

Multilevel Models: Longitudinal Analysis

2017-03-31

Preparation

We will use datasets located in the *minneapolis.csv* file. These data include student-level attributes for $n = 22$ students sampled from the Minneapolis school district. The source of these data is: Long (2012). We will use these data to explore the longitudinal change of student reading scores.

The attributes in the *nminneapolis.csv* file include:

- studentID: Student ID number
- read.5: Student's fifth-grade reading achievement score
- read.6: Student's sixth-grade reading achievement score
- read.7: Student's seventh-grade reading achievement score
- read.8: Student's eighth-grade reading achievement score
- atRisk: Is the student considered either eligible for free/reduced lunch or homeless highly mobile? (0 = no; 1 = yes)
- female: Is the student a female? (0 = Male; 1 = female)
- minority: Is the student a minority student? (0 = white student; 1 = non-white student)
- ell: Is the student an English language learner (ELL)? (0 = non ELL; 1 = ELL)
- sped: Does the student receive special education services? (0 = does not receive special education services; 1 = receives special education services)
- att: Student's attendance rate

```
# Read in player-level and team-level data
library(readr)
mplsWide = read_csv(file = "~/Google Drive/Documents/EPsy-8252/data/minneapolis.csv")
head(mplsWide)
```

studentID	read.5	read.6	read.7	read.8	atRisk	female	minority	ell	sped	att
1	172	185	179	194	1	1	1	0	0	0.94
2	200	210	209	NA	1	1	1	0	0	0.91
3	191	199	203	215	1	0	1	0	0	0.97
4	200	195	194	NA	1	1	1	0	0	0.88
5	207	213	212	213	1	1	1	0	0	0.85
6	191	189	206	195	1	0	1	0	0	0.90

```
# Load other libraries
library(AICcmodavg)
library(broom)
library(dplyr)
library(ggplot2)
library(lme4)
library(sm)
library(tidyr)
```

Wide Data versus Long Data

The `mplsWide` data are referred to as wide data. Wide data has the outcome variable (or predictors, or both) in multiple columns. (For us, the outcome we are interested in is reading achievement score.) When collecting longitudinal data, it is often convenient to enter the data into a spreadsheet in the wide format. Unfortunately, the wide format is not so good for analysis. Functions like `lm()` or `lmer()` require that the outcome variable is in a single column; referred to as long formatted data. Additionally, if there are predictors in the data, these functions require that each predictor is also in its own (single) column.

There are several functions to convert data back and forth between these two formats. We will use the `gather()` function from the **tidyr** package. `gather()` takes multiple columns, and gathers them into key-value pairs. We can also use the same type of piping we use with **dplyr**. We want to have key-value pairs of time points and reading scores. You can read more about using **tidyr** [here](#) and [here](#).

```
# Create long data from wide data (need tidyr package loaded)
mpls = mplsWide %>% gather(
  key = time, # Name of the variable that delineates the time points
  value = read, # Name of the outcome
  c(read.5, read.6, read.7, read.8) # The old variables you will use as values
)

# Arrange by ID number
mpls = mpls %>% arrange(studentID)
head(mpls, 12)
```

studentID	atRisk	female	minority	ell	sped	att	time	read
1	1	1	1	0	0	0.94	read.5	172
1	1	1	1	0	0	0.94	read.6	185
1	1	1	1	0	0	0.94	read.7	179
1	1	1	1	0	0	0.94	read.8	194
2	1	1	1	0	0	0.91	read.5	200
2	1	1	1	0	0	0.91	read.6	210
2	1	1	1	0	0	0.91	read.7	209
2	1	1	1	0	0	0.91	read.8	NA
3	1	0	1	0	0	0.97	read.5	191
3	1	0	1	0	0	0.97	read.6	199
3	1	0	1	0	0	0.97	read.7	203
3	1	0	1	0	0	0.97	read.8	215

The last part in our preparation is to create a numeric predictor for grade (note that currently the time predictor is a character vector). It is good practice to keep the character vector (or, better yet, a factor) and a numeric version of the predictor since some functions prefer one to the other. Here we use `mutate()` to create a numeric predictor called grade. Before we use `as.integer()` to turn the characters to numbers, however, we first turn time into a factor.

```
# Turn time into a factor
mpls = mpls %>% mutate(time = as.factor(time))

# Use as.integer() to turn factor into numbers
mpls = mpls %>% mutate(grade = as.integer(time))
head(mpls)
```

studentID	atRisk	female	minority	ell	sped	att	time	read	grade
1	1	1	1	0	0	0.94	read.5	172	1

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

Note that the way `as.integer()` works is by using the numeric information in the factor level. The first level in the time factor (after we converted it to a factor) is read.5. This was converted to a 1 when we used `as.integer()`; read.6 was converted to a 2; and so on. Because grade makes sense as a number, we will add four to each value so that read.5 will be 5; read.6 will be 6, etc. in the new predictor.

```
# Use as.integer() to turn factor into numbers
mpls = mpls %>% mutate(grade = as.integer(time) + 4)
head(mpls)
```

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

Exploratory Analysis

In longitudinal data analysis we are often interested in how the outcome changes over time. Here we want to know if reading scores are changing over time. When we examine this, we typically look at the mean score across time and the variation in scores over time.

```
mpls %>%
  group_by(time) %>%
  summarize(M = mean(read), SD = sd(read))
```

time	M	SD
read.5	205	20.0
read.6	211	20.1
read.7	216	19.4
read.8	NA	NA

Missingness

Since there are some students that do not have a reading score in the 8th grade, the computation of the mean and standard deviation for these grades defaulted to NA. It is typical to report both the count and percentage of missingness at each time point.

```
mpIs %>%
  group_by(time) %>%
  summarize(
    Miss = sum(is.na(read)),
    Miss_Percent = sum(is.na(read))/length(read)
  )
```

time	Miss	Miss_Percent
read.5	0	0.000
read.6	0	0.000
read.7	0	0.000
read.8	8	0.364

This pattern of later time points show more missing data is quite common in longitudinal data.

We will also compute the means and standard deviations for those students that we have data for. Below we use `filter()` to omit any NA values in the `read` variable before we do any computations.

```
mpIs %>%
  filter(!is.na(read)) %>%
  group_by(time) %>%
  summarize(M = mean(read), SD = sd(read))
```

time	M	SD
read.5	205	20.0
read.6	211	20.1
read.7	216	19.4
read.8	218	19.4

The pattern in the sample data suggests that the mean reading score increases over grade level. The variation in reading scores seems fairly constant over grade levels (equal variances?). We need to be careful about the inferences coming from the last time point. The summary measures are only based on student who had data (recall 36% were missing data at grade 8). This means the estimates are likely biased.

Another way to do this would be to use the optional argument `na.rm=TRUE` in the `mean()` and `sd()` functions. Then you wouldn't have to filter out the missing values.

```
mpIs %>%
  group_by(time) %>%
  summarize(M = mean(read, na.rm = TRUE), SD = sd(read, na.rm = TRUE))
```

Correlation of Reading Scores Over Time

It is probably easiest to compute the correlation between the repeated measures using the wide formatted data. Similar to the computation of the means and standard deviations the `cor()` function will produce NAs if there is any missing data. In the function, we specify how to treat missing values. Using pairwise complete observations, the correlation between reading scores at grade 5 and grade 8, is computed only using students that have data at both time points. Because of this the correlations computed between different sets of reading scores may be based on different students. Here, `mplsWide[2:5]` indicates that the correlation matrix should be based on columns 2 through 5 (the variables `read.5`, `read.6`, `read.7`, `read.8`).

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

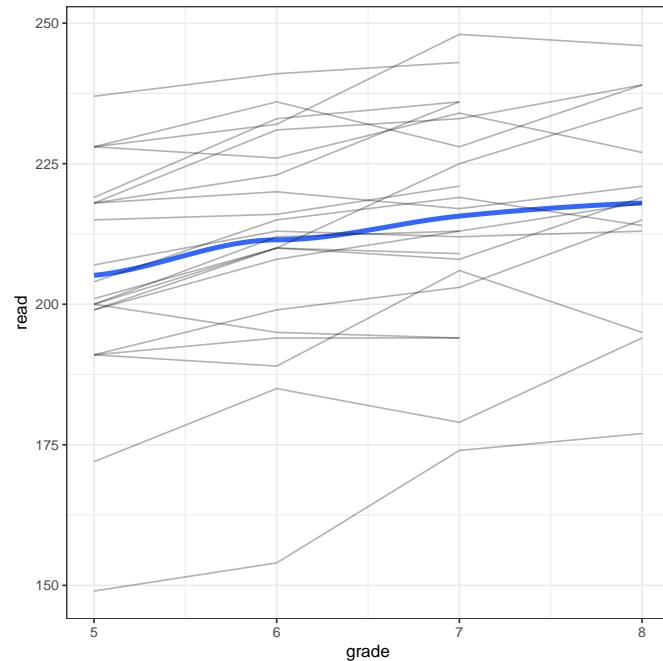
	read.5	read.6	read.7	read.8
read.5	1.000	0.965	0.914	0.883
read.6	0.965	1.000	0.926	0.929
read.7	0.914	0.926	1.000	0.923
read.8	0.883	0.929	0.923	1.000

The correlations show that the reading scores are correlated. They also suggest that reading scores that are further apart in time are less correlated (in statistical terms, a pattern of decay). This pattern in the correlations is quite common in longitudinal data.

Spaghetti Plots

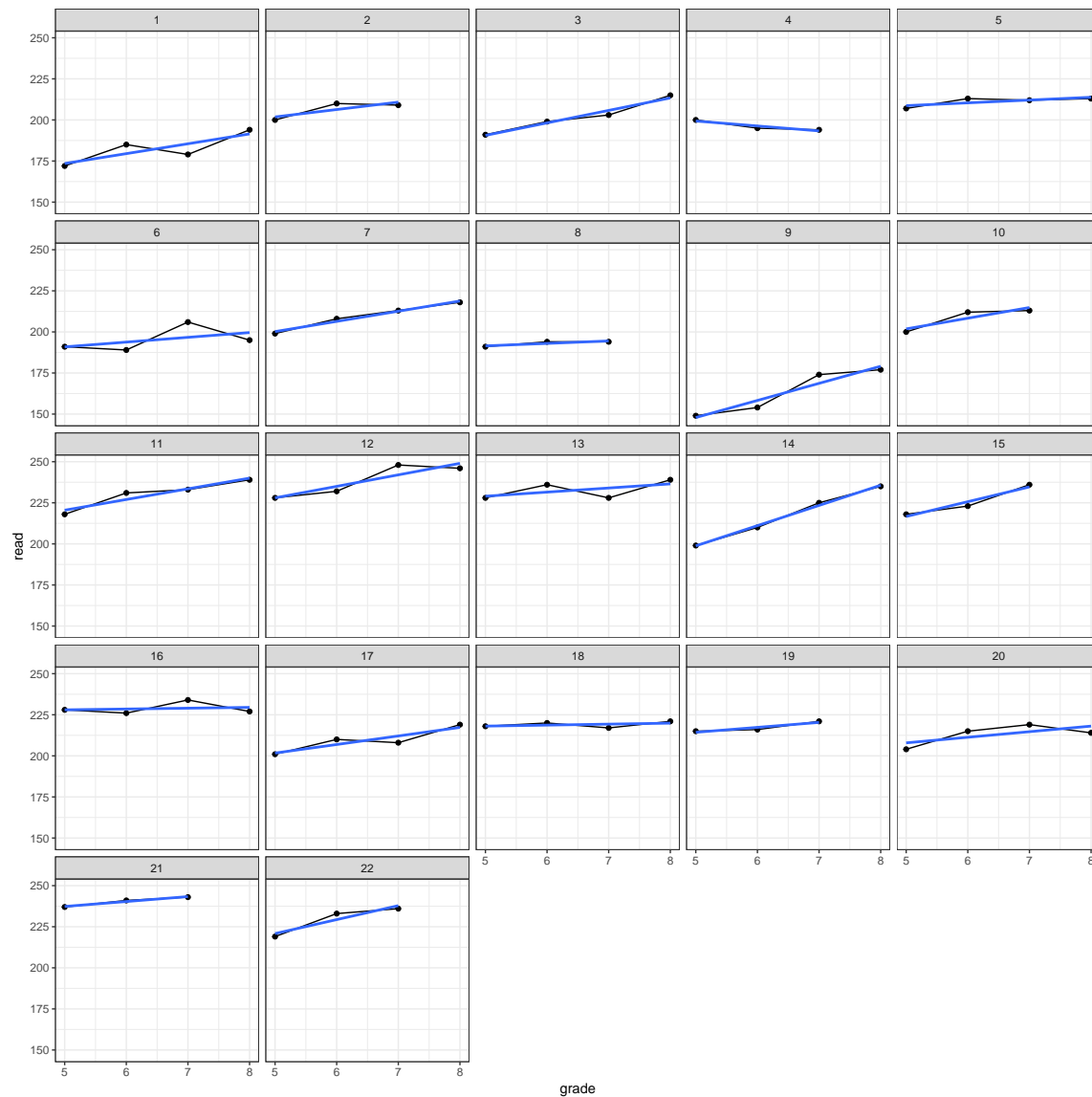
A spaghetti plot is simply a plot showing the trend in the outcome over time for each person. In our case, we will be plotting the reading scores for each student. Here we are looking for the functional form for the relationship between grade level and reading score (e.g., linear, quadratic). We also add the loess smoother to better examine the functional form.

```
ggplot(data = mpls, aes(x = grade, y = read, group = studentID)) +  
  #geom_point() +  
  geom_smooth(aes(group = 1), se = FALSE, lwd = 1.5) +  
  geom_line(alpha = 0.3) +  
  theme_bw()
```



The loess smoother suggests that the average pattern of growth could be linear. We can examine how well this pattern holds for each student by faceting on student ID. This will plot each student's pattern in a separate plot panel. In each panel, we also examine the fit to a linear trend.

```
ggplot(data = mpls, aes(x = grade, y = read, group = studentID)) +
  geom_point() +
  geom_line() +
  geom_smooth(aes(group = 1), se = FALSE, method = "lm") +
  theme_bw() +
  facet_wrap(~studentID)
```



There are several features these plots illustrate:

1. Most patterns are approximately linear (but not always) and the reading scores tend to increase over time.
2. Some OLS linear trajectories fit very well (e.g., Subject #7).
3. Other OLS linear trajectories do not fit as well (e.g., Subject #13).
4. Students vary in their achievement at grade 5.
5. There is variation in the linear rate of change across students (i.e., the slopes vary).

Unconditional Random Intercepts Model

Before we fit a linear function of time, we will first fit the unconditional random intercepts model. The multilevel model for this is:

$$\text{Level-1 Model : } \text{Read}_{ij} = \beta_{0j} + \epsilon_{ij}$$

$$\text{Level-2 Model : } \beta_{0j} = \gamma_{00} + u_{0j}$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $u_{0j} \sim \mathcal{N}(0, \tau_{00}^2)$. In these models, time points (level-1) are nested in students (level-2). As such, the i subscript represents a particular time point, and the j subscript represents a particular student. The composite model is:

$$\text{Read}_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij}$$

Initially we fit this model using REML estimation to get unbiased estimates of the error variances.

```
lmer.0 = lmer(read ~ 1 + (1 | studentID), data = mpls)
summary(lmer.0)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: read ~ 1 + (1 | studentID)
## Data: mpls
##
## REML criterion at convergence: 622
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.2140 -0.4810  0.0244  0.5237  2.2104
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## studentID (Intercept) 335.4    18.31
## Residual                66.2     8.14
## Number of obs: 80, groups: studentID, 22
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   212.24      4.01    52.9
```

The fitted unconditional level-1 model suggests that the mean reading score across students and grade levels is 212.2. Computing the intraclass correlation coefficient (ICC),

$$\begin{aligned} \rho &= \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} \\ &= \frac{335.38}{335.38 + 66.21} \\ &= 0.83 \end{aligned}$$

,

we find that an estimated 83.5% of the total variation in reading scores is attributable to differences between students, and 16.5% of the total variation in reading scores is attributable to differences between time points (within students).

Building the Unconditional Growth Model (Level-1 Model)

Having partitioned the total variation into within-students and between-students, we can now ask: What role does grade-level (time) play? Based on the exploratory analysis, a linear model relating grade-level and reading scores seemed appropriate. Below are the multilevel models that include the linear effect of grade-level in the level-1 model and maintains a random-effect of intercept:

$$\begin{aligned}\text{Read}_{ij} &= \beta_{0j} + \beta_{1j}(\text{Grade Level}) + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10}\end{aligned}$$

The composite (mixed-effects) model is:

$$\text{Read}_{ij} = \gamma_{00} + \gamma_{10}(\text{Grade Level}_{ij}) + u_{0j} + \epsilon_{ij}$$

We also consider a quadratic effect of grade level for completeness (here only the composite model is shown):

$$\text{Read}_{ij} = \gamma_{00} + \gamma_{10}(\text{Grade Level}_{ij}) + \gamma_{20}(\text{Grade Level}_{ij}^2) + u_{0j} + \epsilon_{ij}$$

In both models, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $u_{0j} \sim \mathcal{N}(0, \tau_{00}^2)$. Since the error structure and random effects are the same as that for the unconditional means model, we fit this model and re-fit the unconditional means model using ML estimation. We then use AICc to select the “best” level-1 structure.

```
# Fit models using ML
lmer.0.ml = lmer(read ~ 1 + (1 | studentID), data = mpls, REML = FALSE)
lmer.1.ml = lmer(read ~ 1 + grade + (1 | studentID), data = mpls, REML = FALSE)
lmer.2.ml = lmer(read ~ 1 + grade + I(grade^2) + (1 | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0.ml, lmer.1.ml, lmer.2.ml),
  modnames = c("Unconditional", "Grade level (L)", "Grade level (Q)")
)

myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	Grade level (L)	4	588	0.00	1.000	0.697	-290	0.697
3	Grade level (Q)	5	589	1.66	0.436	0.303	-289	1.000
1	Unconditional	3	633	45.00	0.000	0.000	-313	1.000

Here we adopt the model that includes the linear effect of grade level.

Graphically Examining the Level-1 Residuals: An Alternative Approach to Choosing the Level-1 Structure

Consider Student #1's fitted models for (1) the unconditional means model, (2) the unconditional growth model (linear), and (3) the unconditional growth model (quadratic). To obtain each of these fitted models, substitute in the values of the fixed effects, and of Student #1's random effect. (Base these on models fitted with ML.)

$$\text{Model 1: } \hat{\text{Read}}_{ij} = 183.9 + \hat{\epsilon}_{ij}$$

$$\text{Model 2: } \hat{\text{Read}}_{ij} = 150.8 + 5.0(\text{Grade Level}_{ij}) + \hat{\epsilon}_{ij}$$

$$\text{Model 3: } \hat{\text{Read}}_{ij} = 131.3 + 11.3(\text{Grade Level}_{ij}) - 0.5(\text{Grade Level}_{ij}^2) + \hat{\epsilon}_{ij}$$

Now consider the level-1 residuals for Student #1 for each of these models. To get these we are going to use the `augment()` function from the **broom** package to get the obtain the level-1 residuals from each model. We will also use `select()` to only keep the `studentID` and `.resid` columns, and `mutate()` to create a column (called `model`) that specifies the name of the model (to keep track of which model the residuals came from).

```
# Unconditional means model level-1 residuals
resid.0 = augment(lmer.0.ml) %>% select(studentID, .resid) %>% mutate(model = "lmer.0")

# Unconditional growth model (linear) level-1 residuals
resid.1 = augment(lmer.1.ml) %>% select(studentID, .resid) %>% mutate(model = "lmer.1")

# Unconditional growth model (quadratic) level-1 residuals
resid.2 = augment(lmer.2.ml) %>% select(studentID, .resid) %>% mutate(model = "lmer.2")

# Stack the three sets of residuals
all.resid = rbind(resid.0, resid.1, resid.2)

# Obtain only Student 1's residuals
res.1 = all.resid %>% filter(studentID == 1)
res.1
```

studentID	.resid	model
1	-11.97	lmer.0
1	1.03	lmer.0
1	-4.97	lmer.0
1	10.03	lmer.0
1	-3.68	lmer.1
1	4.34	lmer.1
1	-6.65	lmer.1
1	3.37	lmer.1
1	-3.32	lmer.2
1	3.81	lmer.2
1	-7.08	lmer.2
1	4.00	lmer.2

Student #1 has four level-1 residuals (one for each grade level) in each of the models. If the linear predictor of grade is necessary, then the residuals for the unconditional growth model (linear) should be smaller than those for the unconditional means model. Similarly, if the quadratic predictor of grade is necessary, then the

residuals for the unconditional growth model (quadratic) should be smaller than those for the unconditional growth model (linear).

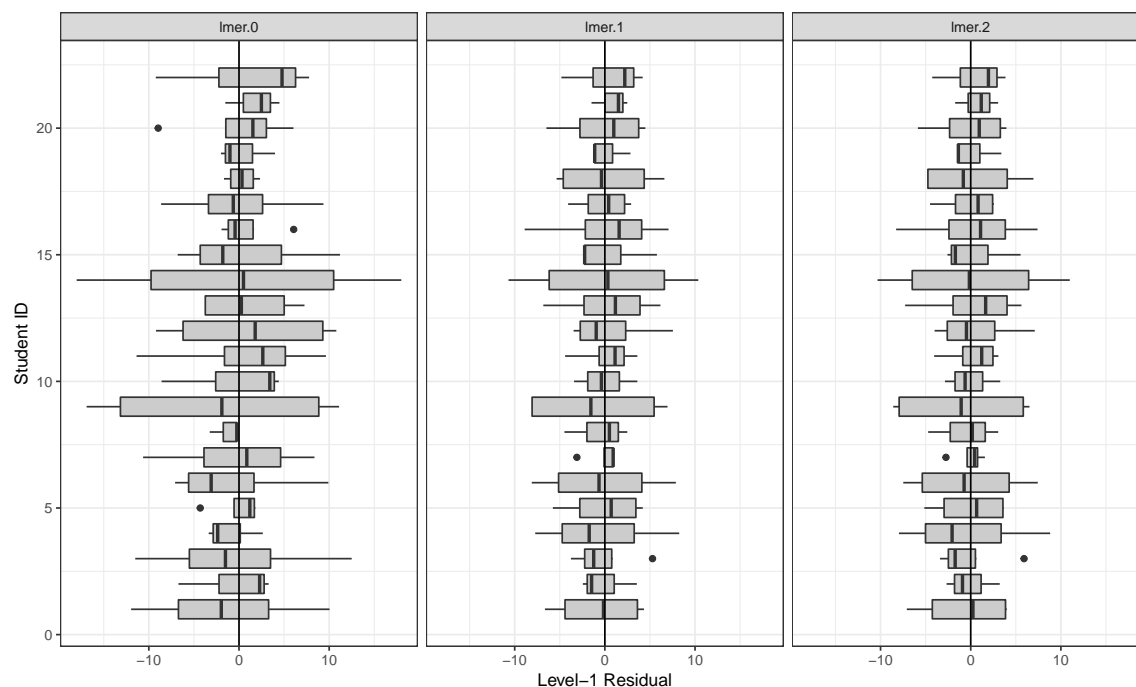
Below we compute the standard deviation for each model's residuals for Student #1.

```
res.1 %>% group_by(model) %>% summarize(SD = sd(.resid))
```

model	SD
lmer.0	9.33
lmer.1	5.36
lmer.2	5.48

For Student #1, the residuals are definitely smaller in the unconditional growth model (linear) than in the unconditional means model. There doesn't seem to be any improvement in the residuals when we add the quadratic predictor. In practice, we would look at the residuals for ALL the students, not just Student #1. Rather than computing the standard deviation and comparing those for all the students, we just plot the residuals for each student, and facet by model. Then we can visually examine the variation.

```
## Plot the residuals by subject ID to
ggplot(data = all.resid, aes(x = studentID, y = .resid, group = studentID)) +
  theme_bw() +
  geom_boxplot(fill = "grey80") +
  geom_hline(yintercept = 0) +
  facet_wrap(~ model) +
  xlab("Student ID") +
  ylab("Level-1 Residual") +
  coord_flip()
```



Adding the grade-level (time) predictor reduces the level-1 residual variation. Again, the quadratic predictor does not seem to really reduce the level-1 residual variation any further. This would suggest that a linear, but not a quadratic predictor should be included in the level-1 model.

This same information is quantified in the residual variance term. Consider the residual variance for each of the three models:

$$\text{Model 0 : } \hat{\sigma}_{\epsilon}^2 = 66.2$$

$$\text{Model 1 : } \hat{\sigma}_{\epsilon}^2 = 29.2$$

$$\text{Model 2 : } \hat{\sigma}_{\epsilon}^2 = 28.9$$

There is a big decrease in level-1 error variance from the unconditional means model to the unconditional growth model (linear). The residual variance from the unconditional growth model (linear) and the unconditional growth model (quadratic) are quite similar.

Choosing Random Effects Structure

Looking back at the spaghetti plot, it seems that not only are students reading scores different at grade 0 (evidence for random intercepts), but also that their growth trajectories are different (evidence for random slopes). We will examine two different random effects structures: (1) random intercepts only; and (2) random intercepts and random slopes. Because we are keeping the fixed-effects the same, it is better to evaluate these models if we have estimated using REML.

```
# Fit models using REML
lmer.1.int = lmer(read ~ 1 + grade + (1 | studentID), data = mpls)
lmer.1.both = lmer(read ~ 1 + grade + (1 + grade | studentID), data = mpls)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.1.int, lmer.1.both),
  modnames = c("Random Intercepts", "Random Intercepts + Slopes")
)
```

```
## Warning in aictab.AIClmerMod(cand.set = list(lmer.1.int, lmer.1.both), modnames = c("Random Intercepts", :
## Model selection for fixed effects is only appropriate with ML estimation:
## REML (default) should only be used to select random effects for a constant set of fixed effects
myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	Res.LL	Cum.Wt
2	Random Intercepts + Slopes	6	579	0.00	1.000	0.85	-283	0.85
1	Random Intercepts	4	582	3.47	0.176	0.15	-287	1.00

The evidence supports adopting a random effects structure that includes random intercepts and random slopes.

Interpretation of the Model

```
summary(lmer.1.both)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: read ~ 1 + grade + (1 + grade | studentID)
## Data: mpls
##
## REML criterion at convergence: 566
##
## Scaled residuals:
## Min      1Q Median      3Q      Max
## -1.592 -0.528 -0.046  0.515  2.023
##
## Random effects:
## Groups      Name      Variance Std.Dev. Corr
## studentID (Intercept) 783.66  27.99
##           grade       7.55   2.75  -0.75
## Residual          18.30   4.28
## Number of obs: 80, groups: studentID, 22
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  181.35      6.71  27.02
## grade         4.88      0.76   6.42
##
## Correlation of Fixed Effects:
##      (Intr)
## grade -0.799
```

- The mean reading score at grade level 0 is predicted to be 181.3.
- Each grade-level difference is associated with a 4.9-point difference, on average, in reading score.
- There is still unexplained within-student residual variance (error variance is not 0). This means we can include time-varying predictors in the model (if we have any).
- There is between-student residual variance in intercepts (reading scores at grade 0) since the RE variance component is not 0.
- There is between-student residual variance in rate of change (slopes are different) since the RE variance component is not 0.
- Estimated residual correlation between intercepts and rate-of-change is negative (students who are lower than average at grade 0 tend to grow faster than average, and vice versa)

It is common to provide covariances (rather than correlations) when reporting results from LMER models. To obtain the variance–covariance matrix of the random effects, we use the `varCorr()` function. Without accessing the `studentID` component, this function would just return the RE part of the `summary()` output.

```
# Get variance-covariance matrix of the random effects
```

```
VarCorr(lmer.1.both)$studentID
```

	(Intercept)	grade
(Intercept)	783.7	-57.32
grade	-57.3	7.55

- $\tau_{00}^2 = 783.7$
- $\tau_{11}^2 = 7.5$
- $\tau_{01} = -57.3$

All three estimates should be reported in a summary or write-up of results.

Centering a Level-1 Predictor

One thing we can do to make the intercept interpretable is to center the grade-level predictor. To do this, we will subtract five from the grade variable

```
mpls$c.grade = mpls$grade - 5
```

```
head(mpls)
```

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

Now we can re-fit the model using the centered predictor.

```
lmer.1.c = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls)
```

```
summary(lmer.1.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: read ~ 1 + c.grade + (1 + c.grade | studentID)
```

```
## Data: mpls
```

```
##
```

```
## REML criterion at convergence: 566
```

```
##
```

```
## Scaled residuals:
```

```
##   Min      1Q  Median      3Q      Max
```

```
## -1.592 -0.528 -0.046  0.515  2.023
```

```
##
```

```
## Random effects:
```

```
## Groups   Name      Variance Std.Dev. Corr
```

```
## studentID (Intercept) 399.12   19.98
```

```
##              c.grade      7.55    2.75   -0.36
```

```
## Residual                18.30    4.28
## Number of obs: 80, groups: studentID, 22
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  205.75      4.33    47.5
## c.grade      4.88       0.76     6.4
##
## Correlation of Fixed Effects:
##           (Intr)
## c.grade -0.360
```

- The mean reading score at grade 5 is predicted to be 205.7 (no more extrapolation).
- Each grade-level difference is associated with a 4.9-point difference, on average, in reading score (doesn't change since we didn't change the scale).

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

- There is still unexplained within-student residual variance (it is the same value as before).
- The intercept now pertains to students' grade 5 reading scores ... 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)

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

```
VarCorr(lmer.1.c)$studentID
```

	(Intercept)	c.grade
(Intercept)	399.1	-19.59
c.grade	-19.6	7.55

- $\tau_{00}^2 = 399.1$
- $\tau_{11}^2 = 7.5$
- $\tau_{01} = -19.6$

Plotting Results from Longitudinal 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). Typically we always plot the fixed-effects portion of the model, since that part of the model informs us of the relationship between grade level (time) and the reading scores (outcome).

$$\widehat{\text{Read}}_{ij} = 205.7 + 4.9(\text{c.grade}_{ij})$$

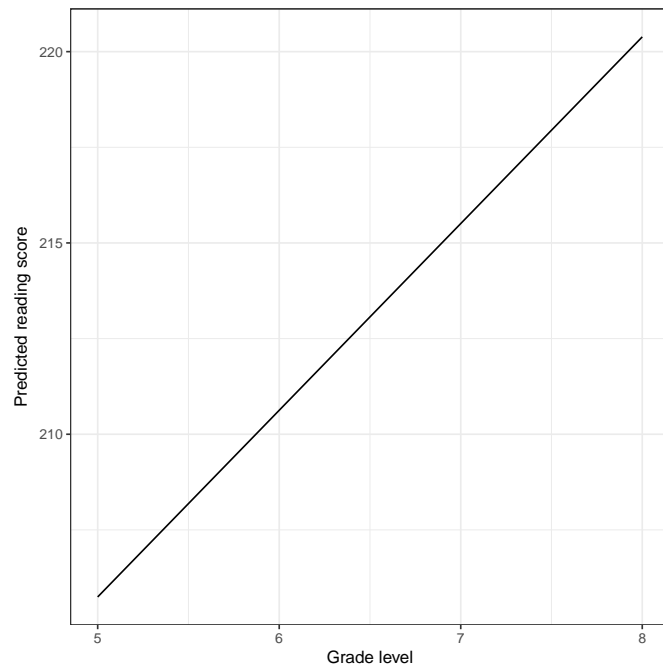
To plot the fixed effects model, we (1) create a dataset with the appropriate predictors; (2) use the model to estimate the \hat{y} values; and (3) plot.

```
# Create data set
plotData = expand.grid(
  c.grade = 0:3
)
```

```
# Compute y-hat values for the fixed-effects (average) model
plotData$yhat = predict(lmer.1.c, newdata = plotData, re.form = NA)
head(plotData)
```

c.grade	yhat
0	206
1	211
2	216
3	220

```
# Plot
ggplot(data = plotData, aes(x = c.grade, y = yhat)) +
  geom_line() +
  theme_bw() +
  scale_x_continuous(name = "Grade level", breaks = 0:3, labels = c("5", "6", "7", "8")) +
  ylab("Predicted reading score")
```



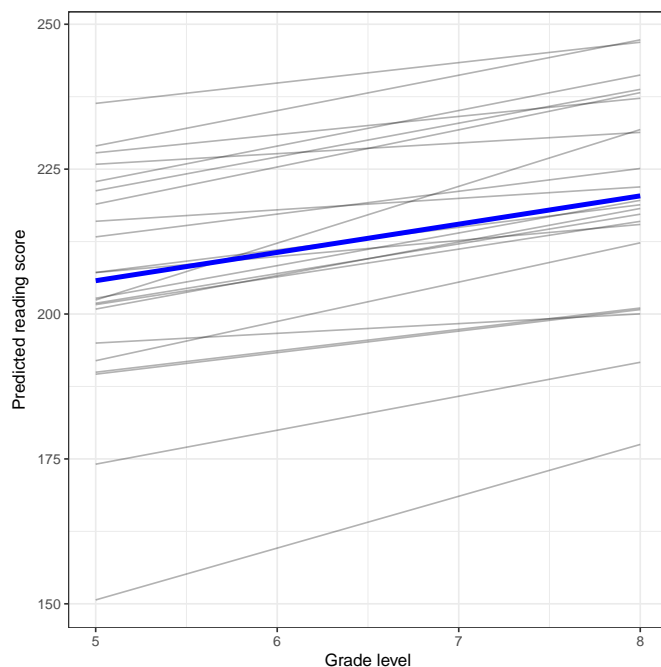
Plotting the Fixed- and Random-Effects

The process for plotting these is similar to plotting the fixed-effects. However, when we set up the initial data frame we also include the level-2 grouping variable (studentID) in `expand.grid()` along with any predictors. When we estimate y -hat values using the `predict()` function, we do not include the `re.form=` argument. Finally, when we plot, we group according to student ID so that we plot a different line for each student (i.e., random effects of intercept and slope). We can add the average line by including a second `geom_line()` layer drawing from the previous `plotData` object.

```
# Create data set
student_data = expand.grid(
  c.grade = 0:3,
  studentID = 1:22
)

# Use predict() to add the y-hat values
student_data$yhat = predict(lmer.1.c, newdata = student_data)

# Plot
ggplot(data = student_data, aes(x = c.grade, y = yhat, group = studentID)) +
  geom_line(alpha = 0.3) +
  geom_line(data = plotData, aes(x = c.grade, y = yhat, group = 1), color = "blue", lwd = 1.5) +
  theme_bw() +
  scale_x_continuous(name = "Grade level", breaks = 0:3, labels = c("5", "6", "7", "8")) +
  ylab("Predicted reading score")
```



Adding Level-2 Predictors

There are several different level-2 predictors we can consider:

- Whether the student is at risk (atRisk)
- Whether the student is female (female)
- Whether the student is a minority student (minority)
- Whether the student is an English Language Learner (ell)
- Whether the student is a special education student (sped)
- The student's attendance rate (att)

In thinking about these predictors, there are several ways to proceed through the model building process. Here, I will first examine each level-2 predictor independent of the other level-2 predictors. For each level-2 predictor, we will examine (1) the unconditional growth model; (2) the unconditional growth model with the level-2 predictor for intercept only; and (3) the unconditional growth model with the level-2 predictor for intercept and slope. Since the effects that are different are the fixed effects, we fit these models using ML. (Note that although the multilevel and composite models are not expressed in these notes, it is good practice to write these out as you are learning multilevel analysis.)

At-Risk

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + atRisk + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + atRisk + atRisk:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "atRisk - Intercept", "atRisk - Intercept + Slope")
)

myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	atRisk - Intercept	7	579	0.00	1.000	0.725	-282	0.725
3	atRisk - Intercept + Slope	8	582	2.33	0.312	0.226	-282	0.951
1	Unconditional Growth	6	585	5.39	0.068	0.049	-286	1.000

Female

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + female + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + female + female:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "female - Intercept", "female - Intercept + Slope")
)

myAIC
```

Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
Unconditional Growth	6	585	0.00	1.000	0.700	-286	0.700
female - Intercept	7	587	2.21	0.331	0.232	-286	0.932
female - Intercept + Slope	8	590	4.68	0.096	0.068	-286	1.000

Minority

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + minority + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + minority + minority:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "minority - Intercept", "minority - Intercept + Slope")
)

myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	minority - Intercept	7	575	0.00	1.000	0.677	-280	0.677
3	minority - Intercept + Slope	8	577	1.51	0.469	0.317	-279	0.994
1	Unconditional Growth	6	585	9.51	0.009	0.006	-286	1.000

English Language Learner

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + ell + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + ell + ell:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "ell - Intercept", "ell - Intercept + Slope")
)

myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
	Unconditional Growth	6	585	0.00	1.000	0.626	-286	0.626
	ell - Intercept	7	587	2.21	0.331	0.207	-286	0.834
	ell - Intercept + Slope	8	587	2.65	0.266	0.166	-285	1.000

Special Education

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + sped + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + sped + sped:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
```

```
# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "sped - Intercept", "sped - Intercept + Slope")
)

myAIC
```

Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
Unconditional Growth	6	585	0.000	1.000	0.462	-286	0.462
sped - Intercept	7	585	0.578	0.749	0.346	-285	0.809
sped - Intercept + Slope	8	587	1.769	0.413	0.191	-284	1.000

Attendance

```
lmer.0 = lmer(read ~ 1 + c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.1 = lmer(read ~ 1 + c.grade + att + (1 + c.grade | studentID), data = mpls, REML = FALSE)
lmer.2 = lmer(read ~ 1 + c.grade + att + att:c.grade + (1 + c.grade | studentID), data = mpls, REML = FALSE)

# AICc selection
myAIC = aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Unconditional Growth", "att - Intercept", "att - Intercept + Slope")
)

myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
3	att - Intercept + Slope	8	582	0.00	1.000	0.543	-282	0.543
2	att - Intercept	7	583	1.03	0.597	0.324	-284	0.867
1	Unconditional Growth	6	585	2.81	0.245	0.133	-286	1.000

Based on the analyses, we should include:

- The level-2 predictor atRisk in the intercept model
- The level-2 predictor minority in the intercept model
- The level-2 predictor att in the intercept and slope models

Here we fit that model using REML.

```
lmer.final = lmer(read ~ 1 + c.grade + atRisk + minority + att +
  att:c.grade + (1 + c.grade | studentID), data = mpls)
summary(lmer.final)$coefficients
```

	Estimate	Std. Error	t value
(Intercept)	346.6	102.30	3.39
c.grade	-30.2	18.53	-1.63
atRisk	-12.7	7.90	-1.61
minority	-22.4	7.98	-2.80
att	-126.2	103.73	-1.22
c.grade:att	36.7	19.35	1.90

From here it is a little uncertain as to how we should proceed. The t -ratios for the at-risk, attendance, and attendance by grade interaction are all less than two. This suggests that maybe these terms could be dropped. To examine this, it is worth fitting a series of models that drops those terms (starting with the interaction term and then the main-effects, starting with the lowest t -value).

```
# Drop att:time interaction term
lmer.final2 = lmer(read ~ 1 + c.grade + atRisk + minority + att +
  + (1 + c.grade | studentID), data = mpls)

# Drop att:time and att main-effect
lmer.final3 = lmer(read ~ 1 + c.grade + atRisk + minority +
  + (1 + c.grade | studentID), data = mpls)

# Drop att:time and att and atRisk main-effects
lmer.final4 = lmer(read ~ 1 + c.grade + minority +
  (1 + c.grade | studentID), data = mpls)
```

Now we will compute AICc information for these four models.

```
# AICc selection
myAIC = aictab(
  cand.set = list(lmer.final, lmer.final2, lmer.final3, lmer.final4),
  modnames = c("Full model", "-att:c.grade", "-att:c.grade, -att", "-att:c.grade, -att, -atRisk")
)
```

```
## Warning in aictab.AIClmerMod(cand.set = list(lmer.final, lmer.final2, lmer.final3, :
## Model selection for fixed effects is only appropriate with ML estimation:
## REML (default) should only be used to select random effects for a constant set of fixed effects
myAIC
```

Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	Res.LL	Cum.Wt
Full model	10	541	0.00	1.000	0.987	-259	0.987
-att:c.grade	9	550	8.64	0.013	0.013	-265	1.000
-att:c.grade, -att	8	559	17.70	0.000	0.000	-271	1.000
-att:c.grade, -att, -atRisk	7	564	23.07	0.000	0.000	-274	1.000

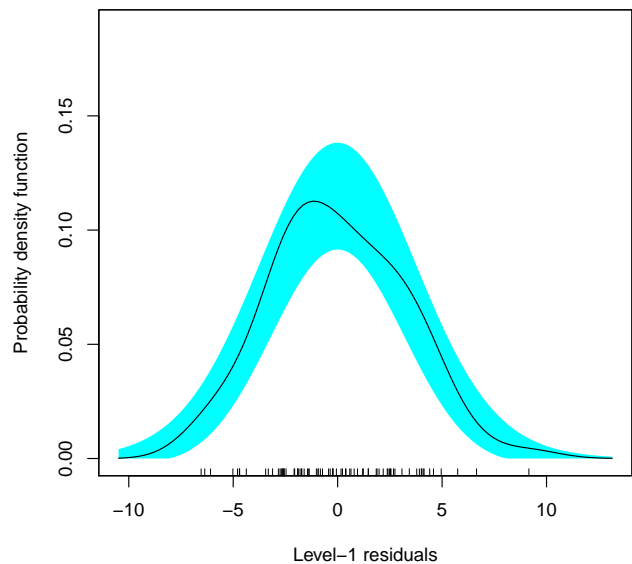
Using information criteria, we would adopt the model that includes the fixed-effects of time, attendance, at-risk, minority and the interaction between time and attendance.

Assumption Checking

Start with the examination of the level-1 residuals.

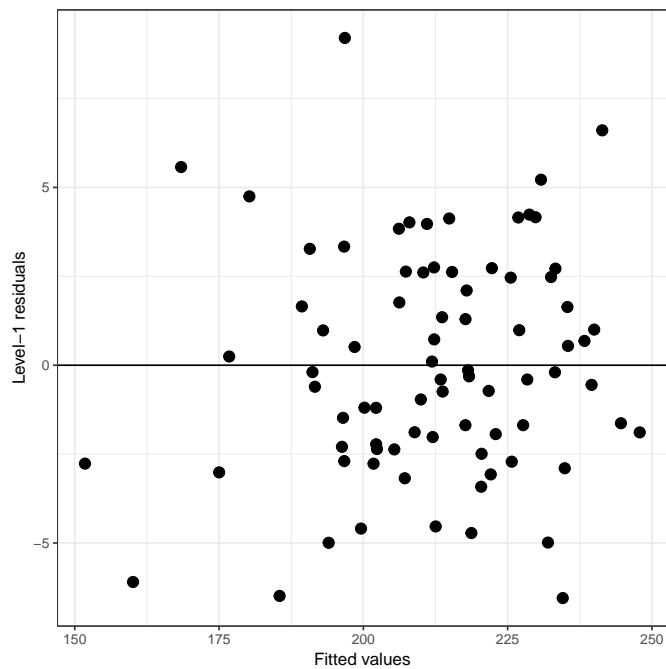
```
out = augment(lmer.final)
#head(out)

# Normality of level-1 residuals
sm.density(out$resid, model = "normal", xlab = "Level-1 residuals")
```



The assumption that the level-1 residuals are normally distributed looks satisfied. Next we will check the linearity and homogeneity of variance assumptions for the level-1 residuals.

```
ggplot(data = out, aes(x = .fitted, y = .resid)) +  
  geom_point(size = 3) +  
  theme_bw() +  
  geom_hline(yintercept = 0) +  
  xlab("Fitted values") +  
  ylab("Level-1 residuals")
```



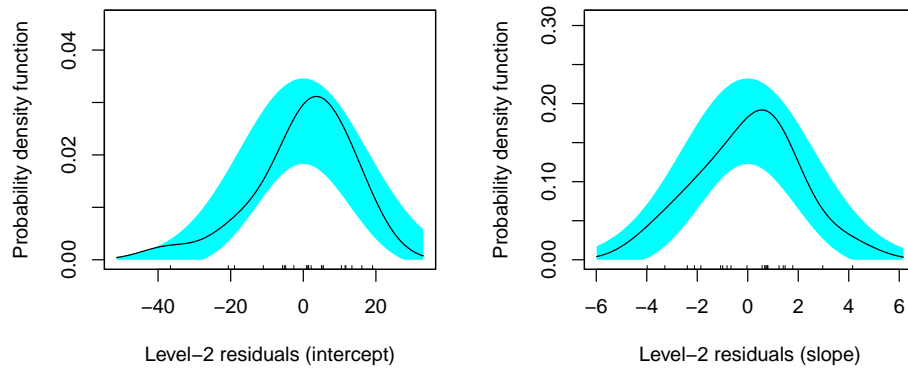
Both of those assumptions also look reasonably satisfied. Now we can examine the level-2 residuals for the

intercept and slope. In practice, typically we only examine the normality assumptions for the level-2 residuals.

```
# Compute the residuals
u0 = ranef(lmer.final)$studentID[ , 1]
u1 = ranef(lmer.final)$studentID[ , 2]

# Density plot of the intercept residuals
sm.density(u0, model = "normal", xlab = "Level-2 residuals (intercept)")

# Density plot of the slope residuals
sm.density(u1, model = "normal", xlab = "Level-2 residuals (slope)")
```



All the assumptions seem reasonably satisfied. The last thing to do would be to produce relevant model summaries and report a reasonable taxonomy of models. Here I would report:

- The unconditional means model
- The unconditional growth model
- The “final” model with the minority level-2 predictor

	Model A	Model B	Model C
<i>Fixed-effects</i>			
Grade level		4.88*** (0.76)	-30.21 (18.53)
At-risk			-12.74 (7.90)
Minority			-22.36** (7.98)
Attendance rate			-126.17 (103.73)
Grade level x Attendance rate			36.69 (19.35)
Intercept	212.24*** (4.01)	205.75*** (4.33)	346.57*** (102.30)
<i>Variance estimates</i>			
Student (Intercept)	335.38	399.12	209.73
Student (Grade level)		7.55	6.90
Covariance		-19.59	-4.07
Residual	66.21	18.30	17.74
<i>Goodness-of-Fit</i>			
AICc	628	579	541

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 25: Statistical models

Table 1. Coefficients (Standard Errors) for a Taxonomy of Fitted Multilevel Models to Predict 5th–8th Grade Reading Achievement Scores for $n = 22$ Students. Results are Reported Based on Fitting Each Model using Restricted Maximum Likelihood Estimation.

Since there is an interaction term, we should plot the model results to help interpret the fixed-effects of grade level and attendance rate. However the other estimated fixed-effects in the model can be interpreted:

- The mean reading achievement score for white, non at-risk students with a 0% attendance rate (extrapolation) at grade 5 is predicted to be 346.6.
- In the 5th grade, students who are at-risk are predicted to have a mean reading achievement scores that is 12.7 points lower then students who are not at-risk, controlling for differences in attendance rate and minority status.
- In the 5th grade, minority students are predicted to have a mean reading achievement scores that is 22.4 points lower then white students, controlling for differences in attendance rate and whether or not they are at-risk.

To interpret the effects of grade level and attendance rate, we would probably want to plot the fixed-effects of the adopted model.

```
# Create data set
fixed_effects = expand.grid(
  c.grade = 0:3,
  att = c(0.85, 0.95),
  atRisk = c(0, 1),
  minority = c(0, 1)
)

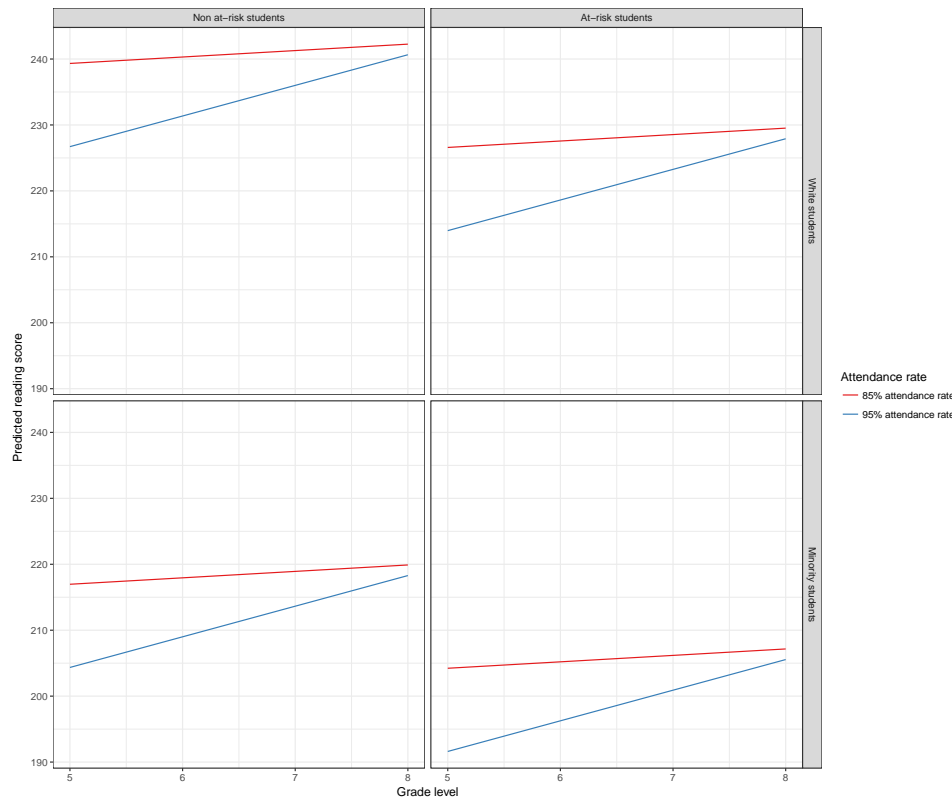
# Use predict() to add the y-hat values
fixed_effects$yhat = predict(lmer.final, newdata = fixed_effects, re.form = NA)

# Coerce any predictors not on the x-axis into factors for better plotting
fixed_effects$att = factor(fixed_effects$att, levels = c(0.85, 0.95),
  labels = c("85% attendance rate", "95% attendance rate"))

fixed_effects$atRisk = factor(fixed_effects$atRisk, levels = c(0, 1),
  labels = c("Non at-risk students", "At-risk students"))

fixed_effects$minority = factor(fixed_effects$minority, levels = c(0, 1),
  labels = c("White students", "Minority students"))

# Plot
ggplot(data = fixed_effects, aes(x = c.grade, y = yhat, color = att)) +
  geom_line() +
  theme_bw() +
  scale_x_continuous(name = "Grade level", breaks = 0:3, labels = c("5", "6", "7", "8")) +
  ylab("Predicted reading score") +
  facet_grid(minority ~ atRisk) +
  scale_color_brewer(name = "Attendance rate", palette = "Set1")
```



This plot displays the positive linear effect of grade level on reading achievement scores (for all groups of students). The magnitude of the grade level effect, however, is different depending on attendance rate. In general, the effect of grade level on reading achievement score is higher for students who have higher attendance rates, controlling for differences in risk and minority status.

References

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Thousand Oaks, CA: Sage Publications, Inc.