Does Gender and Parental Alchoholic Behavior Affect Teenagers' Beverage Consumption ?

Toan Pham

3/16/2021

Abstract

We investigate factors associated with alcohol use in teenagers. A multilevel model is used to model 246 observations by 82 different teenagers. We find the alcohol consumption is highest for male teenagers with alcoholic parent, and is higher for teenagers with friends who also drink . We also find that alcohol consumption tends to increase with age (from 14 to 16). Finally, we find that there is more variability in alcohol consumption over the years by the same teenager than between different teenagers at age 14.

Note:

- This dataset is taken from chapter 9 of **Beyond Multiple Linear Regression** by Roback and Legler.

Background and Significance

In this report, we investigate factors related to teenagers' alcohol consumption during the age of 14-16. A study by National Institute on Alcohol Abuse and Alcoholism (2006) found that being a child of an alcoholic or having several alcoholic family members places a person at greater risk for alcohol problems due to genetics. Children of alcoholics (COAs) are between 4 and 10 times more likely to become alcoholics themselves than are children who have no close relatives with alcoholism . The study also goes on to show that Environmental factors, such as the influence of parents and peers, also play a role in alcohol use. For example, parents who drink more and who view drinking favorably may have children who drink more, and an adolescent girl with an older or adult boyfriend is more likely to use alcohol and other drugs and to engage in delinquent behaviors. In a blog written by Talk It Out Nc(2019), the author shares the same idea: Peer pressure is one of the biggest reasons why teens choose to drink alcohol.

We are interested in exploring certain background experiences are associated with alcohol consumption in teenagers. We also explore whether the age of the teenagers play a role in alcohol consumption. Developing a better understanding of the factors associated with beverage consumption might help parents, instructors, and the teenagers themselves manage the amount of alcohol they consume to a healthier level.

Data

We use data collected by Curran, Stice, and Chassin (1997) on 82 adolescents at three time points starting at age 14. The primary response in the dataset is 'alcuse'. Alcuse is measured by the square root of the sum of four items: (a) drank beer or wine, (b) drank hard liquor, (c) 5 or more drinks in a row, and (d) got drunk—were each scored on an 8-point scale, from 0="not at all" to 7="every day". The data are discussed in **Beyond Multiple Linear Regression** (Roback and Legler, 2021), and are available on Github. (https://github.com /proback/BeyondMLR/blob/master/data/alcohol.csv (https://github.com/proback/BeyondMLR/blob/master /data/alcohol.csv)).

This is a longitudinal multilevel dataset, since it records data of teenagers' alcohol consumption over the time period of 3 years(age 14 to 16). There are variables pertaining to the teenagers, such as gender, whether they have an alcoholic parent, and peer alcohol use. We refer to these as level one variables. There is also a variable that measure the age of the teenagers over the time period. For the sake of clarity in our model, I changed the variable "age" to "age14", to measure year passed since age 14.

Select variables, for the first 9 rows of the dataset are shown below. These pertain to 9 observations by three teenagers.

```
##     X id age coa male      peer   alcuse age14
## 1   1  1  14   1    0 1.2649111 1.732051     0
## 2   2  1  15   1    0 1.2649111 2.000000     1
## 3   3  1  16   1    0 1.2649111 2.000000     2
## 4   4  2  14   1    1 0.8944272 0.000000     0
## 5   5  2  15   1    1 0.8944272 0.000000     1
## 6   6  2  16   1    1 0.8944272 1.000000     2
## 7   7  3  14   1    1 0.8944272 1.000000     0
## 8   8  3  15   1    1 0.8944272 2.000000     1
## 9   9  3  16   1    1 0.8944272 3.316625     2
## 10 10  4  14   1    1 1.7888544 0.000000     0
```

The alcuse score (`alcuse`) variable measures alcohol consumption, with higher values indicating higher

amount of alcohol consumed . Figure 1 displays histograms of average alcuse scores across 3 years for all 82 teenagers.

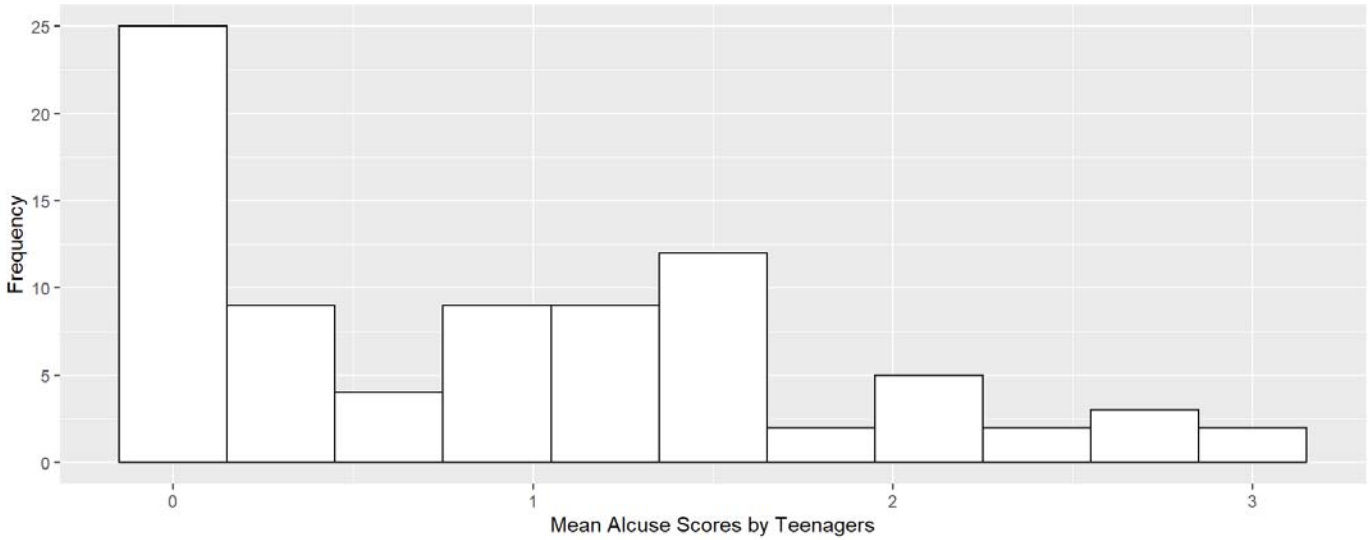

Figure 1

We see that distribution of mean alcuse scores is right-skewed, with scores most frequently lying around 0. This means that the majority of the teenagers in our data actually do not consume any alcohol. For teenagers who do drink, alcuse scores frequently lying bnear 1, with a few outliers near the 3 mark.

Next, we consider level one covariates gender(1 being Male), alcoholic parent(1 being have alcoholic parent), and a measure of peer alcohol use. Figure 2,3, and 4 displays relationships between mean alcuse scores and these variables.

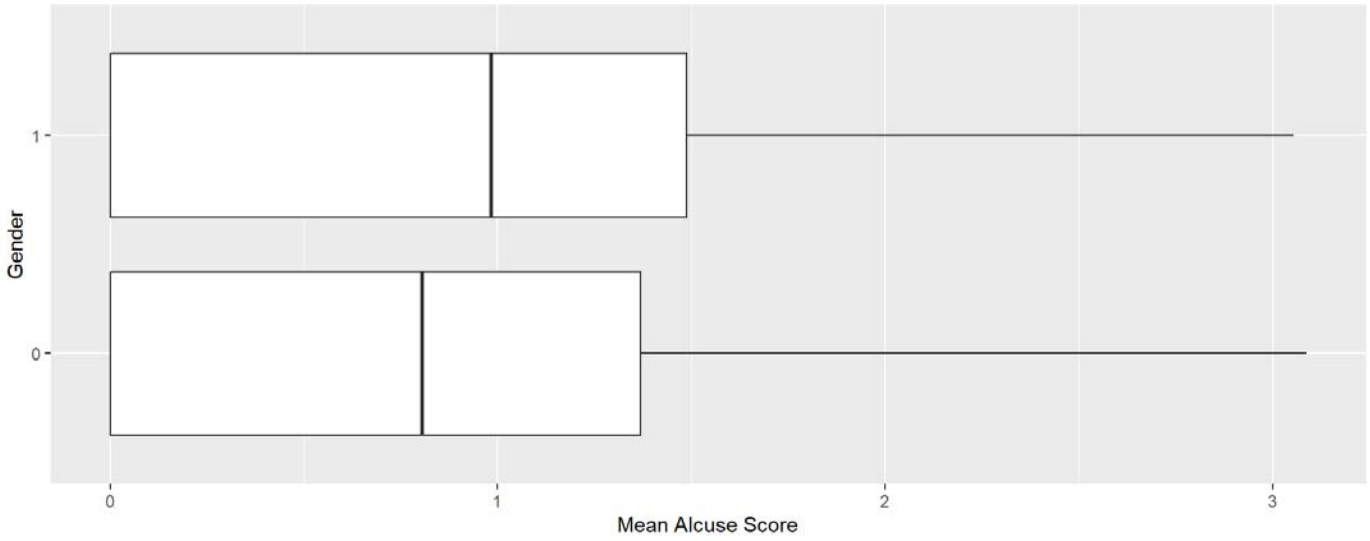

Figure 2

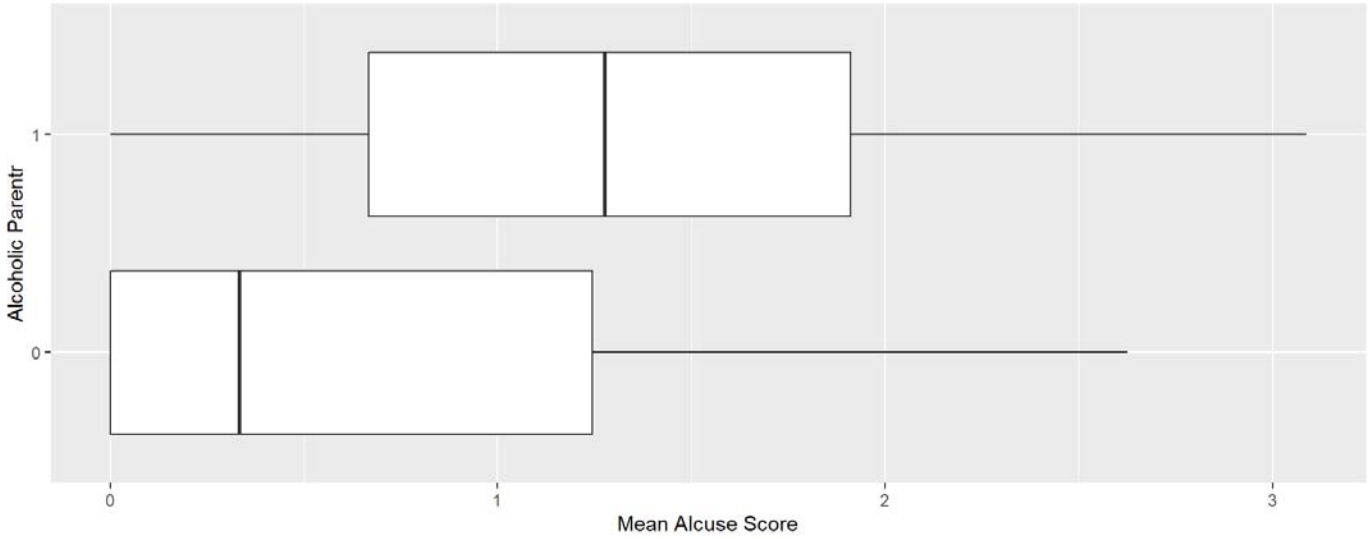

Figure 3

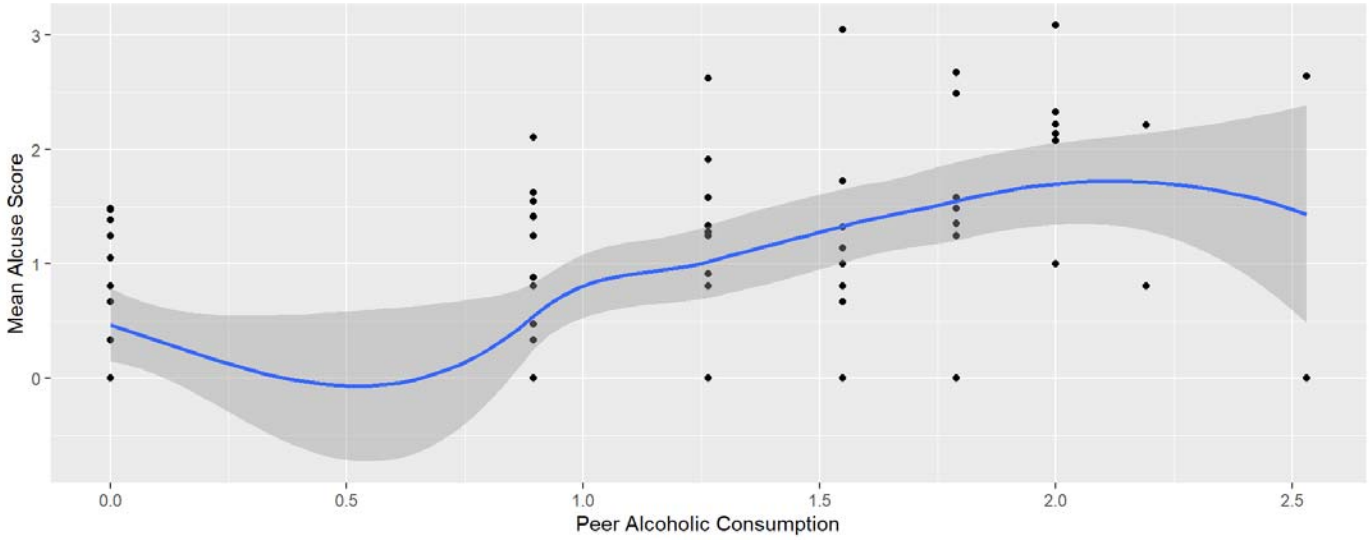

Figure 4

We see that acuse score appears to be slightly higher on average for male teenagers. We also see that alcohol consumption is much when the teenager has an alcoholic parent. There is also some evidence of positive correlation between average alcuse score and peer alcoholic consumption.

To assess the relationship between age and alcoholic consumption, we create individual lattice plots for all 82 teenagers over 3 years, as seen in Figure 5.

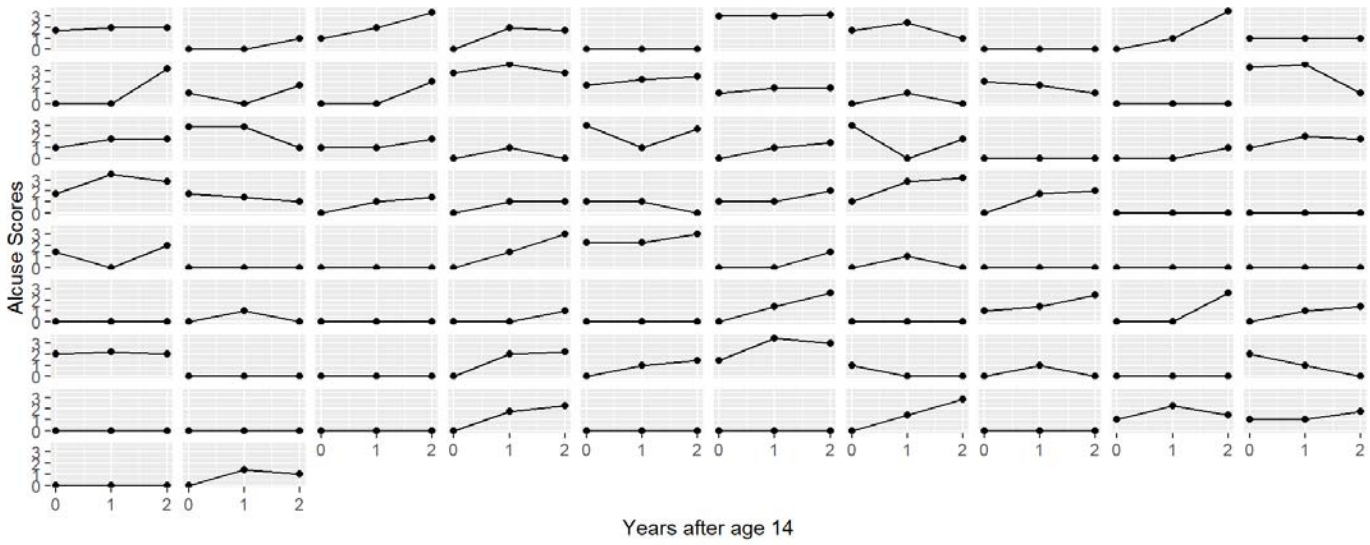

Figure 5

We see that for most teenagers who do not drink at age 14 continue to not drink over the time period.

For the teenagers who do drink, we see that the majority of the trendlines are upward sloping, indicating that alcholic consumption tends to increase with age, but this is not true of all teenagers. Some teenagers actually manage to bring down the alcuse score to 0.

Next, we look at the Latticed Spaghetti plots that illustrate the relationship between alcuse scores and gender/alcholic parent. We are interested in seeing how this relationship change over time.

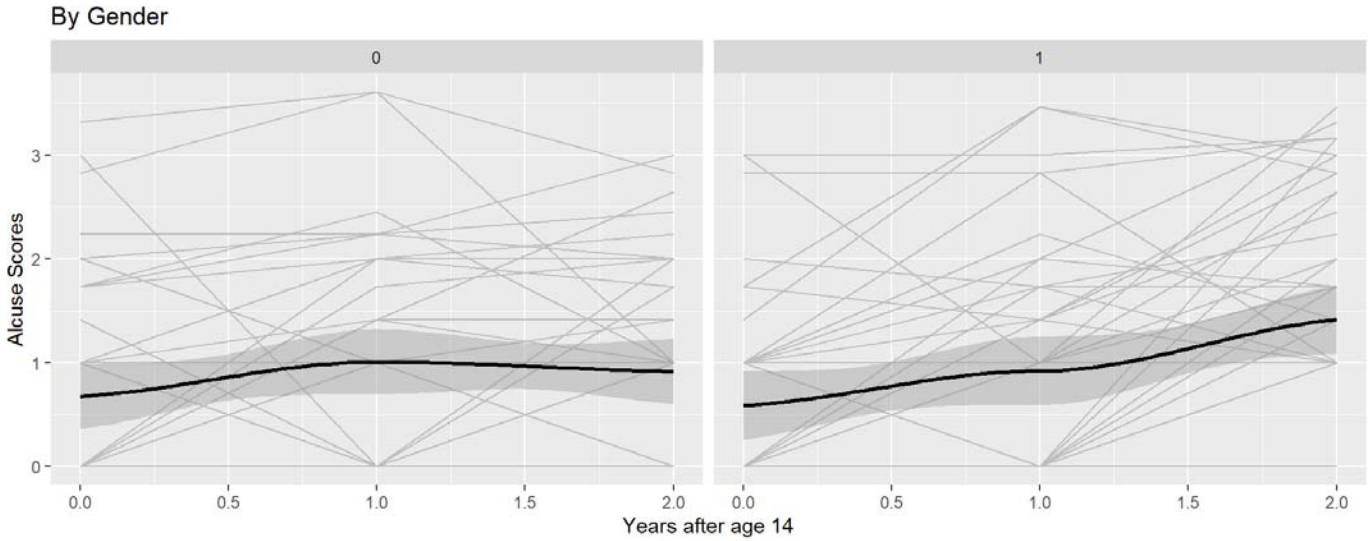

Figure 6

We see that female teenagers keep their alcohol consumption constant over 3 years while male teenagers' consumption is higher over the same time period. This is quite noticeable and should warrant an interaction term between the two in our model.

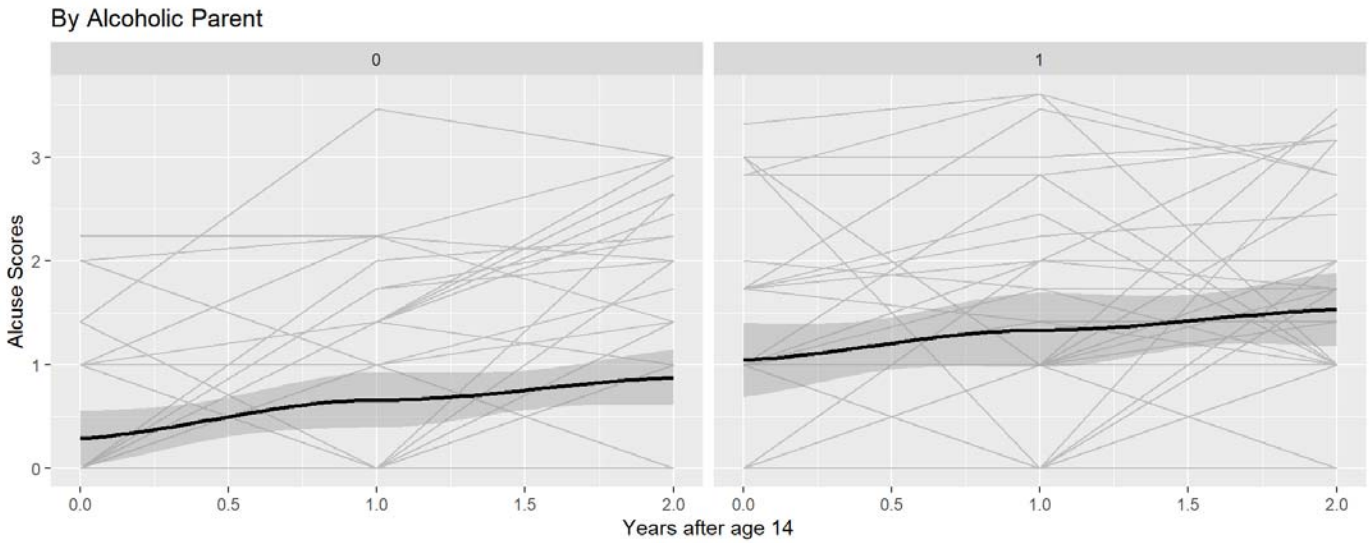

Figure 7

The slopes between having alcoholic parent and not are quite similar, with a slight increase over the time period.

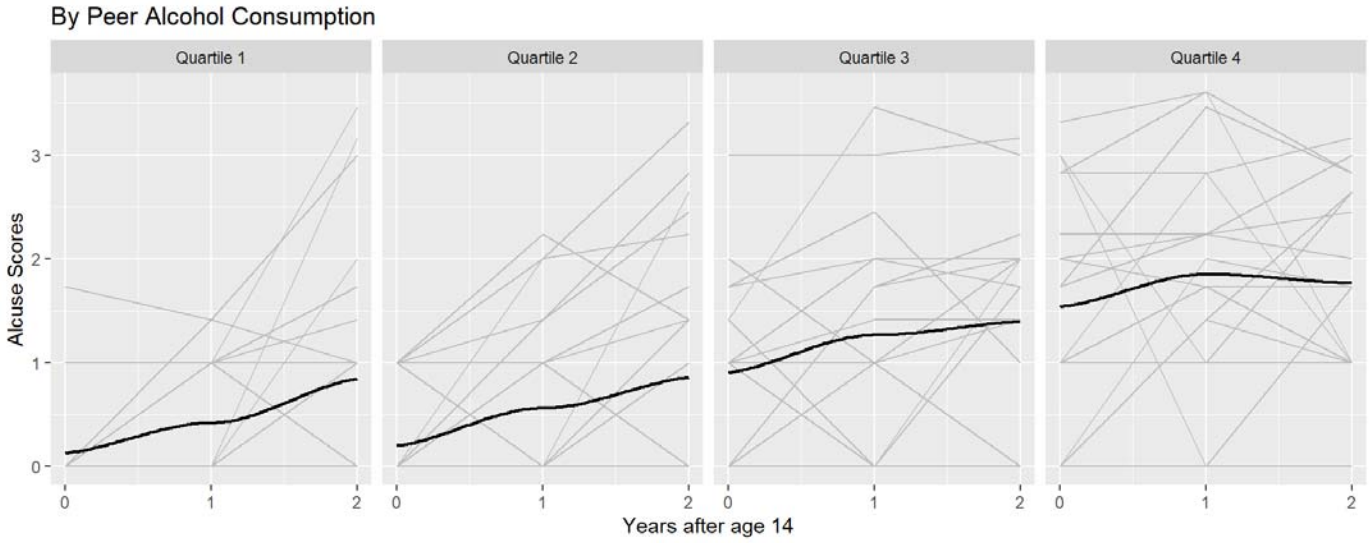

Figure 8

We see that for the first quartile (low peer alcohol consumption), the rate of change in alcohol consumption is much higher than the rest. We have a postive rate of change in the first 3 quartiles, however, in quartile 4, we have a somewhat quadratic relationship between peer alcohol consumption and alcuse score.

Overall, we see that gender, age, alcoholic parent, and peer alcohol consumption all play a role in teenagers' drinking. We will need to consider all these variables, as well as some interaction for our model.

Methods

We now consider three different multilevel models for alcuse scores. Let Y_{ij} be the alcuse score of teenager i at year j after age 14. We begin by fitting a model that accounts for all the explanatory variables without considering their interaction. The model includes a random intercept term, allowing alcuse scores to vary randomly between teenagers, and a random term on the age indicator variable, allowing the yearly rate of change of alcuse scores to also varies between teenagers. Mathematically, we write the model as:

Model 1:

$$
\begin{aligned}
Y_{ij} &= [\alpha_0 + \alpha_1 \text{Male}_i + \beta_0 \text{Age14}_{ij} + \alpha_2 \text{Coa}_i + \alpha_3 \text{Peer}_i] \\
&\quad + [u_i + v_i \text{Age14}_{ij} + \epsilon_{ij}]
\end{aligned}
$$

where,

$$
\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \right)
$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Alternatively, we consider a simplified version of the model that does not include the random effect v_i. This model assumes that alcuse scores vary randomly between teenagers, but the yearly rate of change of alcuse scores does not varies between teenagers. This model is written as

$$
\begin{aligned}
Y_{ij} &= [\alpha_0 + \alpha_1 \text{Male}_i + \beta_0 \text{Age14}_{ij} + \alpha_2 \text{Coa}_i + \alpha_3 \text{Peer}_i] \\
&\quad + [u_i + \epsilon_{ij}]
\end{aligned}
$$

where $u_i \sim N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

We use AIC and BIC to assess the fit of each model.

```
##        df      AIC
## model1  9 622.5627
## model2  7 635.3712
```

```
##        df      BIC
## model1  9 654.1107
## model2  7 659.9085
```

Model 1 achieves lower AIC (622) and BIC (654) scores than Model 2 (AIC=635, BIC=659), suggesting that the random effect of the yearly rate of change of alcuse scores is needed. Thus, we proceed with both random intercept term and random slope.

We now consider adding the interaction that might help explain alcohol consumption in teenagers. In figure 6, we saw that the rate of change in alcuse score between male and female teenagers is noticeably different. The model is:

Model 3:

$$
\begin{aligned}
Y_{ij} &= [\alpha_0 + \alpha_1 \text{Male}_i + \beta_0 \text{Age14}_{ij} + \alpha_2 \text{Coa}_i + \alpha_3 \text{Peer}_i + \beta_1 \text{Male}_i \text{Age14}_{ij}] \\
&\quad + [u_i + v_i \text{Age14}_{ij} + \epsilon_{ij}]
\end{aligned}
$$

where $u_i \sim N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

```
##        df      AIC
## model1  9 622.5627
## model3 10 621.3675
```

```
##        df      BIC
## model1  9 654.1107
## model3 10 656.4208
```

```
## Data: alcohol
## Models:
## model1: alcuse ~ age14 + peer + factor(male) + factor(coa) + (age14 |
## model1:     id)
## model3: alcuse ~ age14 + peer + factor(male) + factor(coa) + age14:factor(male) +
## model3:     (age14 | id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model1    9 608.69 640.24 -295.35   590.69
## model3   10 605.06 640.11 -292.53   585.06 5.6361  1    0.01759 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 3 acheives an AIC value of 621, an improvement over Model 1. On the other hand, the BIC value for Model 3 is 656, which is slightly worse than Model 1. Since Model 1 is a nested submodel of Model 3, we can compare the models with a likelihood ratio test. The null hypothesis is that model 1 adequately fits the data. This test yields a χ^2 test statistic of 5.63 on 1 degrees of freedom, resulting in a low p-value, suggesting that the larger Model 3 is preferred.

Results

We fit Model 3 to the data. Below is the summary of model 3 output.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## alcuse ~ age14 + peer + factor(male) + factor(coa) + age14:factor(male) +
##     (age14 | id)
##    Data: alcohol
##
## REML criterion at convergence: 601.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.57952 -0.42682 -0.09154  0.39138  2.29283
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.2674   0.5171
##           age14       0.1374   0.3707   -0.08
##  Residual             0.3373   0.5808
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)         -0.23496    0.17409 86.66502  -1.350 0.180643
## age14                0.12129    0.08747 79.99976   1.387 0.169411
## peer                 0.63057    0.10502 78.00025   6.004 5.74e-08 ***
## factor(male)1       -0.01234    0.16743 79.50948  -0.074 0.941432
## factor(coa)1         0.55615    0.14949 78.00025   3.720 0.000373 ***
## age14:factor(male)1  0.29161    0.12223 79.99976   2.386 0.019405 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) age14  peer   fctr(m)1 fctr(c)1
## age14       -0.302
## peer        -0.647  0.000
## factor(ml)1 -0.588  0.314  0.207
## factor(co)1 -0.238  0.000 -0.178 -0.081
## ag14:fct()1  0.216 -0.716  0.000 -0.438    0.000
```

We see that $\hat{\sigma}_u^2 = 0.267$ and $\hat{\sigma}^2 = 0.3373$. The fact that $\hat{\sigma}^2$ is approximately larger than $\hat{\sigma}_u^2$ suggests that there is somewhat more variability in alcuse score between different year by the same teenager, than in average alcuse scores for different teenagers. The variance between teenagers in rate of changes in alcuse scores during the observation period = 0.1374 is quite a bit lower than the other two.

The next table shows the estimates of fixed effects and their associated standard errors, t-statistics, and p-values.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-0.2349637	0.1740923	86.66502	-1.3496498	0.1806433

	Estimate	Std. Error	df	t value	Pr(>|t|)
age14	0.1212927	0.0874740	79.99976	1.3866153	0.1694111
peer	0.6305678	0.1050229	78.00025	6.0040973	0.0000001
factor(male)1	-0.0123398	0.1674259	79.50948	-0.0737032	0.9414316
factor(coa)1	0.5561534	0.1494858	78.00025	3.7204438	0.0003732
age14:factor(male)1	0.2916050	0.1222253	79.99976	2.3857992	0.0194055

We see that the coefficient for age14, peer, and coa are all positive, implying a positive correlation with our response variable. In other words, as age increases, teenagers are expected to drink more alcohol. Male teens and teens with alcoholic parents are also expected to have higher alcohol consumption than their counterpart. While male is a negative coefficient(-0.012), it has a high p-value(0.94) thus we can safely ignore it. The interation term age14:male also shows us how male teenagers are expected to have higher rate of change in alcohol consumption over 3 years compared to female teenagers.

Discussion and Conclusions

Our results show that on average, male teenagers who are a child of an alcoholic parent experience higher levels of alcohol consumtion than their counterparts. This supports the findings of National Institute on Alcohol Abuse and Alcoholism (2006). There is evidence that having friends who also drink increases alcohol consumption, which is simlar to what Talk It Out NC claims in their article written in 2019. Other than that, there are also some signs older teenagers having higher alcohol consumption, especially among male teenagers. Finally, we see that there is somewhat more varibility in alcuse score between different year by the same teenager, than in average alcuse scores for different teenagers.

Our results are based on a sample of 82 teenagers, so while they provide insight, we should be careful to not generalize them to all teenagers in the US. Future research might further explore the potential link between teenagers' backgrounds and alcoholic consumption, and investigate psychological reasons for this connection. We may also need more data to find out whether having a alcoholic parents is a genetic influence or a envinronmental one. Using this study, teachers and parents may understand better why their teenagers choose consume alcohol and create a preventation plan to mitigate any health risks associated with underage drinking.

References (5 points)

Roback, P. (2021). Beyond multiple linear regression: Applied generalized linear models and multilevel models in R. Boca Raton, FL: CRC Press.

Underage drinking. (n.d.). Retrieved March 16, 2021, from https://pubs.niaaa.nih.gov/publications /AA67/AA67.htm (https://pubs.niaaa.nih.gov/publications/AA67/AA67.htm)

5 causes of TEEN DRINKING: Why do Kids DRINK? (2019, November 18). Retrieved March 17, 2021, from https://www.talkitoutnc.org/underage-drinking/causes-teen-drinking/ (https://www.talkitoutnc.org/underage-drinking/causes-teen-drinking/)

Appendix (optional)

Include additional figures desired. You should have brief explanations of each figure, but you do not need to go into as much depth as you do in the actual report. Be sure to include your essential findings in the report. In reality, an appendix may or may not be read, so it will have little impact on your grade, but it is an opportunity to show interesting findings that do not fit in the report itself.