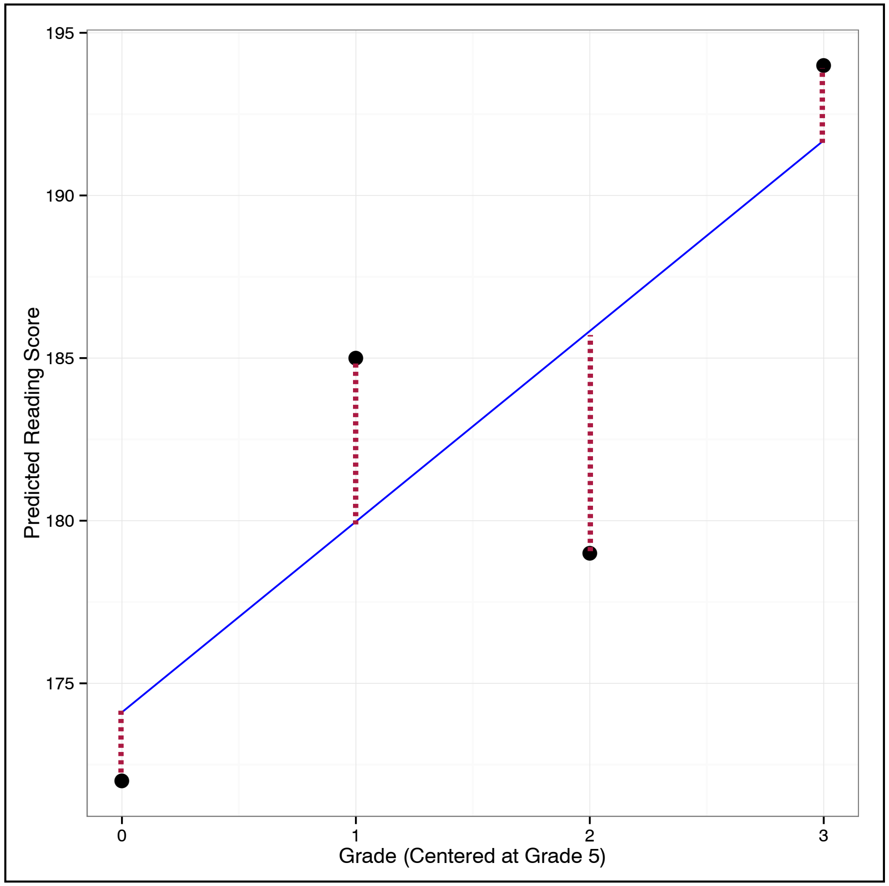
Residuals from Each Subject's Fitted Model

Consider Student #1's level-1 residuals

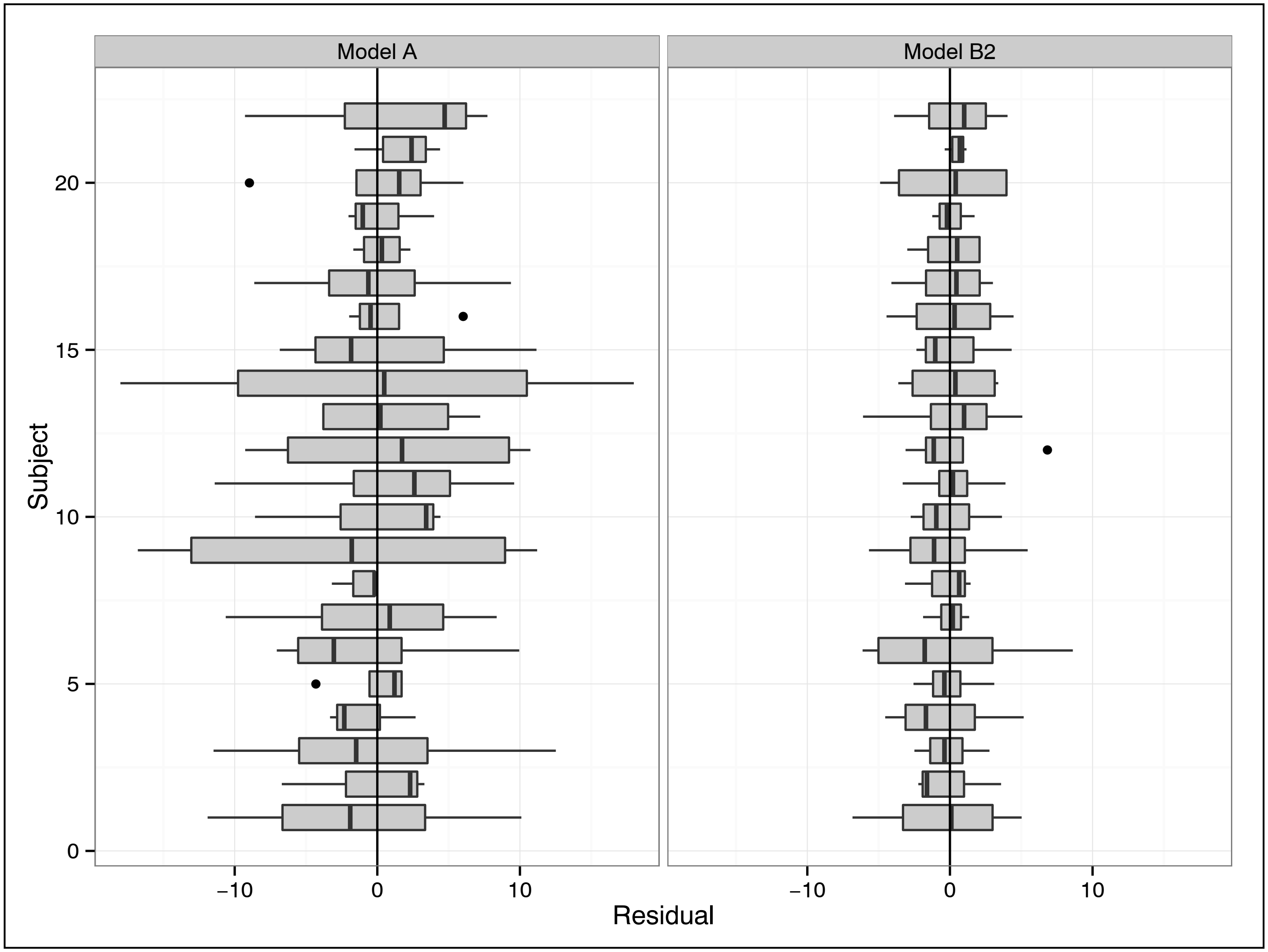
$$\hat{\text{Read}} = 183.9$$



$$\hat{\text{Read}} = 174.1 + 5.88(\text{Centered Grade})$$



Examining All Student's Level-1 Residuals



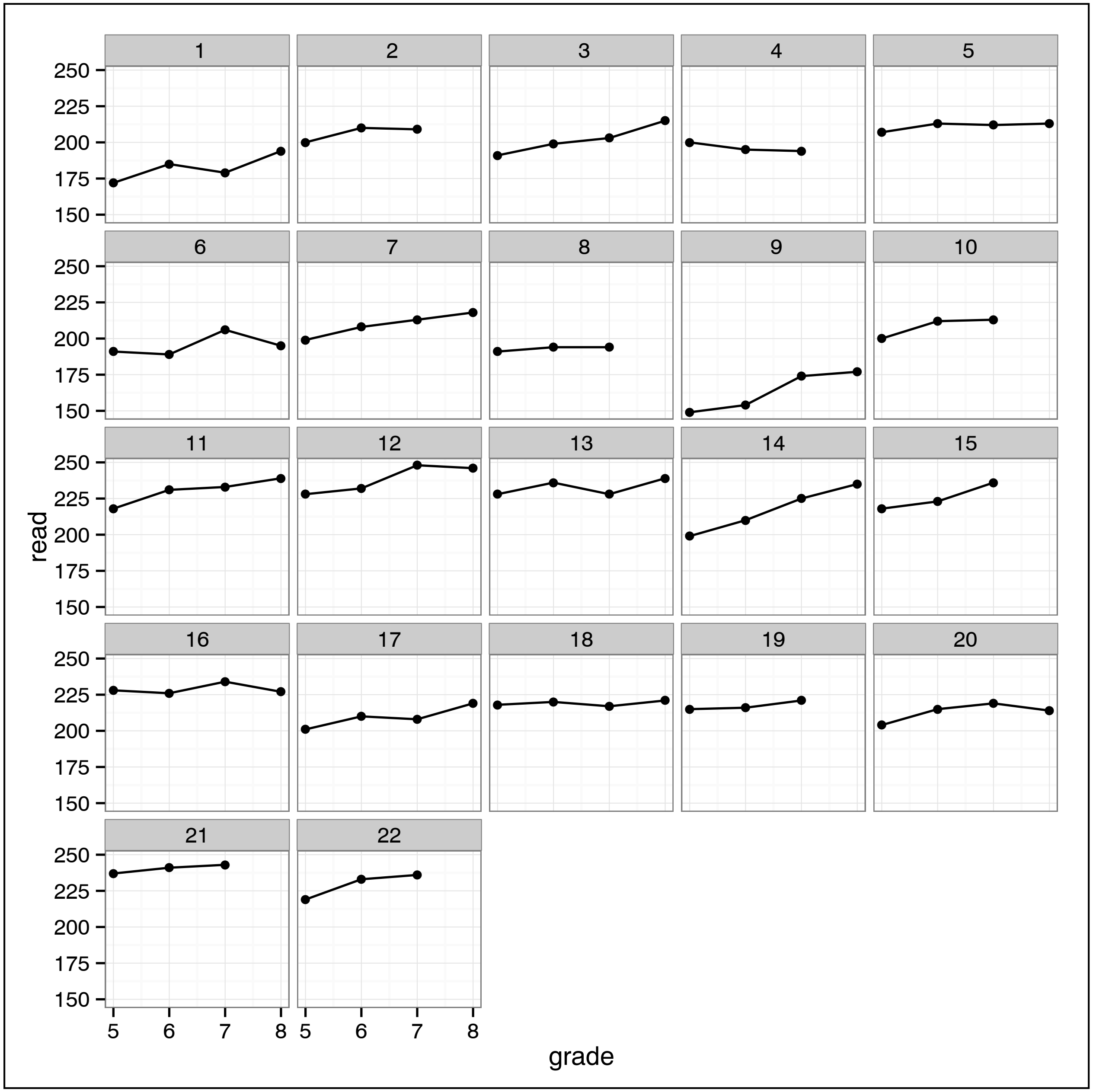
$$\sigma_\epsilon^2(\text{Model A}) = 66.203$$

$$\sigma_\epsilon^2(\text{Model B}) = 18.315$$

Adding the grade-level (time) predictor reduces the level-1 residual variation….it's a keeper!

Quadratic Effect of Time?

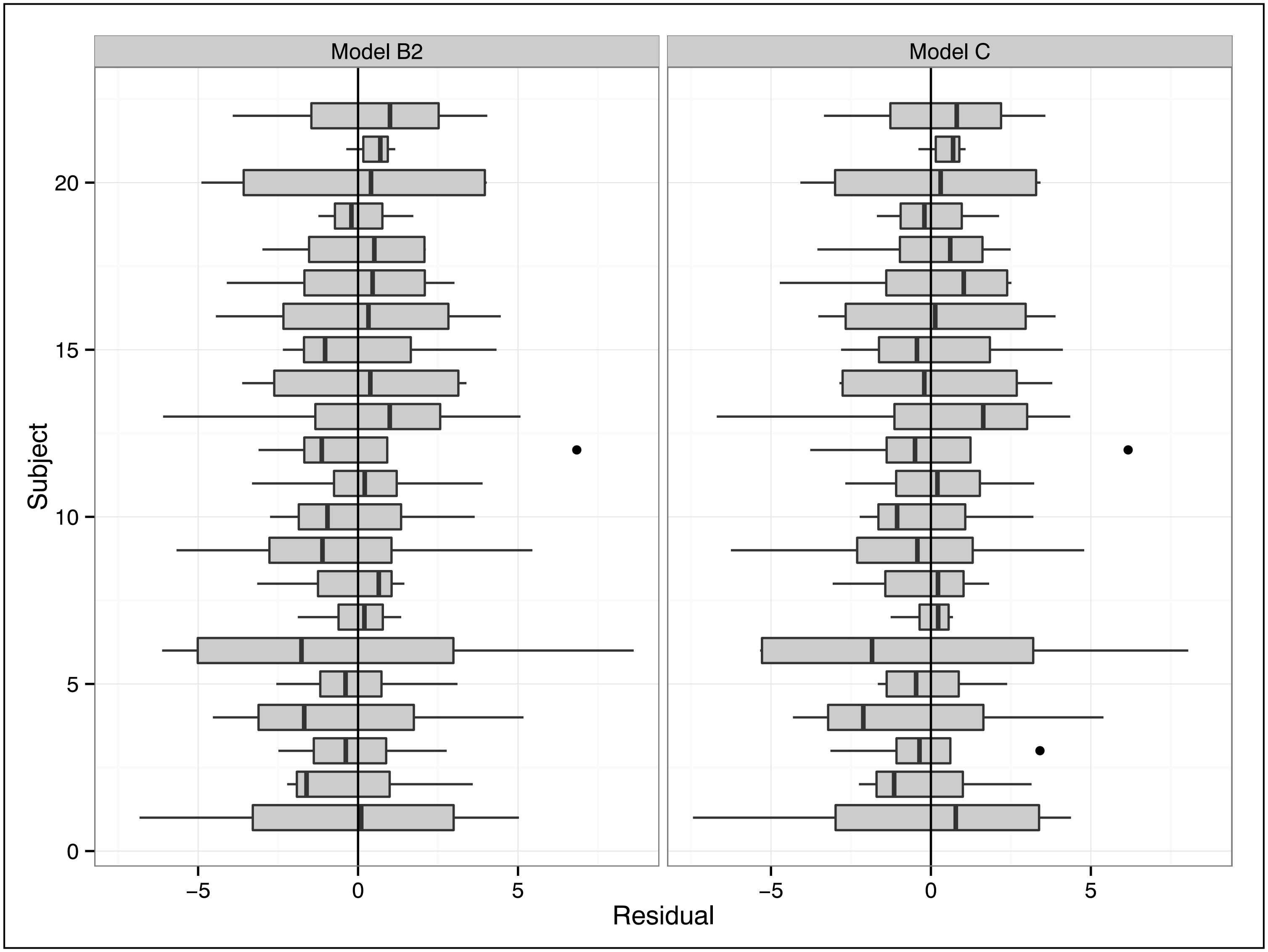
Spaghetti Plot of the Repeated Measures Grouped by ID (Revisited)



Are these quadratic?

```
> model.c = lmer(read ~ 1 + c.grade + I(c.grade ^ 2) +
    (1 + c.grade + I(c.grade ^ 2) | studentID), data = mpls, REML = FALSE)
> summary(model.c)

Random effects:
 Groups    Name         Variance Std.Dev. Corr
 studentID (Intercept)  369.5838 19.2246
           c.grade        6.7526  2.5986   0.01
           I(c.grade^2)   0.1372  0.3703  -1.00 -0.10
 Residual                17.2543  4.1538
Number of obs: 80, groups:  studentID, 22

Fixed effects:
             Estimate Std. Error t value
(Intercept)  205.2102     4.1889   48.99
c.grade        6.6994     1.5538    4.31
I(c.grade^2)  -0.6764     0.5095   -1.33

Correlation of Fixed Effects:
            (Intr) c.grad
c.grade     -0.125
I(c.grad^2) -0.054 -0.879
```

There does not appear to be a quadratic effect of grade on reading scores (*Wald* $t = -1.33$)



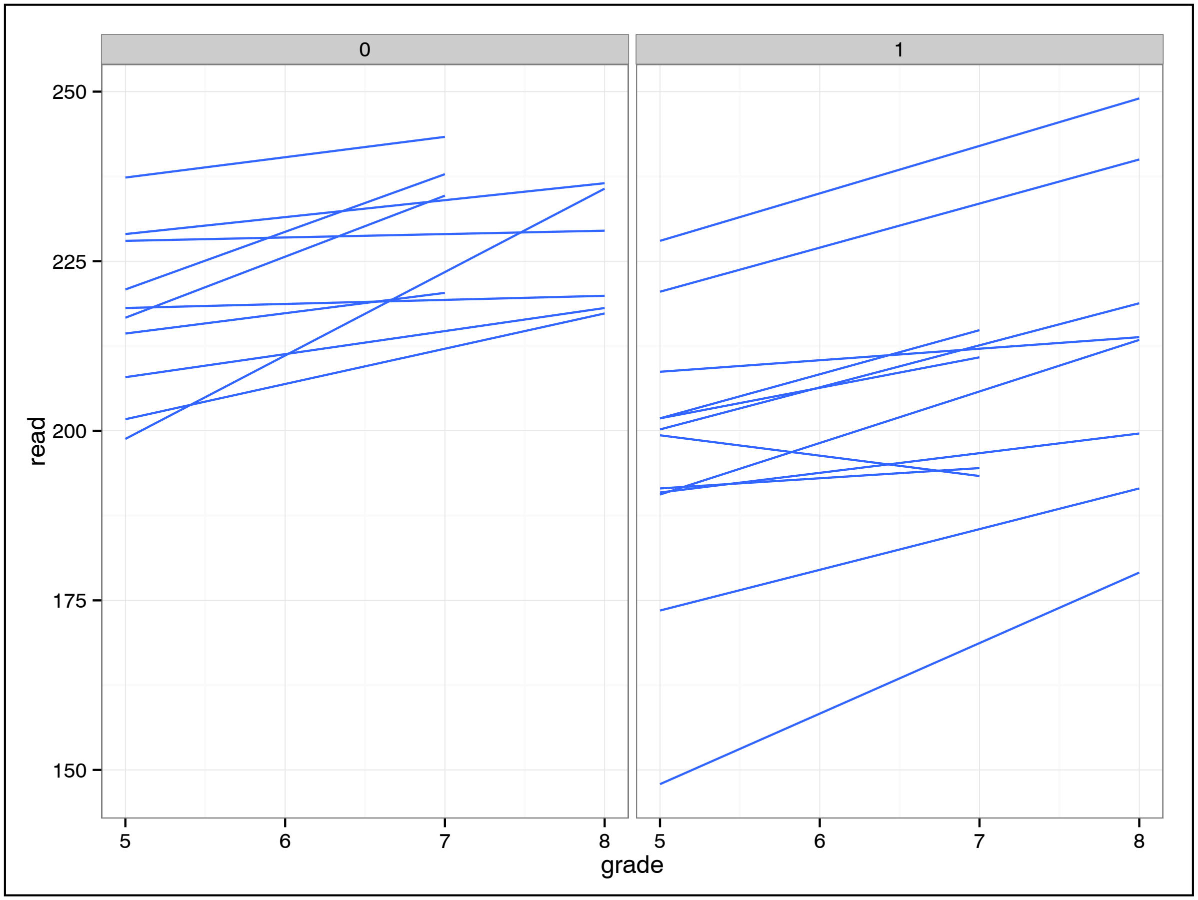
Adding the quadratic term also does not appreciably reduce the size of the level-1 residuals.

Drop the quadratic term.

Adding Level-2 Predictors to Explain Variation in the Intercepts and Slopes

Level-2 Predictors

For categorical predictors, we can look at the fitted models grouped by Student ID and facetted on the predictor.



atRisk

It appears as though there are differences in intercept between at-risk and non-at-risk students.

The slopes look fairly similar between the two groups.

Include atRisk as a Level-2 Predictor

Level-1 Model

$$\hat{\text{Read}}_{ij} = \beta_0^* + \beta_1^*(\text{Centered Grade}) + \epsilon_{ij}$$

Level-2 intercepts

Population average initial status and rate of change for a non-at-risk student

Level-2 slopes

Effect of risk on initial status and rate of change

Level-2 Models

$$\beta_{0i}^* = \beta_{00} + \beta_{01}(\text{risk}_i) + b_{0i} \quad \text{(for initial status)}$$

$$\beta_{1i}^* = \beta_{10} + \beta_{11}(\text{risk}_i) + b_{1i} \quad \text{(for rate of change)}$$

Level-2 residuals

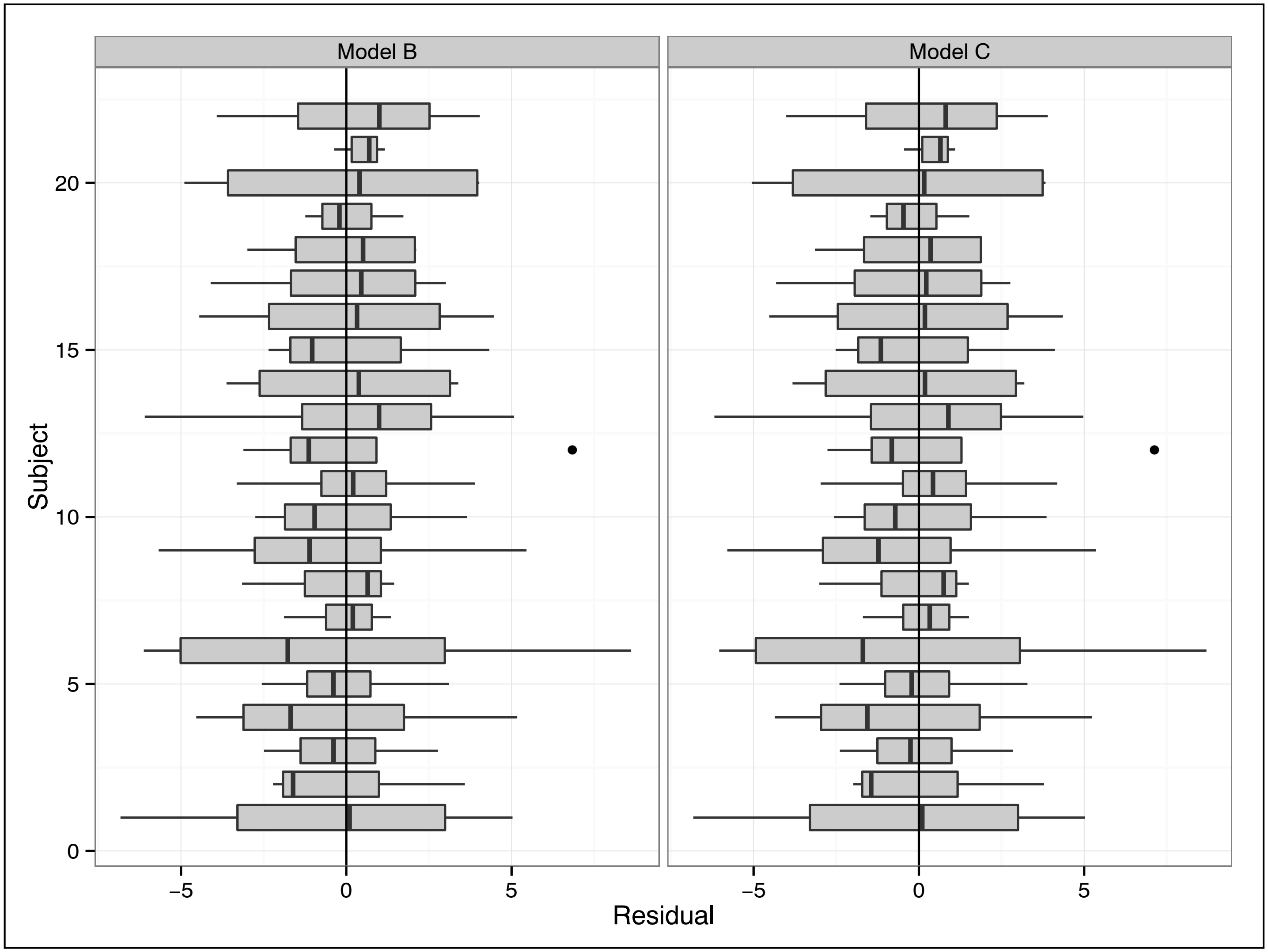
Deviations of individual change trajectories around predicted averages

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

$$\hat{\text{Read}}_{ij} = \beta_{00} + \beta_{01}(\text{Centered Grade}) + \beta_{10}(\text{atRisk}) + \beta_{11}(\text{atRisk})(\text{Centered Grade}) + \eta_0 + \eta_1(\text{Centered Grade}) + \epsilon_{ij}$$

```
> model.d = lmer(read ~ 1 + c.grade + atRisk  + atRisk:c.grade +
     (1 + c.grade | studentID), data = mpls, REML = FALSE)
> summary(model.d)

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 266.711  16.331
           c.grade       6.949   2.636   -0.36
 Residual               18.244   4.271
Number of obs: 80, groups:  studentID, 22

Fixed effects:
               Estimate Std. Error t value
(Intercept)     217.420      5.293   41.07
c.grade           4.570      1.106    4.13
atRisk          -21.403      7.167   -2.99
c.grade:atRisk    0.571      1.489    0.38

Correlation of Fixed Effects:
            (Intr) c.grad atRisk
c.grade     -0.374
atRisk      -0.739  0.276
c.grad:tRsk  0.278 -0.743 -0.374
```

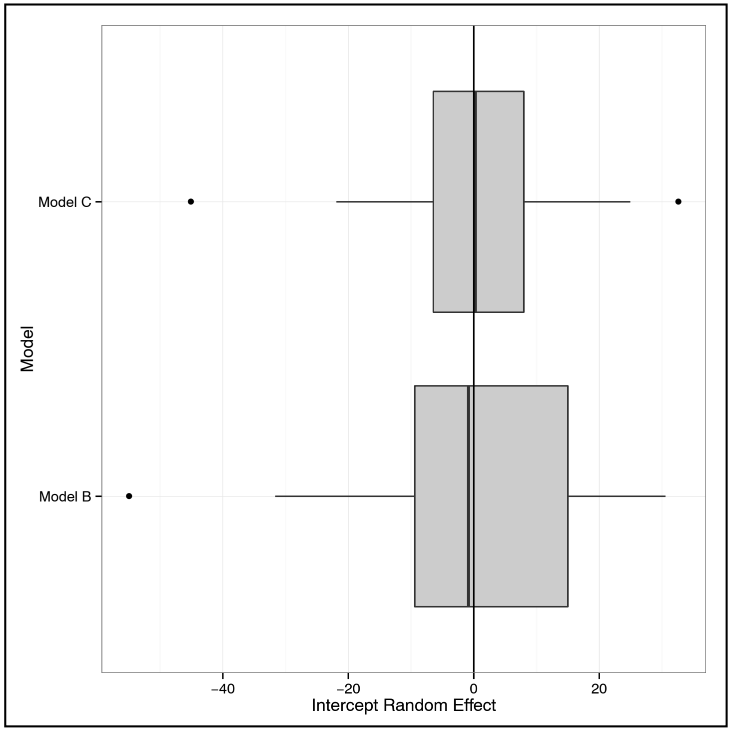
Examining the Change in Variation in the Level-1 Residuals



$$\sigma_{\epsilon}^{2}(\text{Model B}) = 18.315$$

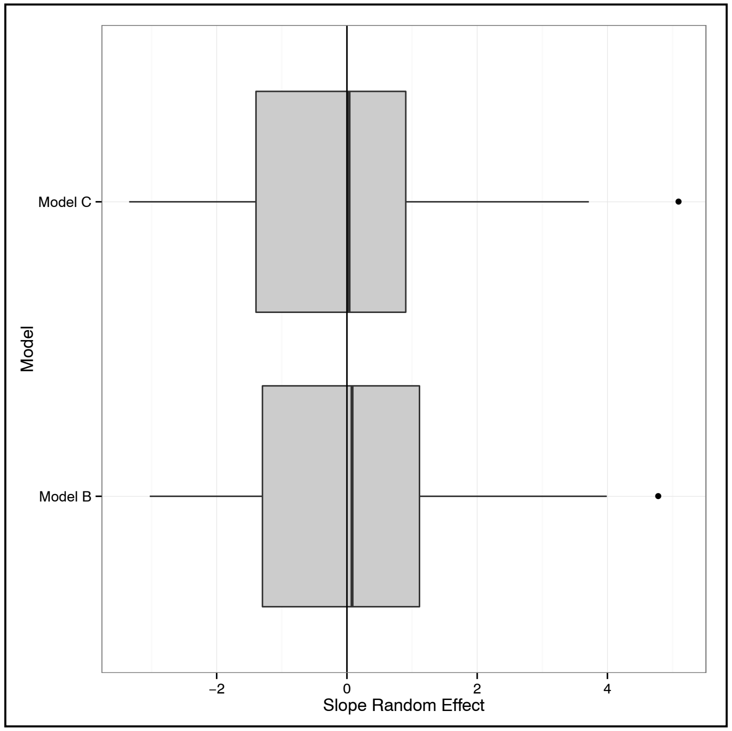
$$\sigma_{\epsilon}^{2}(\text{Model C}) = 18.244$$

Examining the Change in Variation in the Level-2 Residuals



Random effects for intercepts

$$\sigma_0^2(\text{Model B}) = 380.6$$
$$\sigma_0^2(\text{Model C}) = 266.7$$



Random effects for slopes

$$\sigma_1^2(\text{Model B}) = 7.0$$
$$\sigma_1^2(\text{Model C}) = 6.9$$

There interaction effect between grade and atRisk does not seem statistically important (*Wald t* $= 0.38$)

The level-1 **error variance** was **not** appreciably reduced (duh…we added a level-2 predictor) (18.315 to 18.244)

The level-2 variance for **intercepts** was appreciably reduced. (380.586 to 266.711)

The level-2 variance for **slopes** was **not** appreciably reduced. (6.966 to 6.949)

All of this evidence seems to suggest that we **drop the interaction term** from the model. (Note: If atRisk was a focal predictor for answering your RQ, then despite the statistical non-importance, you would keep it in the model at both levels.)

Keeping the main effect of atRisk, but not the interaction, means that we keep atRisk in the level-2 model for intercept but not for slope.

$$\hat{\text{Read}}_{ij} = \beta_0^* + \beta_1^*(\text{Centered Grade}) + \epsilon_{ij}$$

$$\beta_0^* = \beta_{00} + \beta_{10}(\text{atRisk}) + \eta_0$$
$$\beta_1^* = \beta_{01} + \eta_1$$

The model suggests that differences in initial reading scores (at grade 5) are explained by differences in risk, but that differences in the rates-of-change are **not** explained by differences in risk.

$$\hat{\text{Read}}_{ij} = \beta_{00} + \beta_{01}(\text{Centered Grade}) + \beta_{10}(\text{atRisk}) + \eta_0 + \eta_1(\text{Centered Grade}) + \epsilon_{ij}$$

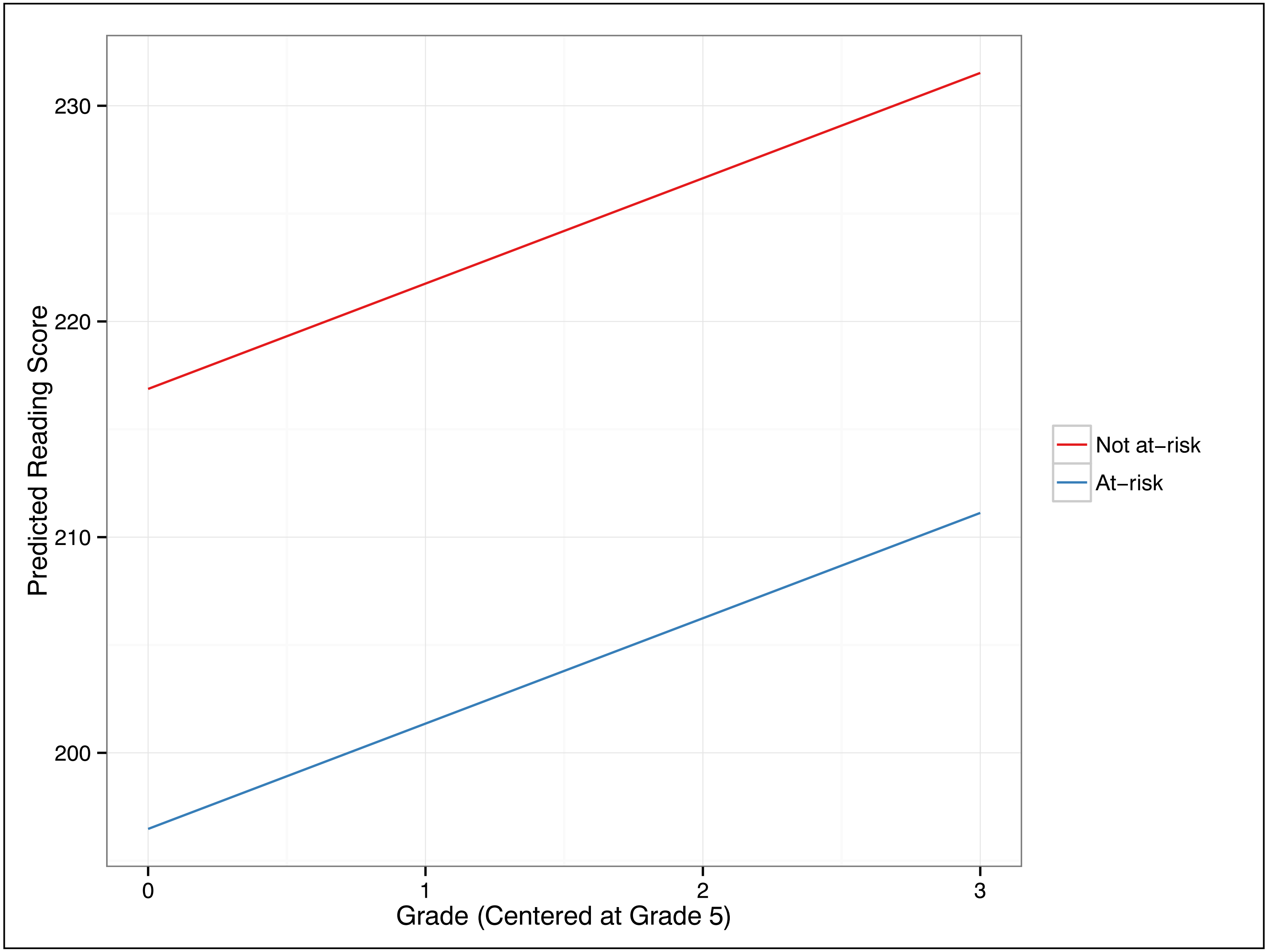
```
> model.e = lmer(read ~ 1 + c.grade + atRisk  + (1 + c.grade | studentID),
     data = mpls, REML = FALSE)
> summary(model.e)

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 studentID (Intercept) 266.912  16.337
           c.grade       7.164   2.677   -0.35
 Residual               18.129   4.258
Number of obs: 80, groups:  studentID, 22

Fixed effects:
            Estimate Std. Error t value
(Intercept) 216.8709     5.0884   42.62
c.grade       4.8836     0.7466    6.54
atRisk      -20.3987     6.6518   -3.07

Correlation of Fixed Effects:
       (Intr) c.grad
c.grade -0.259
atRisk  -0.713 -0.003
```

There main effect of atRisk does seem statistically important (*Wald t* = –3.07)

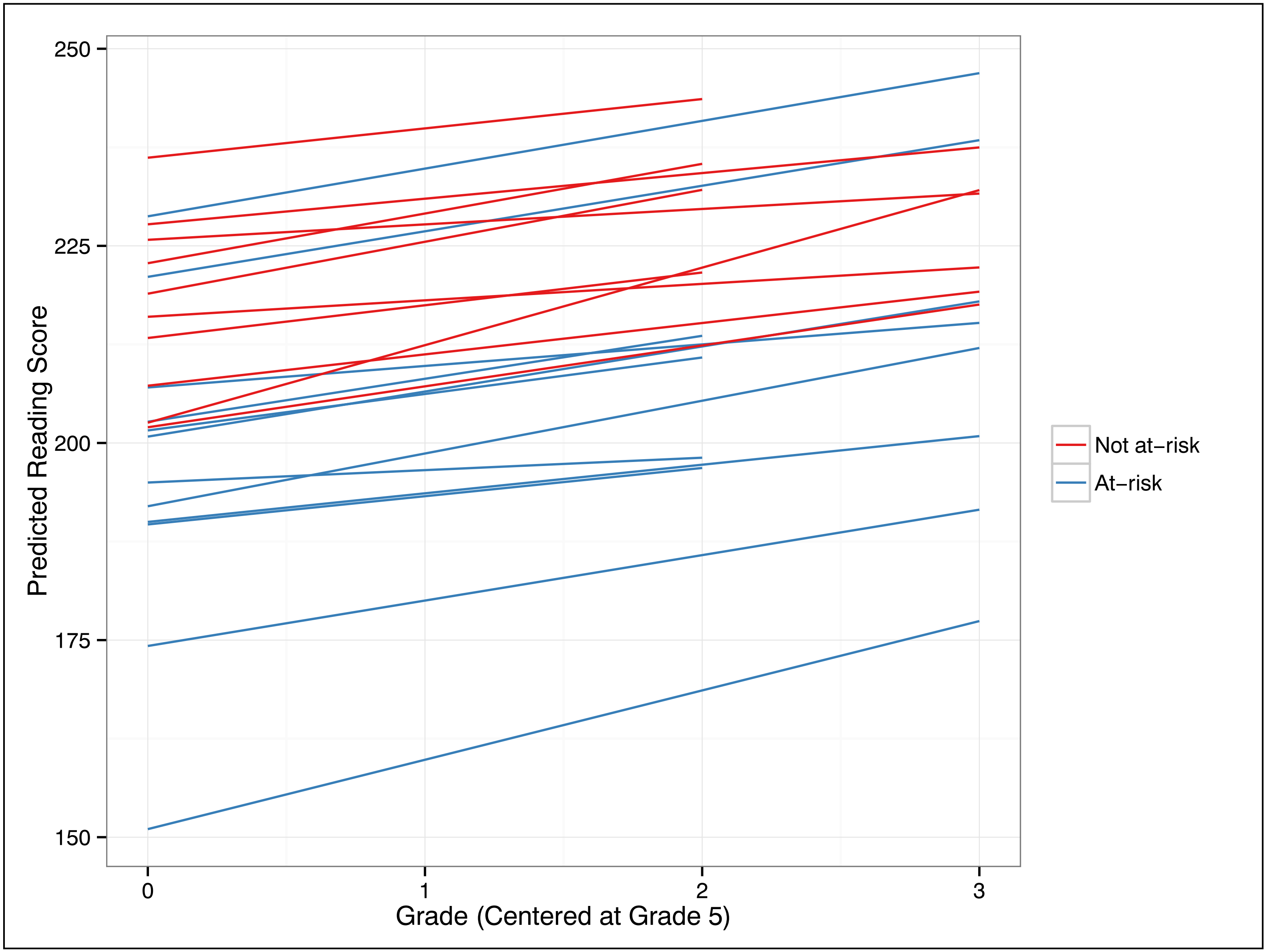
We can also gauge its importance by noting that the level-2 variance for intercepts went down (we explained variation) from the growth model with no predictors (Model B2). 380.586 to 266.912

Keep the atRisk main-effect term.

The plot of the fixed-effects shows this…



The plot of the individual regression lines also show the differences. The differences in slopes that we see are because of the variation between students….nto because of variation in atRisk!



Model Comparisons

Hypothesis Testing

An Approach for Nested Models

Single parameter hypothesis tests

- Simple to conduct and easy to interpret—making them very useful in hands on data analysis
- However, statisticians disagree about their nature, form, and effectiveness
- Disagreement is do strong that some software packages (e.g., R, MLwiN) won't output them
- Their behavior is poorest for tests on variance components

Deviance based hypothesis tests

- Based on the log likelihood (LL) statistic that is maximized under Maximum Likelihood estimation
- Have superior statistical properties (compared to the single parameter tests)
- Special advantage: permit joint tests on several parameters simultaneously
- You need to do the tests "manually" because automatic tests are rarely what you want

$$\text{Deviance} = -2\left[LL_{\text{Current Model}} - LL_{\text{Saturated Model}}\right]$$

Quantifies how much worse the reduced model is in comparison to a full model
A model with a small deviance statistic is nearly as good; a model with large deviance statistic is much worse (we obviously prefer models with smaller deviance)

Simplification: Because a saturated model fits perfectly, its $LL = 0$ and the second term drops out, making Deviance $= -2(LL_{current})$

```
> logLik(model.a)

'log Lik.' -313.2524 (df=3)
```

The log-likelihood for Model A is –313.25.

$$\text{Deviance} = -2\left[LL_{\text{Current Model}} - LL_{\text{Saturated Model}}\right]$$

$$= -2 \times (-313.25 - 0)$$
$$= 626.5$$

```
> logLik(model.b2)

'log Lik.' -285.8487 (df=6)
```

The log-likelihood for Model A is –285.8487.

$$\text{Deviance} = -2\left[LL_{\text{Current Model}} - LL_{\text{Saturated Model}}\right]$$

$$= -2 \times (-285.8487 - 0)$$
$$= 571.7$$

If we had fitted a model with perfect fit, the deviance value would be 0.

$$\begin{aligned}\text{Deviance} &= -2\left[LL_{\text{Current Model}} - LL_{\text{Saturated Model}}\right] \\ &= -2 \times (0 - 0) \\ &= 0\end{aligned}$$

The deviance for Model A is 626.6.
The deviance for Model B2 is 571.7.

This suggests that Model B2 has better fit (the deviance value is closer to 0) than the deviance from Model A.

How much better does Model B2 fit than Model A?
626.6 – 571.7 = 54.8

Note that if Model B2 fit **as well as** Model A, then the difference in their deviances would be 0.

Hypothesis Testing using Deviance Statistics

		Parameter	Model A	Model B
Fixed effects				
Initial status	Intercept	β_{00}	212.2 (3.9)	205.8 (4.2)
Rate of change	Intercept	β_{10}	✓	4.9 (0.7)
Variance components				
Level-1	Within-students	σ_ϵ^2	66.2	18.3
Level-2	In initial status	σ_0^2	319.3	380.6
	In rate of change	σ_1^2	✓	7
	Covariance	σ_{01}	✓	–18.6
Pseudo R^2 statistics and Goodness-of-fit				
	Deviance		626.5	571.7
	AIC		632.5	583.7
	BIC		639.7	598

You can use deviance statistics to compare two models if two criteria are satisfied:

1. Both **models are fit to the same exact data**—beware missing data
2. **One model is nested within the other**—we can specify the less complex model (e.g., Model A) by imposing constraints on one or more parameters in the more complex model (e.g., Model B), usually, but not always, setting them to 0)

If these conditions hold, then:

- Difference in the two deviance statistics is asymptotically distributed as $\chi 2$
- df = # of independent constraints

We can obtain Model A from Model B by invoking three constraints:

✓ $\beta_{10} = 0$
✓ $\sigma^2_0 = 0$
✓ $\sigma_{01} = 0$

H_0: $\beta_{10} = 0; \sigma^2_0 = 0; \sigma_{01} = 0$

Compute **difference in Deviance statistics** and compare to appropriate $\chi 2$ distribution

$\Delta Deviance = 54.81$, $df = 3$ ($p < .001$)
⇛ reject H_0

So…what has been the **effect** of moving from an unconditional model to an growth model?

Hypothesis Testing using Deviance Statistics

		Parameter	Model B	Model C
Fixed effects				
Initial status	Intercept	β_{00}	205.8 (4.2)	217.4 (5.3)
	Risk	β_{01}	✓	–21.4 (7.1)
Rate of change	Intercept	β_{10}	4.9 (0.7)	4.6 (1.1)
	Risk	β_{11}	✓	0.6 (1.5)
Variance components				
Level-1	Within-persons	σ_ϵ^2	18.3	18.2
Level-2	In initial status	σ_0^2	380.6	266.7
	In rate of change	σ_1^2	7	6.9
	Covariance	σ_{01}	–18.6	–15.3
Pseudo R^2 statistics and Goodness-of-fit				
	Deviance		571.7	563.8
	AIC		583.7	579.8
	BIC		598	598.8

Key idea: Deviance statistics are great for simultaneously evaluating the effects of adding predictors to both level-2 models:

We can obtain Model B from Model C by invoking 2 constraints:

- ✓ $\beta_{01} = 0$
- ✓ $\beta_{11} = 0$

H_0: $\beta_{01} = 0; \beta_{11} = 0$

Compute **difference in Deviance statistics** and compare to appropriate $\chi 2$ distribution

$\Delta Deviance = 7.93$, $df = 2$ ($p = 0.019$)
$\Rightarrow$ reject H_0

The pooled test **does not** imply that each level-2 slope is on its own statistically significant

```
> anova(model.a, model.b2, model.d, model.e)

Data: mpls
Models:
model.a: read ~ 1 + (1 | studentID)
model.b2: read ~ 1 + c.grade + (1 + c.grade | studentID)
model.e: read ~ 1 + c.grade + atRisk + (1 + c.grade | studentID)
model.d: read ~ 1 + c.grade + atRisk + atRisk:c.grade + (1 + c.grade |
model.d:     studentID)
         Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
model.a   3 632.50 639.65 -313.25   626.50
model.b2  6 583.70 597.99 -285.85   571.70 54.8073      3  7.548e-12 ***
model.e   7 577.91 594.58 -281.95   563.91  7.7916      1   0.005249 **
model.d   8 579.76 598.82 -281.88   563.76  0.1430      1   0.705276
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Each model is compared to model on the line that immediately precedes it.

- model.b2 is compared to model.a
- model.e is compared to model.b2
- model.d is compared to model.e

Note: If you have fitted a model using REML the anova() function will re-fit the models using ML before outputting the results.

When the models are non-nested, then you have to use information criteria. Hoever, they can also be used (and are better) when the models are nested.

```
> library(MuMIn)

> AICc(model.a)
[1] 632.8206

> AICc(model.b2)
[1] 584.8482

> AICc(model.e)
[1] 579.4614

> AICc(model.d)
[1] 581.791
```

Here we would adopt Model E (atRisk main-effect but no interaction with grade).

A Final Comment about Estimation and Hypothesis Testing

Two common methods of estimation

Maximum likelihood (ML):

Seeks those parameter estimates that maximize the likelihood function, which assesses the joint probability of simultaneously observing all the sample data actually obtained (implemented, e.g., in *HLM* and *SAS Proc Mixed* and *R*).

Generalized Least Squares (GLS) (& Iterative GLS):

Iteratively seeks those parameter estimates that minimize the sum of squared residuals (allowing them to be autocorrelated and heteroscedastic) (implemented, e.g., in *MLwiN*).

A more important distinction: Full vs. Restricted (ML or GLS)

Full: Simultaneously estimate the fixed effects and the variance components.

(Default in *MLwiN* & *HLM* & *R*)

Goodness of fit statistics apply to the entire model (both fixed and random effects)

This is the method we have used in the examples shown so far

Restricted: Sequentially estimate the fixed effects *and then* the variance components

(Default in *SAS Proc Mixed*)

Goodness of fit statistics apply to only the random effects

So we can only test hypotheses about VCs (and the models being compared must have identical fixed effects)

Proportion of Variance Accounted For

Pseudo-R^2

Proportion of Variation Explained

		Parameter	Model A	Model B
Variance components				
Level-1	Within-students	σ_ϵ^2	66.2	18.32
Level-2	In initial status	σ_0^2	319.34	380.59
	In rate of change	σ_1^2		6.97
	Covariance	σ_{01}		–18.59
Pseudo R^2 statistics and Goodness-of-fit				
		$R^2_{Y,\hat{Y}}$		$\boxed{0.973}$
		R^2_e		0.723

Proportion of total variance explained by the model

$$R^2_{Y,\hat{Y}} = 0.987^2$$
$$= 0.973$$

97.3% of the total variation in reading scores is explained by the model

```
## Compute the correlation between the fitted Y-values and the observed Y-values
## Square this correlation
> cor(fitted(lmer.1), mpls$read) ^ 2

[1] 0.9737471
```

Proportion of Variation Explained: Level-1

		Parameter	Model A	Model B
Variance components				
Level-1	Within-students	σ_ϵ^2	66.2	18.32
Level-2	In initial status	σ_0^2	319.34	380.59
	In rate of change	σ_1^2		6.97
	Covariance	σ_{01}		–18.59
Pseudo R^2 statistics and Goodness-of-fit				
		$R^2_{Y,\hat{Y}}$		0.973
		R^2_ϵ		0.723

Proportional reduction in the level-1 variance component

$$R^2_\epsilon = \frac{66.20 - 18.32}{66.20}$$
$$= 0.723$$

72.3% of the within-student variation in reading scores is associated with linear time

In general: $$R^2_\epsilon = \frac{\sigma^2_\epsilon(\text{Unconditional Model}) - \sigma^2_\epsilon(\text{New Model})}{\sigma^2_\epsilon(\text{Unconditional Model})}$$

Pseudo R^2 for the Level-2 Variance Components

We can extend this same idea of proportional reduction in variance to Level-2 (to estimate the percentage of between-student variation in reading scores associated with predictors)

$$R_0^2 = \frac{\sigma_0^2(\text{Growth Model}) - \sigma_0^2(\text{New Model})}{\sigma_0^2(\text{Growth Model})}$$

In general:

$$R_1^2 = \frac{\sigma_1^2(\text{Growth Model}) - \sigma_1^2(\text{New Model})}{\sigma_1^2(\text{Growth Model})}$$

Where We have Been and Where We are Going…

What these unconditional models tell us:

1. Almost all of the total variation in reading scores is attributable to differences among students
2. About 72% of the within-student variation in reading scores is explained by linear time
3. There is significant variation in both initial status and rate of change— so it pays to explore substantive predictors (risk and ethnicity)

How do we build statistical models?

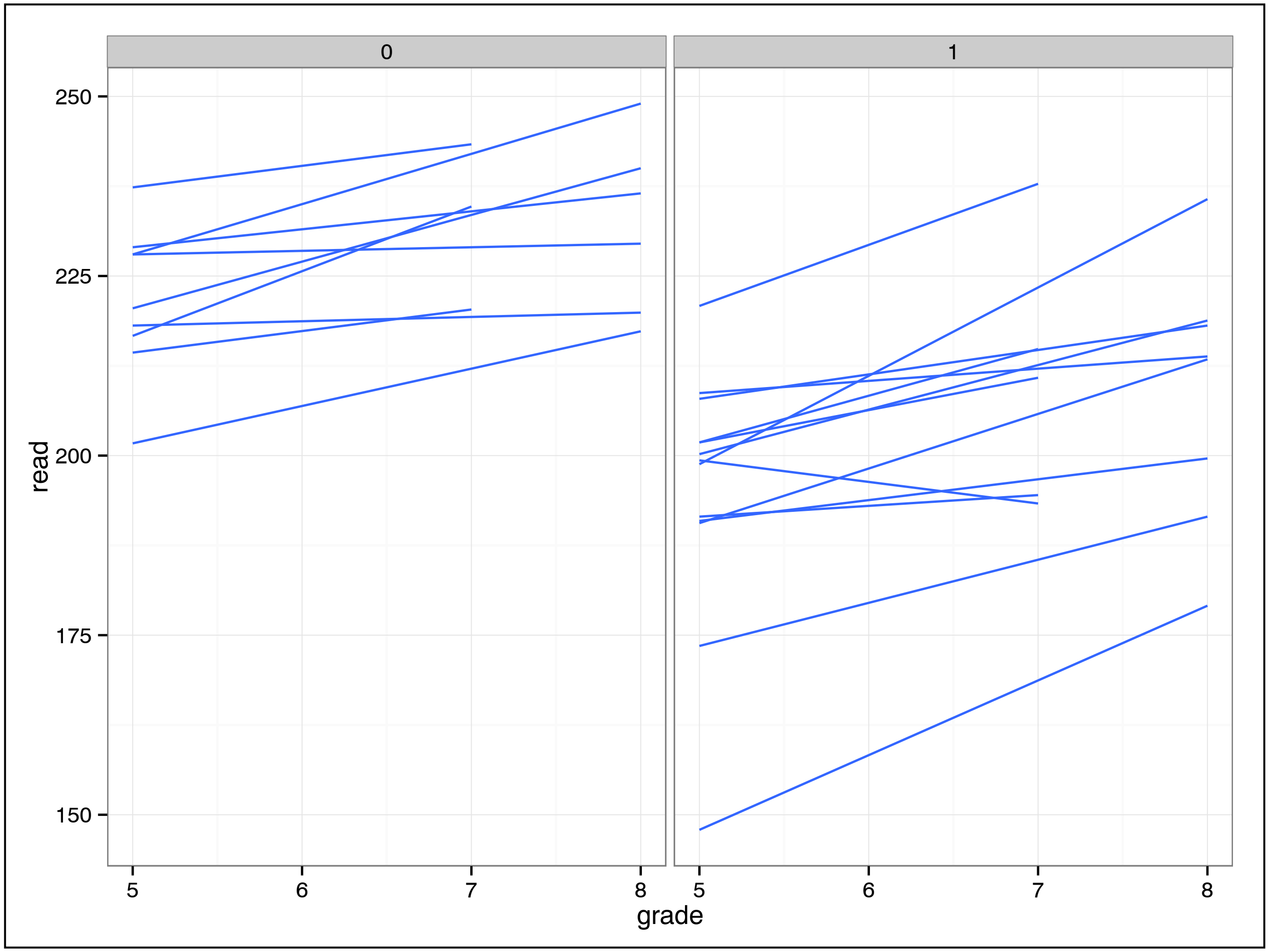
- <u>Use all your intuition and skill you bring from the cross sectional world</u>
  - ✓ Examine the effect of each predictor separately
  - ✓ Prioritize the predictors,
    - ➡ Focus on your focal predictors
    - ➡ Include interesting and important control predictors
- <u>Progress towards a "final model"</u> whose interpretation addresses your research questions

But because the data are longitudinal, we have some other options…

- <u>Multiple level-2 outcomes</u> (the individual growth parameters)—each can be related separately to predictors
- <u>Two kinds of effects being modeled:</u>
  - ✓ Fixed effects
  - ✓ Variance components
  - ✓ *Not all effects are required in every model*

		Parameter	Model B	Model C
Fixed effects				
Initial status	Intercept	β_{00}	205.8 (4.2)	217.4 (5.3)
	Risk	β_{01}		–21.4 (7.1)
Rate of change	Intercept	β_{10}	4.9 (0.7)	4.6 (1.1)
	Risk	β_{11}		0.6 (1.5)
Variance components				
Level-1	Within-persons	σ_ϵ^2	18.3	18.2
Level-2	In initial status	σ_0^2	380.6	266.7
	In rate of change	σ_1^2	7	6.9
	Covariance	σ_{01}	–18.6	–15.3
Pseudo R^2 statistics and Goodness-of-fit				
	$R^2_{Y,\hat{Y}}$		0.973	0.974
	R^2_e		0.723	0.725
	R^2_0			0.299
	R^2_1			0.002
	Deviance		571.7	563.8
	AIC		583.7	579.8
	BIC		598	598.8

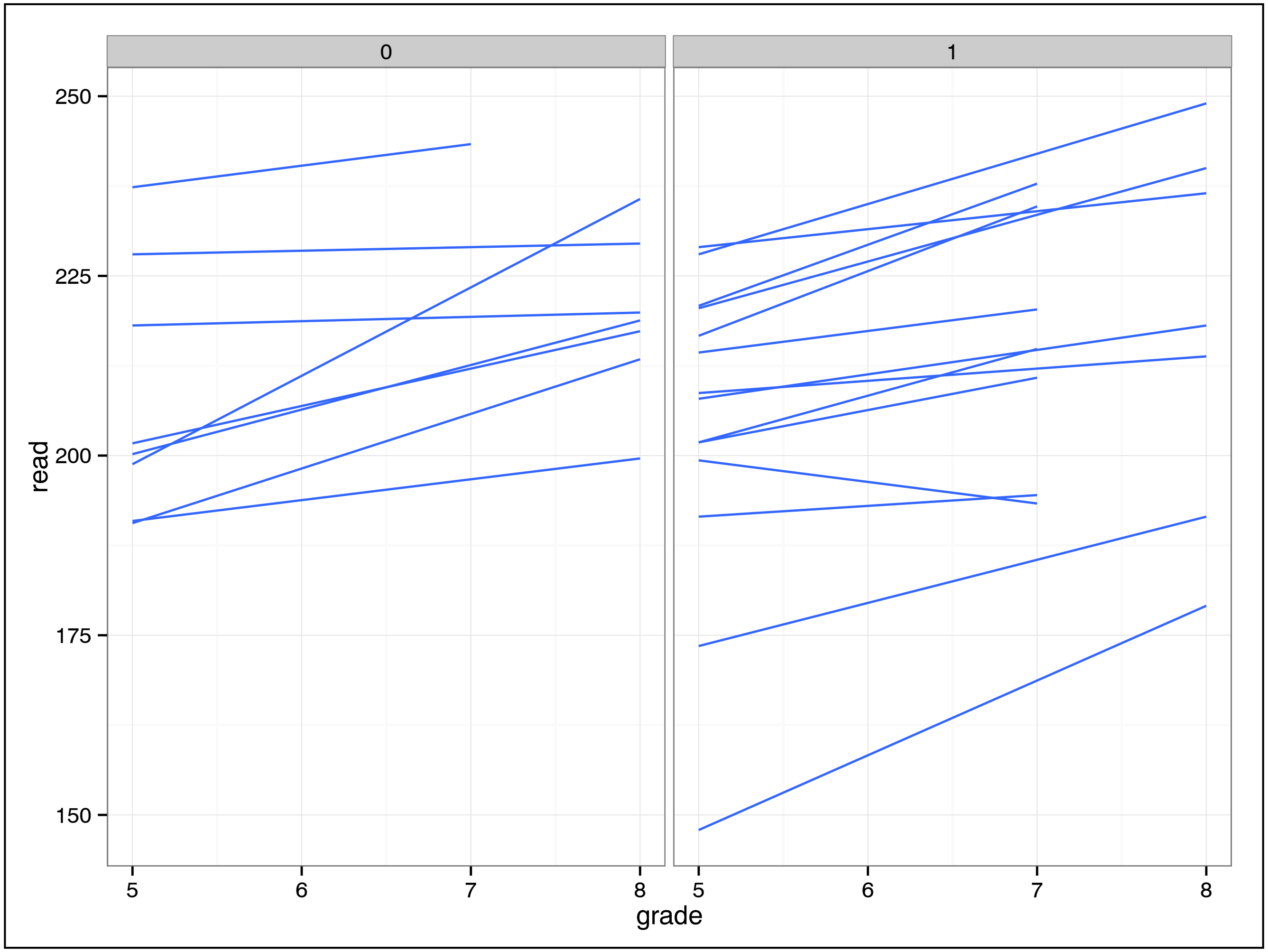
What Next?



minority

It appears as though there are differences in intercept between minority and non-minority students.

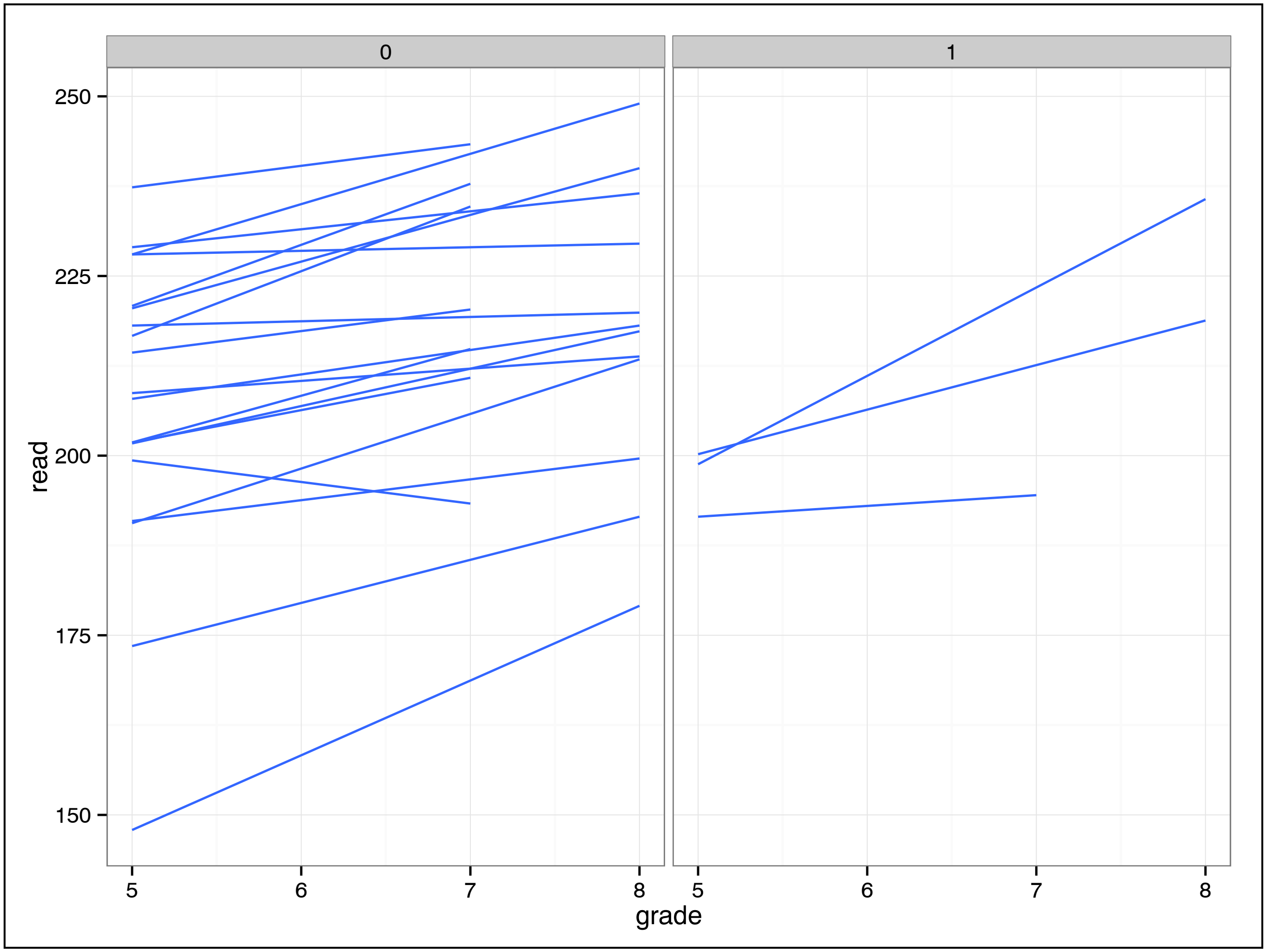
The slopes look fairly similar between the two groups.



female

It appears as though there are differences in intercept between female and male students.

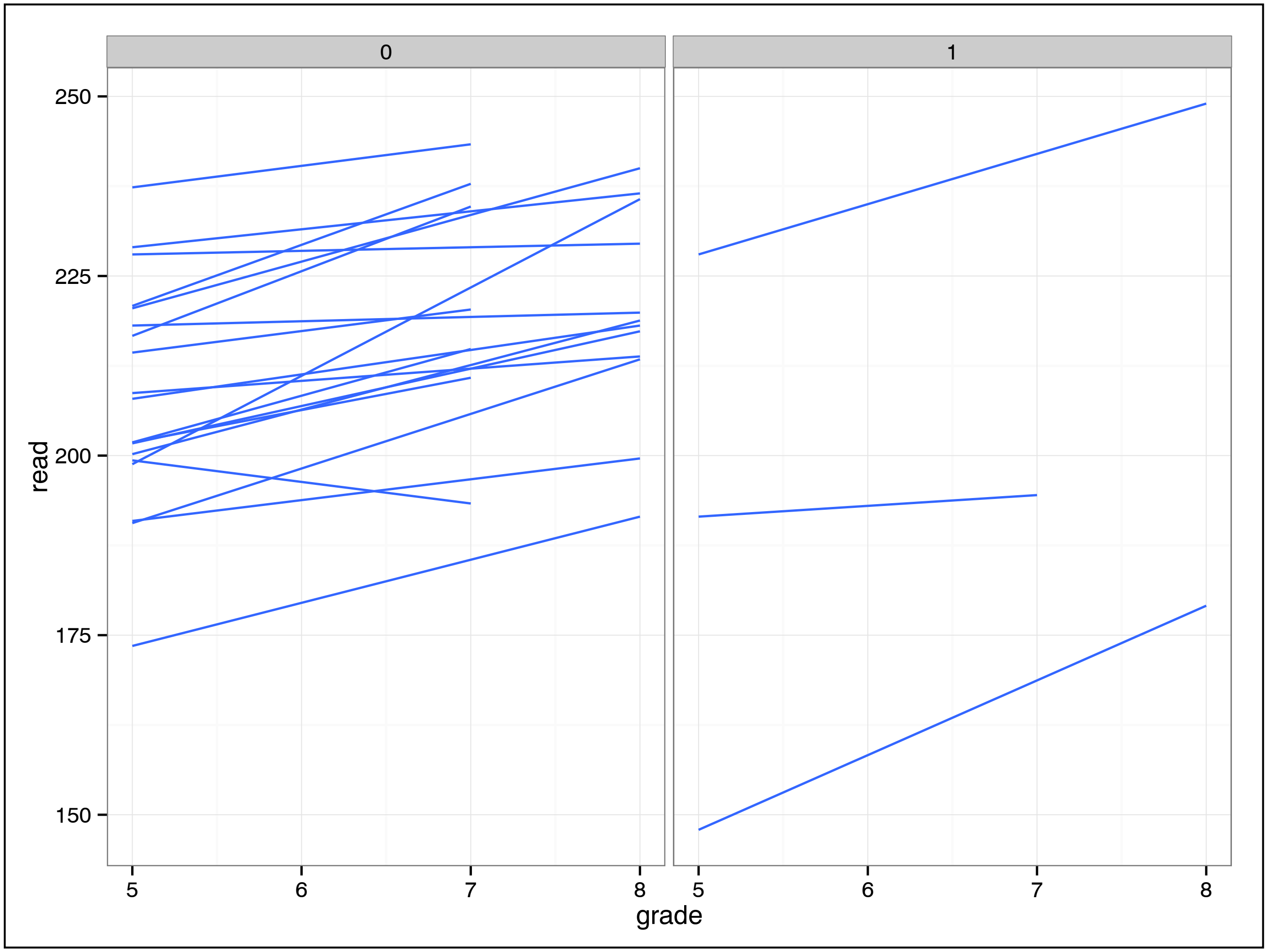
The slopes look fairly similar between the two groups, although maybe they are slightly flatter for males..



ell

It appears as though there **no** differences in intercept between English Language Learners (ELL) and non-ELL students.

The slopes look fairly similar between the two groups.



sped

It appears as though there **slight** differences in intercept between special education and and non-special education students.

The slopes look fairly similar between the two groups although maybe flatter (?) for the special education students.

Examine Assumptions

Check the following assumptions for your adopted model:

<u>Density Plots (Normality)</u>

- Level-1 residuals
- Level-2 residuals for the intercepts
- Level-2 residuals for the slopes

<u>Residual Plots (Linearity, homogeneity of variance, etc.)</u>

- Level-1 residuals vs. fitted values
-

Presentation of Model Results

Some artifacts that you may want to present in a paper

Plots of the Adopted Model

- Fixed-effects model
- Individual regression lines

Table

- Table of the regression results for selected models
- Fixed-effects, random-effects, variances, covariances, model fit indices, pseudo-R^2
-