Application

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.

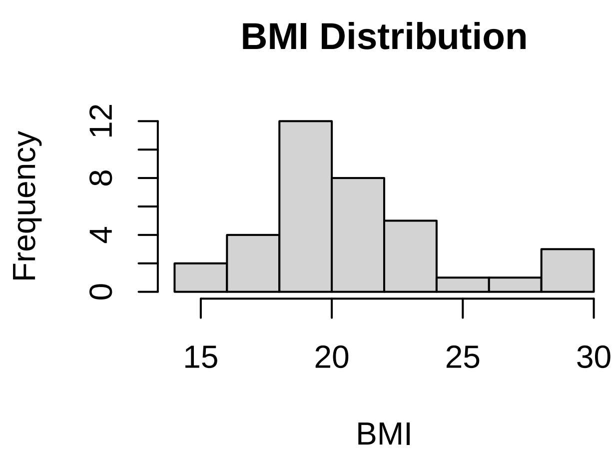
Background

The National Longitudinal Study of Adolescent to Adult Health is a longitudinal study spanning 1994 to 2008 that surveyed 15,000 U.S students in 7-12th grade in the 1994-95 school year [@harris_national_2022]. Four waves of data were collected in 1994, 1996, 2001, and 2008; 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 follow a right-skewed distribution [@penman_changing_2006]. We can see this in the distribution of children's BMI in figure @ref(fig:intercept). 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_youth_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.

Exploration



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. The skewness and kurtosis values for align most closely with that of the lognormal distribution $X \sim Lognormal(0, .25)$ used in the simulation study.

Figure @ref(fig:alcohol) 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, there are slight variations in pattern of change in BMI over time. At least two individuals in the sample have a pattern of BMI change that is noticeably different from the rest. Given that the sample size is small, it may not be necessary to examine implementing random effects for time, but it will be explored briefly.

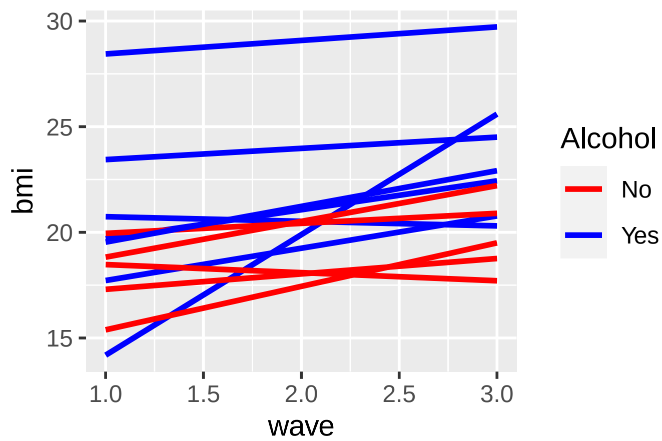


Figure 1: Changes in BMI by Individual and Alcohol Use

In a longitudinal study examining youth substance use and body composition, @pasch_youth_2012 found that baseline alcohol use is associated with lower values of BMI at follow-up. The opposite pattern can be seen in Figure @ref(fig:alcohol). Individuals who 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.

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.

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 1: Linear Model Output

term	estimate	std.error	statistic	p.value
(Intercept)	17.688	0.981	18.035	0.000
alcohol1	3.134	1.037	3.022	0.005
age_at_start_scaled	0.420	0.286	1.468	0.152
time	0.547	0.248	2.203	0.035

Table @ref(tab:lm) 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.

Intercept Only Model

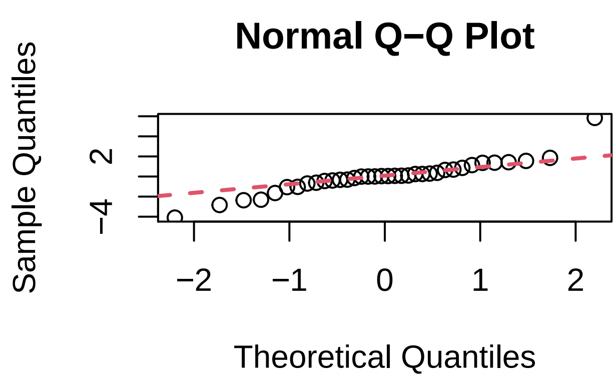
Table 2: Fixed Effects for Random Intercept Model Using KR

term	estimate	std.error	statistic	df	p.value
(Intercept)	17.688	1.313	13.475	10.547	0.000
alcohol1	3.134	1.653	1.897	9.000	0.090
age_at_start_scaled	0.420	0.456	0.921	9.000	0.381
time	0.547	0.157	3.491	23.000	0.002

Table 3: Random Effects for Random Intercept Model

group	term	estimate
AID	sd___(Intercept)	2.585
Residual	sd___Observation	1.931

Table @ref(tab:interceptKR) shows that time is the only significant effect in our model with only a random intercept. There is no significant relationship between early alcohol use and BMI. For individuals that have the same value for the random intercept, each year after the initial wave increases the predicted BMI by .5. In this model, an individual who is 16 years old at the initial wave and hasn't consumed alcohol has a predicted BMI of 17.7. Next, we turn to the random effects output in table @ref(tab:interceptr). The variance for individuals (represented by `AID`), which depicts variability across individuals, is 2.6, while the residual variance, representing within-subject variability is 1.9. The significantly larger variance across individuals compared to within individuals suggests that distinguishing different types of variation is useful. The intraclass correlation is .63, which indicates that weight measurements taken of the same individual have slightly higher similarity than those of different individuals.



When checking to see if the assumptions of the linear mixed model hold, we see that in the QQ plot the normality assumption of the residuals is not met. This is to be expected as the random effects generated are nonnormal, as well as the continuous outcome variable. We will proceed with evaluating the fixed effects in order to see how DF methods perform when normal assumptions do not hold.

Table 4: Comparison of Fixed Effects P-Values in Random Intercept Model

term	Satterthwaite_p.value	KR_p.value
(Intercept)	0.0000000553755002743052	0.0000000553754719963339
alcohol1	0.0903585631404512	0.090358561560752
age_at_start_scaled	0.380940525477476	0.380940524349758
time	0.0019689955685938	0.00196899558572261

The summary output referenced above uses Kenward-Roger DF approximation. Table @ref(tab:compareintercept) 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.

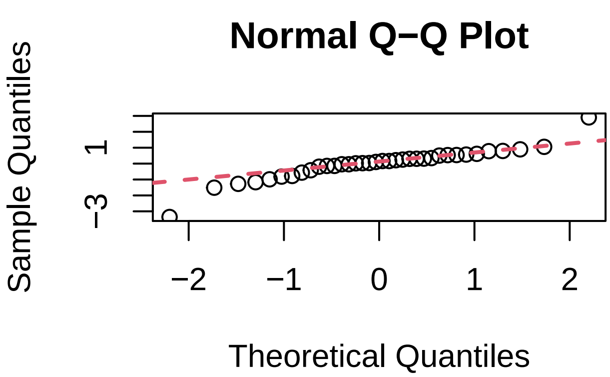
Intercept and Random Slope

While we explore the random slope model, it is important to recognize that this model may not be optimal given the sample size. Smaller sample sizes make parameter estimation more complicated; considering that the random slope model is more complex, we turn to the random intercept model as the most parsimonious.

Table 5: Fixed Effects for Random Slope Model Using KR

term	estimate	std.error	statistic	df	p.value
(Intercept)	17.500	1.448	12.083	11.485	0.000
alcohol1	3.458	1.756	1.969	9.000	0.080
age_at_start_scaled	0.346	0.485	0.714	9.000	0.493
time	0.547	0.193	2.835	11.000	0.016

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. Similar to the random intercept model, there is no significant relationship between early alcohol use and BMI. Variability in BMI across individuals is 11.1, and the residual variance is .3. The variance for time is 1.9, which represents variability across individual's BMI rates of change. We see that imposing variability between each individual's relationship of time vs BMI does not affect the residual variance.



Similar to the random intercept model, the normality condition is not met. We will proceed to evaluation.

Table 6: Comparison of Fixed Effects P-Values in Random Slope Model

term	estimate	std.error	statistic	df	p.value
(Intercept)	17.500	1.381	12.674	11.725	0.000
alcohol1	3.458	1.589	2.177	9.000	0.057
age_at_start_scaled	0.346	0.439	0.789	9.000	0.450
time	0.547	0.193	2.835	11.000	0.016

Earlier evaluation of fixed effects uses KR DF method, and @ref(tab:compareslope) shows significance of fixed effects when using the Satterthwaite. In the random slope model, alcohol not a significant effect when using KR, nor is it significant when using Satterthwaite. However, the p-value for the fixed effect of alcohol use is lower when using Satterthwaite DF compared to KR. While the lower p-value was not enough to create differences in significance of predictors between the two methods, it is enough to justify why a simulation study comparing KR and Satterthwaite is relevant. It is likely that in other models, a lower p-value produced by Satterthwaite method could cause a fixed effect to be interpreted as significant, while it is interpreted as not significant by KR. By using the Satterthwaite method, we could possibly falsely conclude that a predictor is significant when it isn't. This aligns with our simulation results that demonstrated that Satterthwaite is slightly more anti-conservative than KR.

Discussion

We have demonstrated implementing two linear mixed models on a data set that is small and with a non-normal continuous outcome. Imposing random effects structure to appropriately account for the correlation between repeated measures within each person improved the model and reduced the number of predictors that were significant. One key result was that comparing KR and Sattherthwaite DF methods resulted in varying p-values for fixed effects in the random slope model. Which DF method is preferable?

Looking at the results from our simulation study, we can identify which condition most closely resembles the children's health data, and determine if KR or Satterthwaite would produce more robust results. There was no signficant difference in the p-values of 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 lognormal 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 @ref(fig:fig5), 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.

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 conclusions with the same fitted model.