Formatting

* Resize figures again (fig.width and fig.height)
* Update tables with kable(booktabs = TRUE, digits = 2 to 4)   (no more than 4)
* Update tables with kable_styling(latex_options = "HOLD_position")
* Customize table column names and row names (mutate(fct_recode()))

Analysis

* Consider mean-centering age for interpretability of the intercept (and interaction term if you end up including it).

* Should be fitting the model with continuous time values instead of categorical wave treated continuously (i.e., recode wave as time = time-from-study-start and re-fit model)

* I thought about the model some more, and with such a small sample size I am (re)questioning the addition of the random slope term. I think you should still fit and include the model with random slopes, but similar to our argument for reducing the number of fixed effects, maybe we should favor the more parsimonious model (random intercepts only) with fewer parameters to estimate.

* Revisit interpretations. Remember that early alcohol use is the primary predictor, so should be the initial focus, followed by how BMI changes over time. Even though alcohol use is not significant, you should still address the question (e.g., "There does not appear to be a difference in BMI between…, however we did find a significant increase in average BMI over time. Specifically,….").

For the linear model, the coefficients are the same for either cluster-specific (for a person whose random effects are all 0, whom you might call the "typical" person) or population-averaged (for the average person) interpretations. So you can take either approach. I emphasized the cluster-specific version because I have been in the GLMM world for so long, but you can actually stick with the standard (STAT 135/230) population-averaged interpretation for these coefficients. So sorry!

Chapter 3 Application

3.1 Application to Longitudinal Study about Children's Health

In this chapter, we will apply linear mixed models to a longitudinal study about children's health, and explore how inference of fixed effects can possibly change when using various degrees of freedom approximation methods.

3.1.1 Background

Condense and clarify, e.g.: "surveyed a US sample of X students across 7-12th grade (ages 13 to 18) in the 1994-95 school year. Four waves of data collected in 1994, 199…."

The National Longitudinal Study of Adolescent to Adult Health is a longitudinal study spanning 1994 to 2008 that surveyed a U.S sample of students in 7-12th grade in the 1994-95 school year (Harris & Udry, 2022). Four waves of data were collected, in which the sample during the first wave was aged 13-18. Questions about mental health, socioeconomic status, and family background were collected, as well as physical measurements of height and weight.

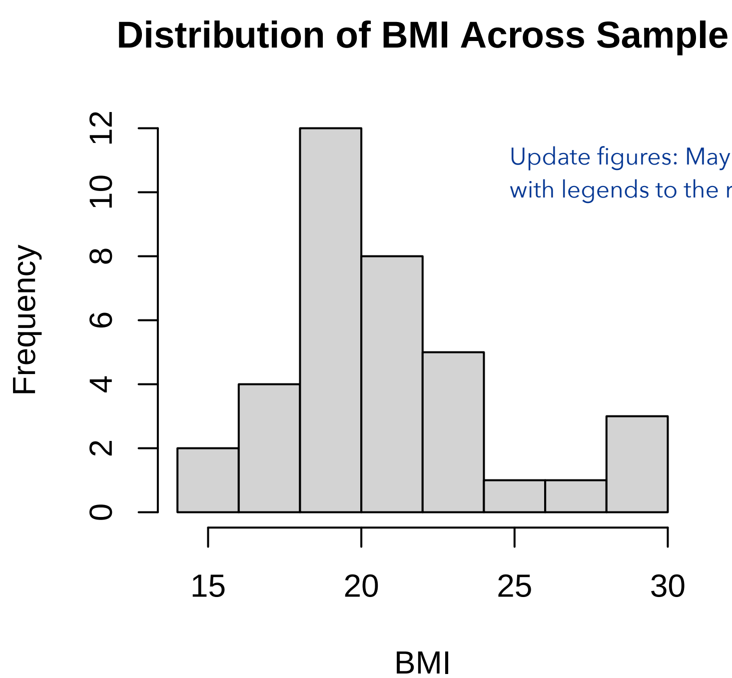
One question of interest to consider is how salient life experiences that occur during adolescence, such as being exposed to alcohol or being in a physical altercation, may impact changes to one's physical health over time. One way to capture physical health is through BMI, which tends to? ~~is shown~~ to follow a skewed Right-skewed? (Non-normal is implied) ~~nonnormal~~ distribution (Penman & Johnson, 2006). As we can also see in our dataset (Figure Y) With this in mind, we can employ methods of degrees

The adjustment is for small sample sizes, not non-normal distributions, so this does not follow from the previous statement.

of freedom adjustment. While this dataset is large and encompasses approximately 5,000 students, the scope of this application will be narrowed in order to examine the performance of degrees of freedom methods, which are sensitive to sample size.

We will be focusing on Chinese female respondents who completed at least three waves of the study, which amounts to a sample size of 12. It is hypothesized that early exposure to substances could potentially affect changes in weight. Previous studies have demonstrated a significant relationship between alcohol consumption and BMI (Pasch, Velazquez, Cance, Moe, & Lytle, 2012). Thus, our model will use alcohol use as the predictor of BMI. At the first wave of the study in 1994, children were asked whether they had consumed alcohol before. This, along with their age at the start of the first wave, will be used to model changes in BMI across three waves.

3.1.2 Exploration



Update figures: Maybe 2.4 high by 3.6 wide (or ~4 in for figures with legends to the right)

From this histogram depicting distribution of BMI, we can see that it follows a nonnormal trend. The skewness is .95 and the kurtosis is 3.39. How does this compare to any of the scenarios you set up in your simulation

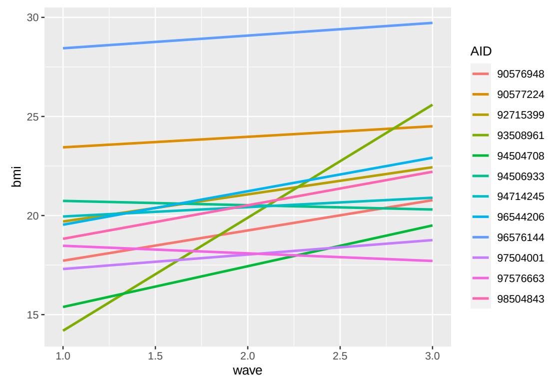
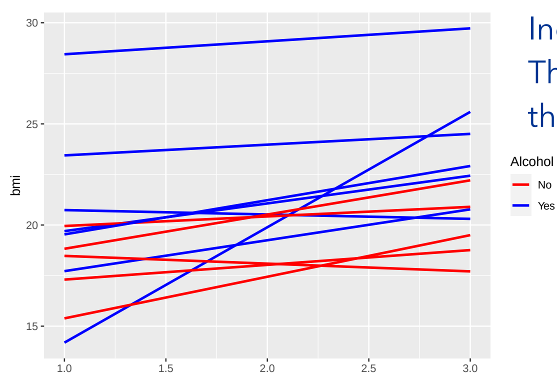


Figure 3.1: Changes in BMI by Individual

Figure 3.2 (will update automatically to 3.1 after you delete the current Figure 3.1, but change the reference tag to the tag for the current Fig 3.2)

Figure 3.1 depicts changes in BMI over time for each individual in the study. The wide variation in BMI values between individuals suggest that including random effects for the intercept would be beneficial. Additionally, different rates of change in BMI across the 3 waves between individuals would suggest that an additional random effect for time is worth exploring.



Include just this graph! The previous graph shows the same information.

Figure 3.2: Changes in BMI by Individual and Alcohol Use

In a longitudinal study examining youth substance use and body composition, Pasch et al. (2012) found that baseline alcohol use is associated with lower values of BMI at follow-up. The opposite pattern can be seen in Figure 3.2. Individuals who

Include argument in kable_styling(latex_options = “HOLD_position”) to put tables right after corresponding code chunks and in kable(book_tabs = TRUE) to clean up the tables (get rid of most vertical and horizontal lines).

Table 3.1: Linear Model Output

Customize the terms (mutate(fct_recode()), I think) and the column names.

Consider centering age to aid in interpretability

term	estimate	std.error	statistic	p.value
(Intercept)	9.71	4.736	2.05	0.049
alcohol1	3.13	1.044	3.00	0.005
age_at_start	0.42	0.288	1.46	0.154
wave	1.32	0.629	2.10	0.044

Are waves evenly spaced? Use years (time) instead of wave for interpretability (time = 0 for wave 1, time = X for wave 2 if wave 2 is X years later, …)

had consumed alcohol during the first wave had higher BMI values across all time points. While the data may not be consistent with other research, it highlights the importance of substance use as a potential predictor of BMI.

This paragraph is perfect!

While our sample size is small, our initial exploration of alcohol use suggests that it may be salient predictor for BMI. In addition, there are unique patterns of BMI change over time by individual, which suggests that creating a mixed effects model with random effects for time and intercept may be useful.

3.1.3 Linear Mixed Model

Because we have repeated measurements of the same individual in this study, a regular linear model would not be appropriate as it assumes independence; implementing this model would inflate Type I error rates. Table 3.1 shows the results of implementing a linear model with alcohol, age at first wave, and time as predictors. Alcohol use and time are significant predictors in this model. This will only used as a comparison to the other linear mixed models.

I’ve gone back and forth on this a bit and finally landed on yet a different recommendation (sorry!!): it looks like slopes are effectively parallel for 9 or so of the individuals, and only 2 or 3 have slightly flat/negative or extremely steep slopes. Given the small sample size, perhaps we should think about a simpler model and only include random intercepts. Justify this above with discussion of spaghetti plot, and then you don’t need to re-justify it here; just go into discussing the results. Then you can still present the results of the model with the random slopes, and just refer back to the potential concerns with over-parameterization given the small sample size.

3.1.4 Intercept Only Model

~~We would hypothesize that changes in weight may vary less within a person, as opposed to comparing changes in weight across two separate individuals. Thus, we fit an intercept only model, allowing the intercept to vary by individual. Alcohol, age at first wave, and time are the fixed effects.~~ ~~First, we can examine the output of the fixed~~

Customize these tables

Table 3.2: Fixed Effects for Random Intercept Model Using KR

term	estimate	std.error	statistic	df	p.value
(Intercept)	9.71	7.274	1.335	9.22	0.214
alcohol1	3.13	1.653	1.897	9.00	0.090
age_at_start	0.42	0.456	0.921	9.00	0.381
wave	1.32	0.403	3.272	23.00	0.003

Table 3.3: Random Effects for Random Intercept Model Using KR

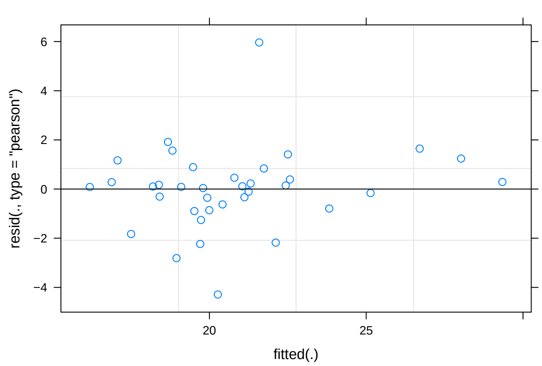
group	term	estimate
AID	sd__(Intercept)	2.57
Residual	sd__Observation	1.97

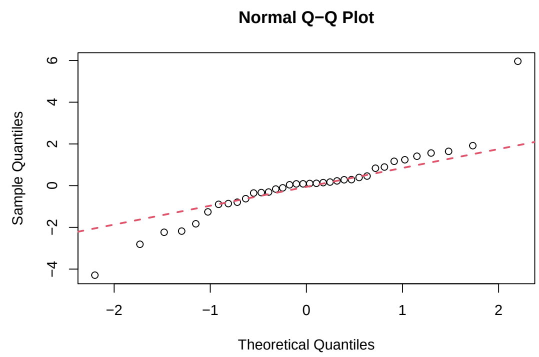
Revisit the fixed effect(s) interpretation after updating the model. What about impact of alcohol use? Careful not to be overly precise (0 or 1 decimal place is sufficient when talking changes in BMI). Also, there's only one random effect here, so this is an interpretation for individuals who share the same value of the random intercept.

~~effects similar to how a regular linear regression model is interpreted~~. Table 3.2 shows that time is the only significant effect. This means that conditional on the random effects, each wave increases the predicted BMI by 1.32. Next, we turn to the random effects output in table 3.3. The variance for individuals (represented by `AID`), which depicts variability across individuals, is 6.626, while the residual variance, representing within-subject variability is 3.895. ~~The significantly larger variance across individuals compared to within individuals suggests that this model is more optimal than a regular linear regression model since differences in variability are apparent.~~ The intraclass ~~interclass~~ correlation is .63, which indicates that BMI ~~weight~~ measurements taken of the same individual have slightly higher similarity than those of different individuals.

variability of what? Be specific

Any model that doesn't assume independence is necessary because of the repeated measures. (I think you've already made that argument in the beginning of the chapter, so no need to repeat).





The residual-vs-fitted plot doesn't add anything, so I would exclude it. Generally we don't talk about or show residual plots, but it's useful here to focus in particular on the Q-Q plot since it is validating, again, that not only is the outcome non-normal but even after fitting our model our residuals are still non-normal (sometimes regressing a skewed variable on another skewed variable results in residuals that are just fine, but not here! In terms of organization, I might lead with the residuals so that you can continue on with interpretation without the brief interruption to discuss residuals only to go back to interpretation.

The residuals vs fitted plot indicates no heterogeneity in the residuals, however, the normality assumption of the residuals is not met as seen in the QQ plot. This is to be expected as the random effects are nonnormal. We will proceed with evaluating the fixed effects in order to see how DF methods perform when normal assumptions do not hold.

Since the coefficients themselves don't change, I would be interested in a table like this but may be too complicated or time-consuming to figure out:

Term	Coef	SE KR	SE Sat	df KR	df SAT	p-value KR	p-value SAT
Intercept							
Alcohol use							
Age							
Time							

Table 3.4: Comparison of Fixed Effects P-Values in Random Intercept Model

term	Satterthwaite_p.value	KR_p.value
(Intercept)	0.214028860057237	0.214028858602635
alcohol1	0.0903585620561796	0.0903585607408565
age_at_start	0.380940524102144	0.380940522986174
wave	0.00335031625142439	0.00335031627116489

For the sake of time, I would take the full coefficient table for both models instead (cleaned up) instead of Table 3.4 (so make Table 3.3 the Satterthwaite results, and Table 3.4 a comparison of both Satterthwaite and KR random effect estimates).

The summary output referenced above uses Kenward-Roger DF approximation. Table 3.4 compares the summary output comparing Satterthwaite and Kenward-Roger. There is no significant difference between performance of the two DF methods, which aligns with results of our random intercept models from our simulation study

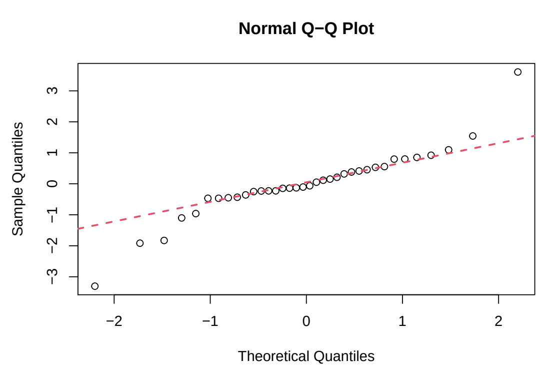
I would introduce this model by saying why addition of a random slope may be problematic given the sample size, and then just in a sentence or two simply describe how the results compare (e.g., time is no longer significant as a fixed effect, and alcohol use is significant only if we use Satterthwaite's df; however, KR is likely most appropriate here, which highlights the importance of having done the simulation in the first place (because that's what the simulation would recommend, and choosing otherwise may result in possibly falsely concluding that early alcohol use has an impact on BMI over time, after adjusting for age and person-specific variability in BMI). I would not re-interpret effects (3.2 vs 3.4 is not really different anyway).

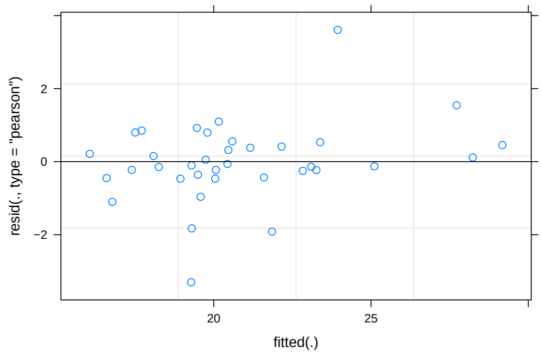
3.1.5 Intercept and random slope

Table 3.5: Fixed Effects for Random Slope Model Using KR

term	estimate	std.error	statistic	df	p.value
(Intercept)	11.644	2.881	4.041	25.71	0.000
alcohol1	3.461	1.666	2.077	9.64	0.066
real_age	0.394	0.209	1.885	16.52	0.077
wave	-0.060	0.862	-0.069	26.70	0.945

In this model, we add a random effect of time as well as random intercept. This means that we are assuming that for each individual, the relationship between time and BMI is unique. There are no significant effects when using the KR method. Variability in BMI across individuals is 16.388, and the residual variance is 2.352. The variance for time is 1.324, which represents variability across individual's BMI rates of change. We see that imposing variability between each individual's relationship of time vs BMI reduces some of the residual variance. When we accounted for some of the variation through each individual's weight changes over time, time as a predictor was no longer significant as it was in the random intercepts only model.





Similar to the random intercept model, there is homogeneity in variance of the residuals, but normality is not met. We will proceed to evaluation.

Table 3.6: Comparison of Fixed Effects P-Values in Random Slope Model

term	Satterthwaite_p.value	KR_p.value
(Intercept)	0.000213283766249562	0.000426895805163504
alcohol1	0.0459153731793183	0.0655425035327326
real_age	0.0572004159149207	0.0771903662035183
wave	0.942264459160284	0.945433676863754

Earlier evaluation of fixed effects uses KR DF method, but 3.6 shows that significance of fixed effects differ when using Satterthwaite versus KR. In the random slope model, alcohol not a significant effect when using KR, but is significant when using Satterthwaite. This difference affects our interpretation of the model and conclusions that are drawn.

3.2 Discussion

We have demonstrated implementing two linear mixed models on a data set that is small and with a ~~nonnormal~~ right-skewed continuous outcome. Imposing random effects structure to appropriately account for the correlation between repeated measures within each person

~~to isolate variation between individuals apart from overall variation improved the model and~~ reduced the number of predictors that were significant. One key result was that comparing KR and Sattherthwaite DF methods resulting in varying significant predictors in the random slope model. One additional predictor, alcohol, was a significant fixed effect when using Satterthwaite, the default output summary for *lmerTest*, in comparison to KR. Which method is preferable?

Bold package names

Looking at the results from our simulation study, we can identify which condition most closely resembles the children data, and determine if KR or Satterthwaite would produce more robust results. There was no difference in the fixed effects of the random intercept model, so we only focus on the random slopes model. The skewness and kurtosis values for the application data align most closely with that of the log-normal distribution $X \sim Log(0, .25)$. We will look at the performance of the two DF methods in this distribution, in a random slope model with a sample size of 10 and 4 measurements per individual. In reference to Figure 2.5, we see that the performance of KR and Satterthwaite are virtually the same. If this is the case, our preference will still be towards KR as it tends to produce slightly less anti-conservative Type I error rates.

Log-normal?

Ultimately, our application study supports the idea that linear mixed models can be applied to small samples when using KR or Satterthwaite DF methods. While performance between the two can be equally robust in theory, they can possibly lead to different ~~significant fixed effects and~~ conclusions ~~about data overall.~~ with the same fitted model.