

Assumptions Underlying LMER Models

2017-11-05

The general model for the linear mixed-effects model is:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i$$

where

$$\epsilon_{ij} \stackrel{i.i.d}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$$

For the linear mixed-effects model with a linear fixed-effect of time and random-effects of both intercept and slope, the model would be:

$$Y_{ij} = \beta_0 + \beta_1(\text{Time}_{ij}) + b_0 + b_1(\text{Time}_{ij}) + \epsilon_{ij}$$

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

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

These assumptions state that:

- The level-1 errors are normally distributed.
- The level-1 errors are independent of one another.
- The level-1 errors have a mean of 0 (linearity).
- The level-1 errors have constant variance.
- All random-effects (intercept and slope) are normally distributed.
- All random-effects (intercept and slope) have mean of zero (linearity).
- There is a variance for each of the random-effects and a covariance between them.

Any time a model is fitted we need to evaluate these assumptions to ensure the validity of any results we interpret from the model.

Example: Vocabulary Data

Consider again the vocabulary data that we used in previous notes. (The importing and preparation of these data is not shown.) Here we fit the model that includes a fixed-effect of grade and random-effects of both intercept and slope to the data.

We can use the `augment()` function from the **broom** package to obtain the fitted-values and level-1 residuals from an LMER model.

```
library(broom)

# Augment the model to obtain residuals
out1 = augment(lmer.1)
head(out1)
```

##	score	grade	id	.fitted	.resid	.fixed	.mu	.offset	.sqrtXwt
## 1	1.75	0	1	1.797158	-0.04715805	1.413297	1.797158	0	1
## 2	2.60	1	1	2.545693	0.05430726	2.159953	2.545693	0	1
## 3	3.76	2	1	3.294227	0.46577257	2.906609	3.294227	0	1
## 4	3.68	3	1	4.042762	-0.36276212	3.653266	4.042762	0	1
## 5	0.90	0	2	1.206275	-0.30627544	1.413297	1.206275	0	1
## 6	2.47	1	2	1.951919	0.51808138	2.159953	1.951919	0	1

##	.sqrttrwt	.weights	.wtres
## 1	1	1	-0.04715805
## 2	1	1	0.05430726
## 3	1	1	0.46577257
## 4	1	1	-0.36276212
## 5	1	1	-0.30627544
## 6	1	1	0.51808138

In this output, the `.fitted` values are the predicted vocabulary scores based on the fixed- and random-effects. The `.mu` values are the predicted vocabulary scores based on the fixed-effects only. The `.resid` values are the residuals based on the difference between the observed vocabulary scores and the `.fitted` values (the level-1 errors).

To evaluate the assumptions, we will look for similar patterns in the residual plots that we do when we are evaluating the assumptions from a fixed-effects regression (however, we look at a slightly different set of plots to do this). Since we will need to look at the random-effects, it is useful to put the random-effects in a data frame for easy access in the plot syntax.

```
# Put RE in data frame
r_effects = ranef(lmer.1)$id

# Change the name of the first column from "(Intercept)" to "Intercept"
names(r_effects)[1] = "Intercept"

# Add ID (row number) to the data frame
r_effects = r_effects %>% mutate(ID = as.integer(row.names(r_effects)))

# Inspect the data frame
head(r_effects)
```

##	Intercept	grade	ID
## 1	0.3838612	0.001878439	1
## 2	-0.2070214	-0.001013067	2
## 3	-1.3312643	-0.006514591	3
## 4	1.2062818	0.005902985	4
## 5	-3.6299692	-0.017763390	5
## 6	-1.7404154	-0.008516788	6

Normality of the Level-1 Residuals and the Random Effects

```
# Load sm library
library(sm)

# Level-1 residuals
sm.density(out1$.resid, model = "normal", xlab = "Level-1 residuals")

# Random effect of intercept
sm.density(r_effects$Intercept, model = "normal", xlab = "Intercept random-effect")

# Random effect of slope
sm.density(r_effects$grade, model = "normal", xlab = "Linear slope random-effect")
```

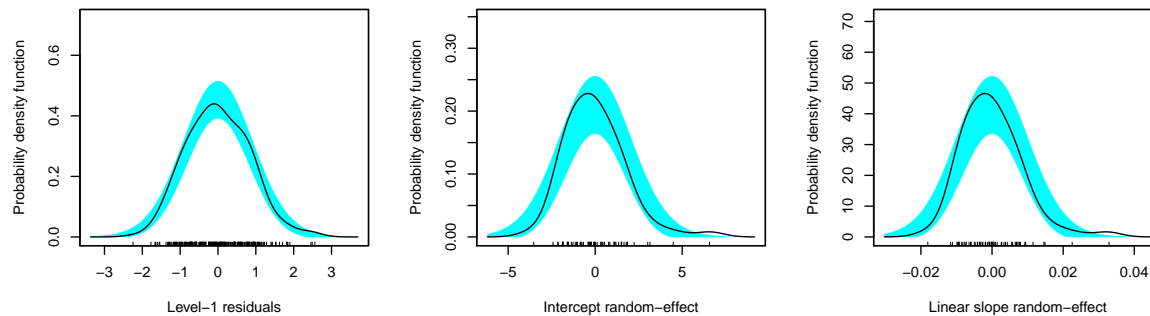


Figure 1. Density plots of the level-1 residuals (left), intercept random-effects (middle), and linear slope random-effects (right). The plots indicate that the normality assumptions are all tenable.

None of the three distributions show evidence that the normality assumption has been violated. The observed distributions all lie within the confidence band consistent with random sampling from a normal distribution. We would probably indicate that the assumptions about normality are reasonably satisfied for this model.

The right-hand side of the density plots of the random-effects maybe suggests some small deviation from normality (where the observed curve does not lie within the confidence band). Looking for these observations in the output of the random-effects, we find that it is a single observation (Row 36) that is deviating in both plots.

```
## Intercept      grade ID
## 1      6.6072 0.03233258 36
```

If we were concerned about this deviation, we could remove the observation, re-fit the model, and the re-check the assumptions. For now, I will leave the observation in, and check the other assumptions.

Linearity

The “linearity” assumption is about the growth trajectory for each individual. In a residual plot, checking this is about examining that the mean level-1 residual for each person is zero (random scatter around the line $y = 0$).

```
ggplot(data = out1, aes(x = id, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()
```

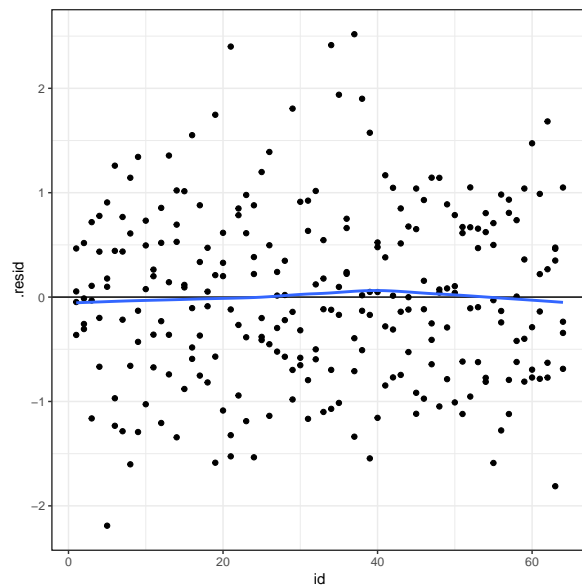


Figure 2. Plot of the level-1 residuals versus the ID values. The plot shows random scatter around the $y = 0$ line. The empirical loess smoother also shows little deviation from the $y = 0$ line.

The evidence from the residual plot of the level-1 residuals versus the ID values suggests that the linearity assumption is tenable and the mean residual for all individuals is zero.

```
p1 = ggplot(data = r_effects, aes(x = ID, y = Intercept)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()

p2 = ggplot(data = r_effects, aes(x = ID, y = grade)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()

gridExtra::grid.arrange(p1, p2, nrow = 1)
```

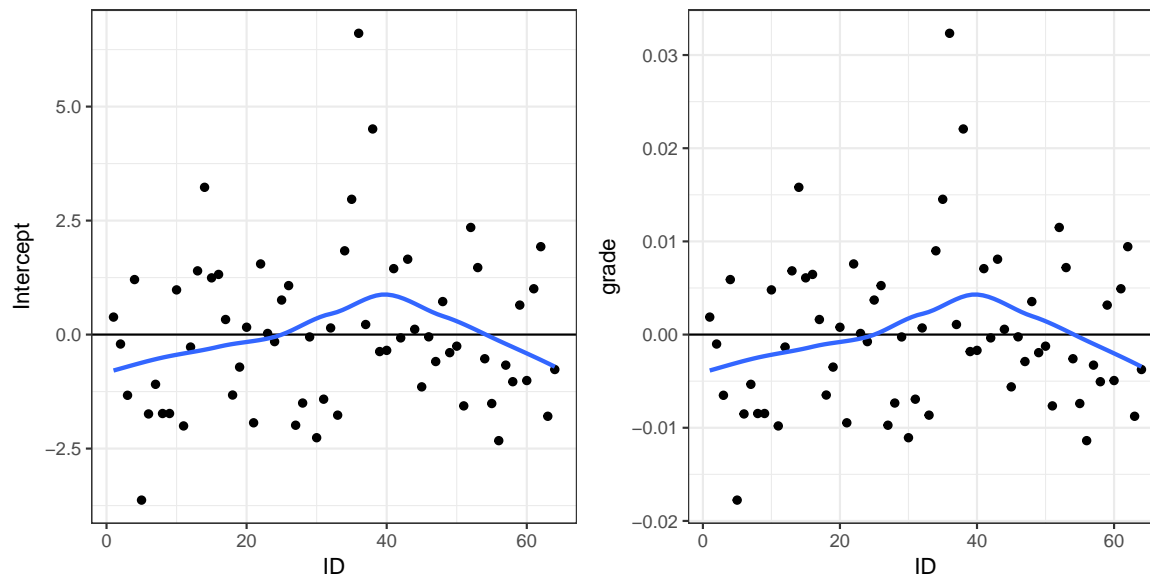


Figure 3. Plot of the random-effects versus the ID values. The intercept random-effects (left) and the linear slope random-effects (right) suggest that the linearity assumption is tenable.

The evidence from these two plots suggestst that the linearity assumption for the random-effects is tenable. The plots, in general, both show random scatter around the $y = 0$ line. The empirical loess smoother also shows little deviation from the $y = 0$ line. ID=36 again shows up as an extreme residual in these plots. This is where the loess smoother seems to deviate most from the $y = 0$ line.

Scaling Residuals and Random-Effects

Just like in fixed-effects regression, standardizing the residuals from the mixed-effects models can help us better identify outlying observations. To standardize the level-1 residuals we used the `resid()` function with the argument `scaled=TRUE`. Here we add a column of scaled residuals to the augmented data frame.

```
# Add column of scaled level-1 residuals
out1 = out1 %>%
  mutate( scl_resid = resid(lmer.1, scaled = TRUE))

head(out1)
```

```
##   score grade id  .fitted  .resid  .fixed  .mu .offset .sqrtXwt
## 1  1.75    0  1  1.797158 -0.04715805  1.413297 1.797158      0      1
## 2  2.60    1  1  2.545693  0.05430726  2.159953 2.545693      0      1
## 3  3.76    2  1  3.294227  0.46577257  2.906609 3.294227      0      1
## 4  3.68    3  1  4.042762 -0.36276212  3.653266 4.042762      0      1
## 5  0.90    0  2  1.206275 -0.30627544  1.413297 1.206275      0      1
## 6  2.47    1  2  1.951919  0.51808138  2.159953 1.951919      0      1
##   .sqrttrwt .weights  .wtres  scl_resid
## 1         1         1 -0.04715805 -0.04981056
## 2         1         1  0.05430726  0.05736189
## 3         1         1  0.46577257  0.49197096
## 4         1         1 -0.36276212 -0.38316646
## 5         1         1 -0.30627544 -0.32350257
## 6         1         1  0.51808138  0.54722199
```

```
# Plot of the scaled level-1 residuals vs. ID
ggplot(data = out1, aes(x = id, y = scl_resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()
```

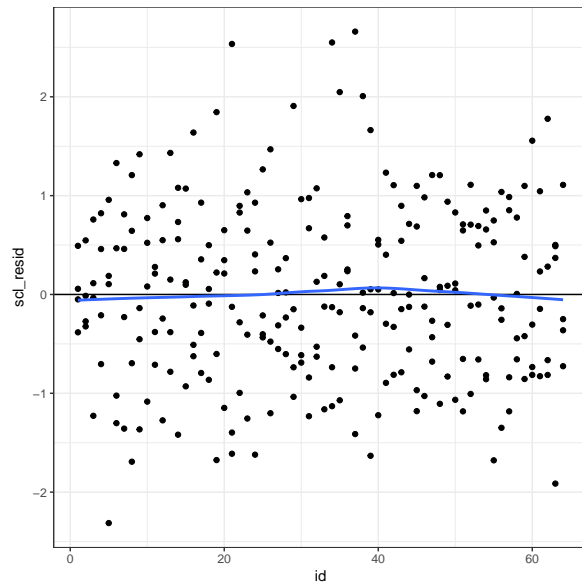


Figure 4. Plot of the scaled level-1 residuals versus the ID values.

Evaluating the assumption about linearity, we draw the EXACT same conclusion from this plot as we did from the plot using the unscaled residuals. Now, however, it is easier for us to say whether there are EXTREME residuals since the scale on the y -axis is now in standard errors. In this plot there does not appear to be any observations with extreme level-1 residuals.

To standardize the random-effects, we will divide the random-effects by the respective standard deviation of the RE. These can be obtained from the `summary()` output.

```
# Obtain the scaled REs
r_effects = r_effects %>%
  mutate(
    scl_intercept = Intercept / 1.771334,
    scl_grade = grade / 0.946748
  )

head(r_effects)
```

##	Intercept	grade	ID	scl_intercept	scl_grade
## 1	0.3838612	0.001878439	1	0.2167074	0.001984096
## 2	-0.2070214	-0.001013067	2	-0.1168732	-0.001070049
## 3	-1.3312643	-0.006514591	3	-0.7515603	-0.006881019
## 4	1.2062818	0.005902985	4	0.6810019	0.006235012
## 5	-3.6299692	-0.017763390	5	-2.0492856	-0.018762532
## 6	-1.7404154	-0.008516788	6	-0.9825450	-0.008995834

```
p1 = ggplot(data = r_effects, aes(x = ID, y = scl_intercept)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()

p2 = ggplot(data = r_effects, aes(x = ID, y = scl_grade)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()

gridExtra::grid.arrange(p1, p2, nrow = 1)
```

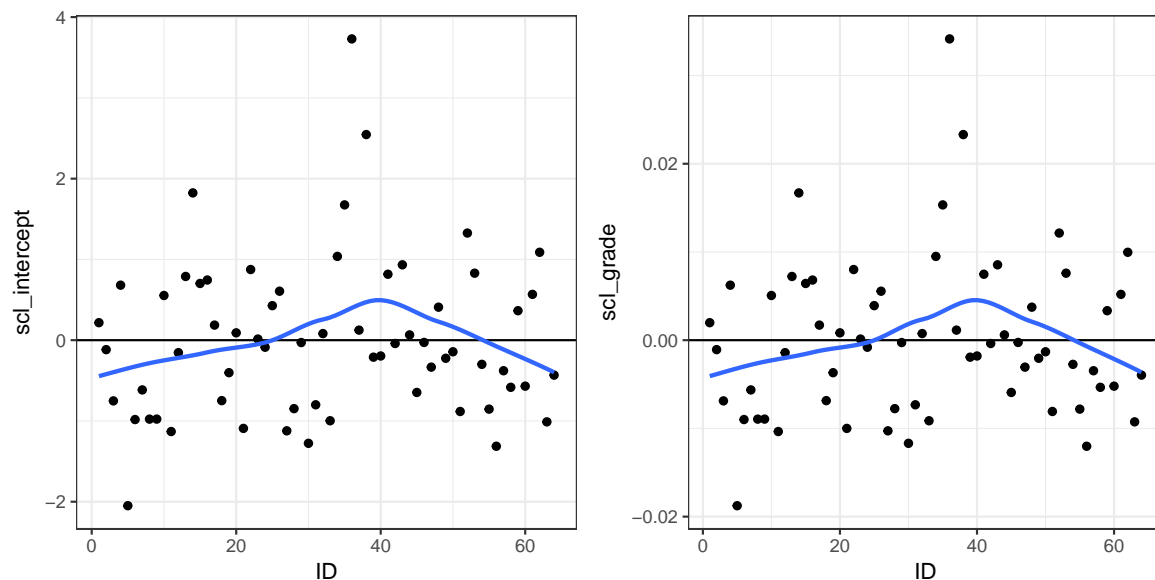


Figure 5. Plot of the scaled random-effects versus the ID values. Observation 36 seems to have an extremely high random-effect of intercept.

The random-effect of intercept for observation 36 now, indeed, looks to be extreme; it is almost four standard errors higher than we would expect. Again, we will hold off on removing this observation, although the evidence is starting to build to this decision.

Plots of the Residuals and Random-Effects versus the Predictors

We should also (1) plot the level-1 residuals versus each of the predictors in the level-1 model (grade) and (2) plot each of the random-effects versus each of the predictors in the respective level-2 models. Here we will also use the scaled residuals and random-effects.

```
# Plot scaled level-1 residuals vs. predictor in level-1 model
ggplot(data = out1, aes(x = grade, y = scl_resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw()
```

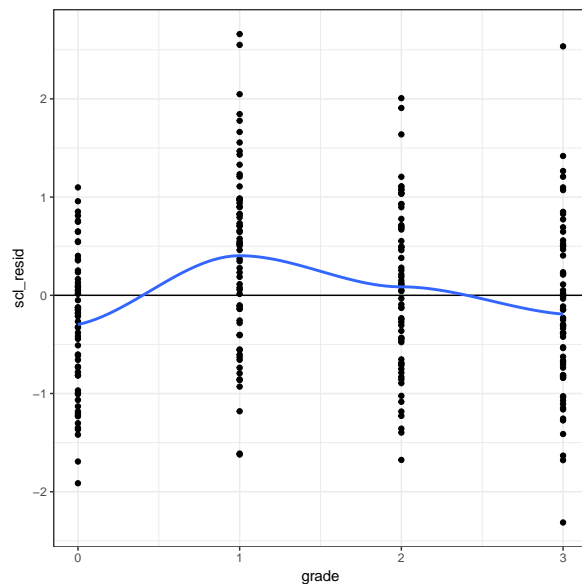


Figure 6. Plot of the scaled level-1 residuals versus grade.

There are no level-2 predictors, so we don't need to examine any plots for the REs here.

Constant Variance

To evaluate the constant variance assumptions, we will use the same plots we used to evaluate linearity. For the most part, in all of these plots, it seems that the constant variance assumption is satisfied. The exception might be in the plots of the random-effects where ID=36 seems to have more variation because of its large RE relative to the others.

Remove Observation 36

Here we will remove observation 36 from the data and re-fit the model. Then, we can re-evaluate the assumptions based on the re-estimated REs and residuals.

```
# Create new data set that removes ID=36
vocab_long_2 = vocab_long %>%
  filter(id != 36)

lmer.2 = lmer(score ~ 1 + grade + (1 + grade | id), data = vocab_long_2, REML = FALSE)
```

Now we will examine the same set of plots. The plots below are based on the scaled residuals and scaled random-effects. (Below I show the syntax to obtain the scaled residuals, but not the syntax to create the plots. This syntax would essentially be identical to previous syntax.) Here when we are adding the ID column to the random-effects data frame, we need to

```
# Level-1 residuals
out2 = augment(lmer.2) %>%
  mutate(scl_resid = resid(lmer.2, scaled = TRUE))
```

```
## Warning: Deprecated: please use `purrr::possibly()` instead
```

```
## Warning: Deprecated: please use `purrr::possibly()` instead
```



```
## Warning: Deprecated: please use `purrr::possibly()` instead
## Warning: Deprecated: please use `purrr::possibly()` instead
## Warning: Deprecated: please use `purrr::possibly()` instead
```

```
# Random-effects
r_effects = ranef(lmer.2)$id

# Change column name for Intercept
names(r_effects)[1] = "Intercept"

# Add ID column and scaled REs
r_effects = r_effects %>%
  mutate(ID = as.integer(row.names(r_effects))) %>%
  mutate(
    scl_intercept = Intercept / 1.541329,
    scl_grade = grade / 0.953593
  )

head(r_effects)
```

```
##      Intercept      grade ID scl_intercept  scl_grade
## 1  0.47754738  0.0027081036  1    0.30982833  0.0028398945
## 2 -0.09980751 -0.0005659942  2   -0.06475419 -0.0005935386
## 3 -1.19873138 -0.0067978359  3   -0.77772583 -0.0071286554
## 4  1.28111162  0.0072650025  4    0.83117337  0.0076185569
## 5 -3.44553188 -0.0195391231  5   -2.23542921 -0.0204900027
## 6 -1.59800591 -0.0090620651  6   -1.03677146 -0.0095030743
```

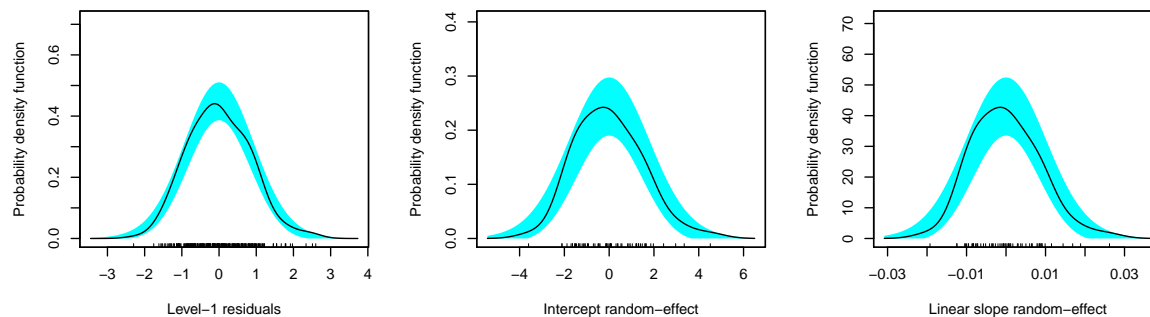


Figure 7. Density plots of the level-1 residuals (left), intercept random-effects (middle), and linear slope random-effects (right). The plots indicate that the normality assumptions are all tenable.

```
# Plot of the scaled level-1 residuals vs. ID
p1 = ggplot(data = out2, aes(x = id, y = scl_resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw() +
  ggtitle("Scaled Level-1 Residuals vs. ID")

p2 = ggplot(data = out2, aes(x = grade, y = scl_resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
```

```
theme_bw() +
ggtitle("Scaled Level-1 Residuals vs. Grade")
```

```
gridExtra::grid.arrange(p1, p2, nrow = 1)
```

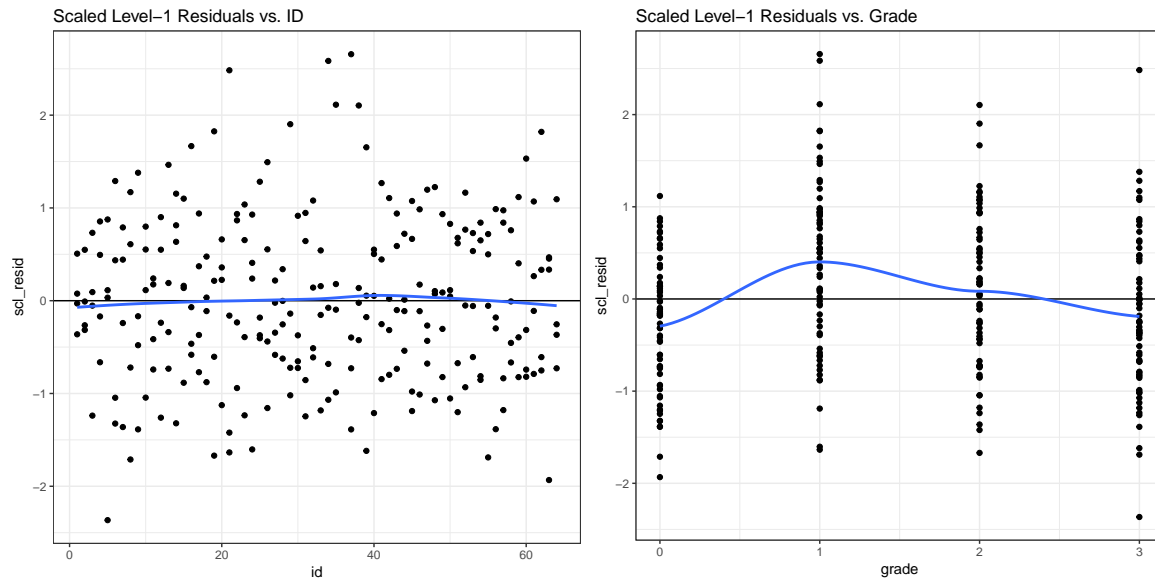


Figure 8. Plot of the scaled level-1 residuals versus ID value (left) and grade (right). The plots suggest that the assumptions of linearity and constant variance are both tenable.

```
p1 = ggplot(data = r_effects, aes(x = ID, y = scl_intercept)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw() +
  ggtitle("Scaled RE of Intercept vs. ID")

p2 = ggplot(data = r_effects, aes(x = ID, y = scl_grade)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth(se = FALSE) +
  theme_bw() +
  ggtitle("Scaled RE of Grade vs. ID")

gridExtra::grid.arrange(p1, p2, nrow = 1)
```

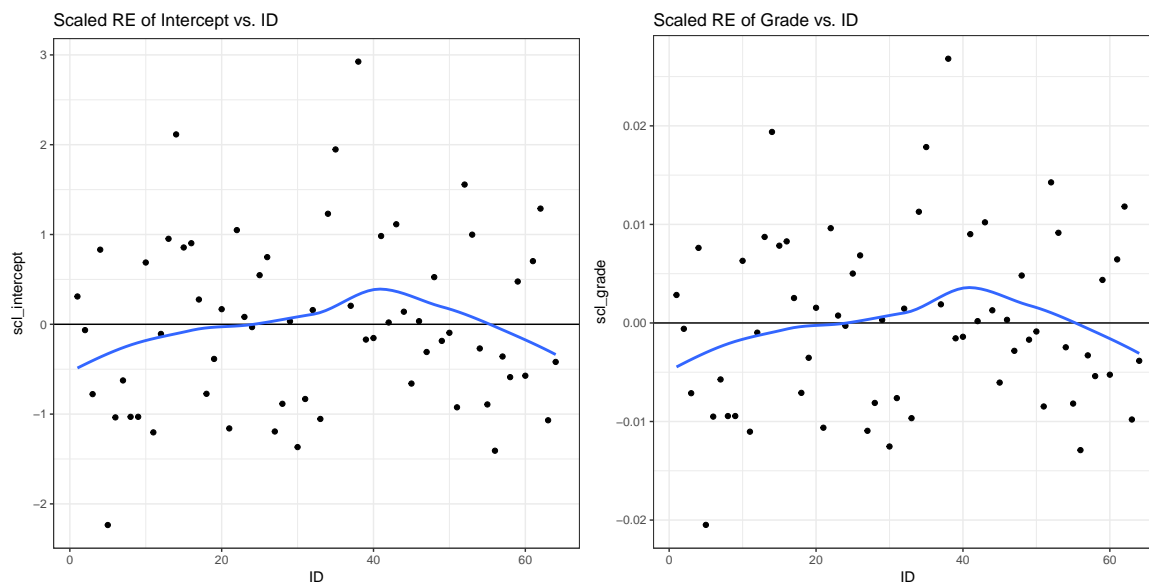


Figure 9. Plot of the scaled random-effects versus ID. The plots suggest that the assumption of linearity and constant variance are both tenable. The loess line does indicate some departure from linearity, even after removing Observation 36.

Here are the results from both models. (Note that we cannot use AIC or BIC to compare these models since they were fitted to different data sets.)

Table 1.

Results from Fitting a Mixed-Effects Regression Model to Predict Vocabulary Scores Over Time. Coefficient Estimates, (Standard Errors), and t-Values are Presented.

	All Obs. ($N = 256$)	Removing Obs. 36 ($N = 252$)
Grade	0.747 (0.053) 14.11	0.746 (0.054) 13.89
Constant	1.413 (0.243) 5.83	1.301 (0.219) 5.95

Removing the observations from ID=36 does not change the results much. In both sets of results, we would find an effect of grade on vocabulary scores. Since removing the observations associated with ID 36 does not seem to improve the residual fit of the model assumptions that much, it seems like we might be better reporting the results from the full data set.

Example 2: Model that Includes Sex as a Covariate

Consider if we had included the main-effect of gender into our model, the multi-level models for this are:

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

$$\beta_0^* = \beta_{00} + \beta_{01}(\text{Female}_i) + b_{0i}$$

$$\beta_1^* = \beta_{10} + b_{1i}$$

In this model, when we are evaluating the assumptions, we would look at:

- Density plot of the level-1 scaled residuals
- Density plot of the scaled intercept random-effects
- Density plot of the scaled linear grade random-effects
- Scatterplot of the level-1 scaled residuals versus the IDs
- Scatterplot of the scaled intercept random-effects versus the IDs
- Scatterplot of the scaled linear grade random-effects versus the IDs
- Scatterplot of the level-1 scaled residuals versus the grade
- Scatterplot of the scaled intercept random-effects versus female

Since we have included a level-2 predictor in the intercept equation, we would also evaluate the residuals of that model (the RE of the intercept) versus the included predictors from that model.

If we had included the main-effect of gender and the interaction between gender and grade into our model, the multi-level models are:

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

$$\beta_0^* = \beta_{00} + \beta_{01}(\text{Female}_i) + b_{0i}$$

$$\beta_1^* = \beta_{10} + \beta_{11}(\text{Female}_i) + b_{1i}$$

In this model, when we are evaluating the assumptions, we would look at:

- Density plot of the level-1 scaled residuals
- Density plot of the scaled intercept random-effects
- Density plot of the scaled linear grade random-effects
- Scatterplot of the level-1 scaled residuals versus the IDs
- Scatterplot of the scaled intercept random-effects versus the IDs
- Scatterplot of the scaled linear grade random-effects versus the IDs
- Scatterplot of the level-1 scaled residuals versus the grade
- Scatterplot of the scaled intercept random-effects versus female
- Scatterplot of the scaled linear grade random-effects versus female

References

- Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Thousand Oaks, CA: Sage Publications, Inc.
- Loy, A., & Hofmann, H. (2014). HLMdiag: A suite of diagnostics for hierarchical linear models in R. *Journal of Statistical Software*, 56. Retrieved from <https://www.jstatsoft.org/article/view/v056i05>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.