Alcohol Consumption in Adolescents

Bobby Soule, Daniel Chen & Tangni Chen

The Data

Background

- The NIH states that underage drinking is a leading public health problem
- Both nationwide surveys and local studies show adolescent drinking continues to be widespread
- About 5,000 deaths per year from underage drinking
  - 1,900 deaths from motor vehicle crashes
  - 1,600 as a result of homicide
  - 300 suicides
  - Hundreds more from falls, burns, drownings, etc.

Background Cont.

- Research shows that many adolescents start drinking at very young ages
- In 2003, the average age of first use of alcohol was about 14
    - In 1965, this statistic was about 17.5
- Those who reported drinking before the age of 15 were four times more likely to report meeting the criteria for alcohol dependence at some point in life

Goal of Analysis

We seek to examine the association between school relationships (student-student and teacher-student) and alcohol consumption behavior

Motivation

- At risk adolescents are more likely to attend schools with poorer quality environments and less resources
- Adolescents spend considerable amounts of time in schools
- Interested in seeing if schools can play a unique role in preventing risky alcohol consumption behavior
- Interested in school metrics that can make a difference; looked at:
    - Teacher-student relationships
    - Student-student relationships

ADD Health

- The National Longitudinal Study of Adolescent Health (Add Health)is a longitudinal study of a nationally representative sample of adolescents in grades 7- 12 in the United States during the 1994-95 school year
- Wave I: focus on the forces that may influence adolescents' health and risk behaviors
    - N = 6504
    - **School factors**
    - Personal traits, families, friendships, romantic relationships, peer groups, neighborhoods, and communities

Outcome of Interest (Response)

- Alcohol use:
  - 2 level nominal variable:
  - Non-drinker and light drinker vs. risky drinker

Alcohol Use (Non-Drinker)

Question: Have you ever had a drink of beer, wine, liquor - not just a sip or taste of someone else's drink -- more than 2 or 3 times in your life?

Answer:

- If NO; classified as a 'non-drinker'
- If YES; asked follow up questions

Alcohol Use (Drinker, Risky Drinker)

Classified as a “risky drinker” if:

- Males:
  - 5+ drinks per drinking occasion
  - 14+ drinks per week
- Females:
  - 4+ drinks per drinking occasion
  - 7+ drinks per week

All others were classified as “light drinkers”

Predictors

Main Predictors:

- Score of student-teacher relationship
- Score of student-student relationship

Additional Predictors:

- Race
- Age
- Sex
- Parent’s education
- Urbanicity
- Income

VARIABLE OPERATIONALIZATIONS

RELATIONSHIPS AT SCHOOL STUDENT-TEACHER AND STUDENT-STUDENT RELATIONSHIP

- How often did you have trouble getting along with your teachers
- How often did you have trouble getting along with other students 5 level Likert scale 0 - 4 (never → everyday)

COVARIATES

RACE (4): WHITE, BLACK, HISPANIC/LATINO, ASIAN

AGE (3): 10-14, 15-17, 18+

SEX (2): MALE, FEMALE

PARENT'S EDUCATION (3): HS-, HS, HS+ (High School)

URBANICITY (3): RURAL, SUB-URBAN, URBAN

INCOME (4): 0-20K, 20-35K, 35-70K, 70K+

Reference variable is the first one listed

Logistic Regression

Logistic Regression (in Bayesian Paradigm)

In logistic regression, our response variable can only take on values of zero or one (cannot range continuously from negative to positive infinity)

So, we must use a different model

Classical Logistic Regression Cont.

The model can range from negative to positive infinity,but the expected value of the yi's can only range from zero to one

We must apply a "link" function to the expected value of the yi's:

$$g\left(E\left[y_i\right]\right) = \beta_0 + \beta_1 x_{1i}$$

which now gives us $g\left(E\left[y_i\right]\right) \in (-\infty, \infty)$

In logistic regression, the "link" function g(x) is called the logit function:

$$g\left(E\left[y_i\right]\right) = logit\left(p_i\right) = log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i}$$

Classical Logistic Regression Cont.

Next, we must solve for pi:

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i}$$

$$\Rightarrow \frac{p_i}{1-p_i} = exp(\beta_0 + \beta_1 x_{1i})$$

$$\Rightarrow p_i = exp(\beta_0 + \beta_1 x_{1i}) - p_i \cdot exp(\beta_0 + \beta_1 x_{1i})$$

$$\Rightarrow p_i \cdot (1 + exp(\beta_0 + \beta_1 x_{1i})) = exp(\beta_0 + \beta_1 x_{1i})$$

$$\Rightarrow p_i = \frac{exp(\beta_0+\beta_1 x_{1i})}{1 + exp(\beta_0+\beta_1 x_{1i})} = \frac{1}{exp(-(\beta_0+\beta_1 x_{1i})) + 1}$$

Logistic Regression (Bayesian Paradigm)

In the Bayesian paradigm, we have the following logistic regression model:

$$y_i \mid p_i \sim Bernoulli\,(p_i) = Bernoulli\left(\frac{1}{exp(-(\beta_0+\beta_1 x_{1i})) + 1}\right)$$

Some common priors for logistic regression are:

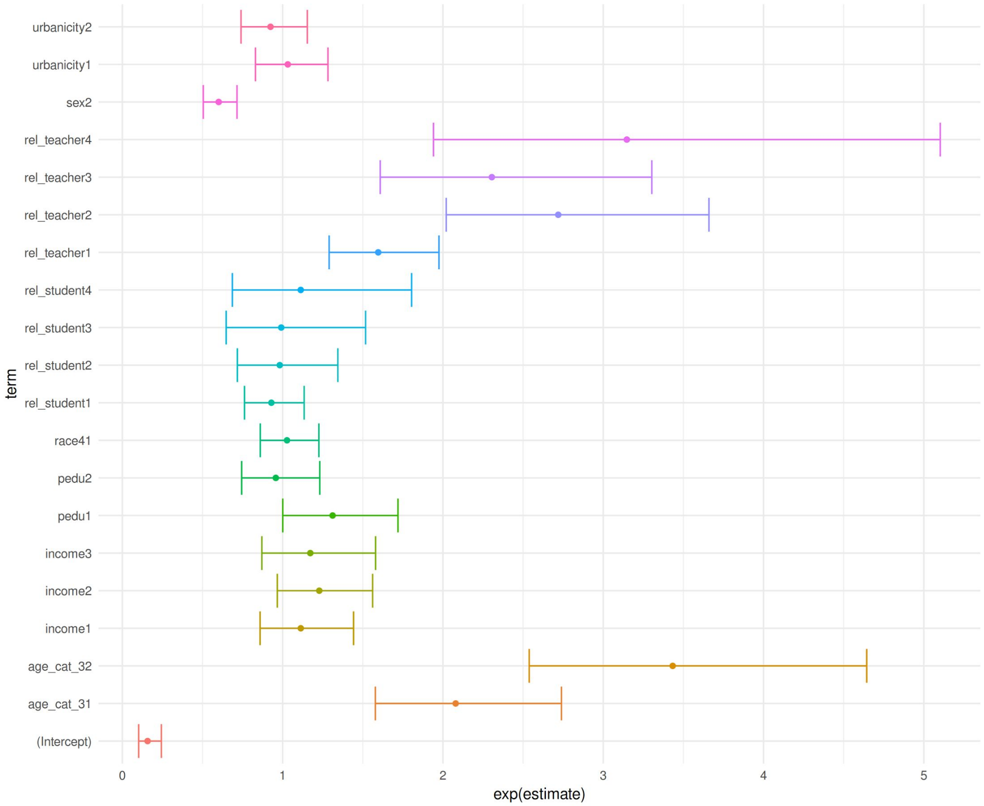
$$\beta_0 \sim N\left(0,\ \varphi^{-1}\right)$$

$$\beta_1 \sim N\left(0,\ \varphi^{-1}\right)$$

$$\varphi \sim Gamma\,(a,\ b)$$

Analysis

Odds Ratio Plots (Classical Log. Reg.)



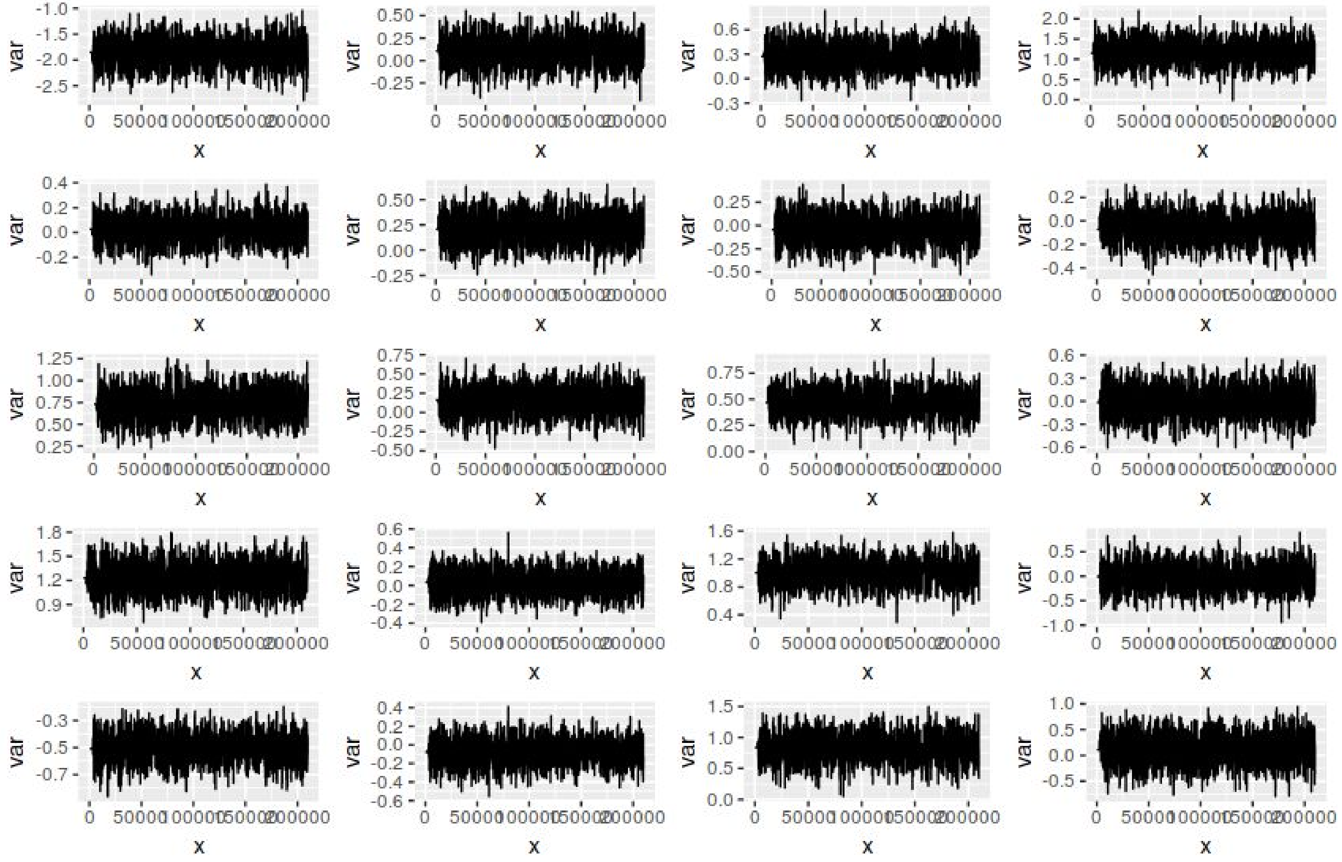
MCMClogit Function

Used the MCMClogit function from the MCMCpack package:

```
MCMClogit(formula, data=NULL, burnin = 1000, mcmc = 10000, thin=1, tune=1.1, verbose = 0,
    seed = NA,  beta.start = NA, b0 = 0, B0 = 0, user.prior.density=NULL, logfun=TRUE,
    marginal.likelihood=c("none", "Laplace"), ...)
```

- Generates sample from posterior distribution of logistic regression model
- Uses a random walk Metropolis algorithm (not a Gibbs sampler)
- By default, assumes all beta priors to be multivariate normal
- Used a burn-in of 1000 and 210,000 draws
- Intercept plus 19 variables (8 categorical variables plus response)

Trace Plots



var
-1.0 -1.5 -2.0 -2.5
0.50 0.25 0.00 -0.25
0.6 0.3 0.0 -0.3
2.0 1.5 1.0 0.5 0.0
0.4 0.2 0.0 -0.2
0.50 0.25 0.00 -0.25
0.25 0.00 -0.25 -0.50
0.2 0.0 -0.2 -0.4
1.25 1.00 0.75 0.50 0.25
0.75 0.50 0.25 0.00 -0.25 -0.50
0.75 0.50 0.25 0.00
0.6 0.3 0.0 -0.3 -0.6
1.8 1.5 1.2 0.9
0.6 0.4 0.2 0.0 -0.2 -0.4
1.6 1.2 0.8 0.4
0.5 0.0 -0.5 -1.0
-0.3 -0.5 -0.7
0.4 0.2 0.0 -0.2 -0.4 -0.6
1.5 1.0 0.5 0.0
1.0 0.5 0.0 -0.5
0 50000 100000 150000 200000
x

MCMC Summary

	variable	mean	sd	ci_lower	ci_upper
1	(Intercept)	-1.85771870	0.21358367	-2.280752665	-1.4221874
2	age_cat_31	0.73673968	0.14016571	0.455124692	1.0110882
3	age_cat_32	1.24440697	0.15452089	0.944743190	1.5481549
4	income1	0.10457602	0.12840397	-0.152488370	0.3516683
5	income2	0.20604745	0.11774313	-0.027596127	0.4370216
6	income3	0.15582943	0.14917422	-0.137812201	0.4360424
7	pedu1	0.27350734	0.13455988	0.006180259	0.5325361
8	pedu2	-0.04022461	0.12652968	-0.291405046	0.2095747
9	race41	0.02252508	0.08891678	-0.153875231	0.1965895
10	rel_student1	-0.06702976	0.09955636	-0.264161479	0.1339143
11	rel_student2	-0.01042233	0.15872697	-0.321808558	0.2996826
12	rel_student3	-0.01461795	0.21502412	-0.444683248	0.4057518
13	rel_student4	0.11621265	0.25162256	-0.390571001	0.6045833
14	rel_teacher1	0.46063324	0.10663646	0.252750763	0.6687060
15	rel_teacher2	0.99739939	0.15056273	0.706760722	1.2862407
16	rel_teacher3	0.82038171	0.18351097	0.461517252	1.1719004
17	rel_teacher4	1.16174484	0.24639553	0.671359888	1.6457717
18	sex2	-0.50820484	0.08883851	-0.689104921	-0.3340304
19	urbanicity1	0.03061319	0.10887015	-0.185085215	0.2442152
20	urbanicity2	-0.07970572	0.11531155	-0.307089588	0.1405449

Significant Parameters

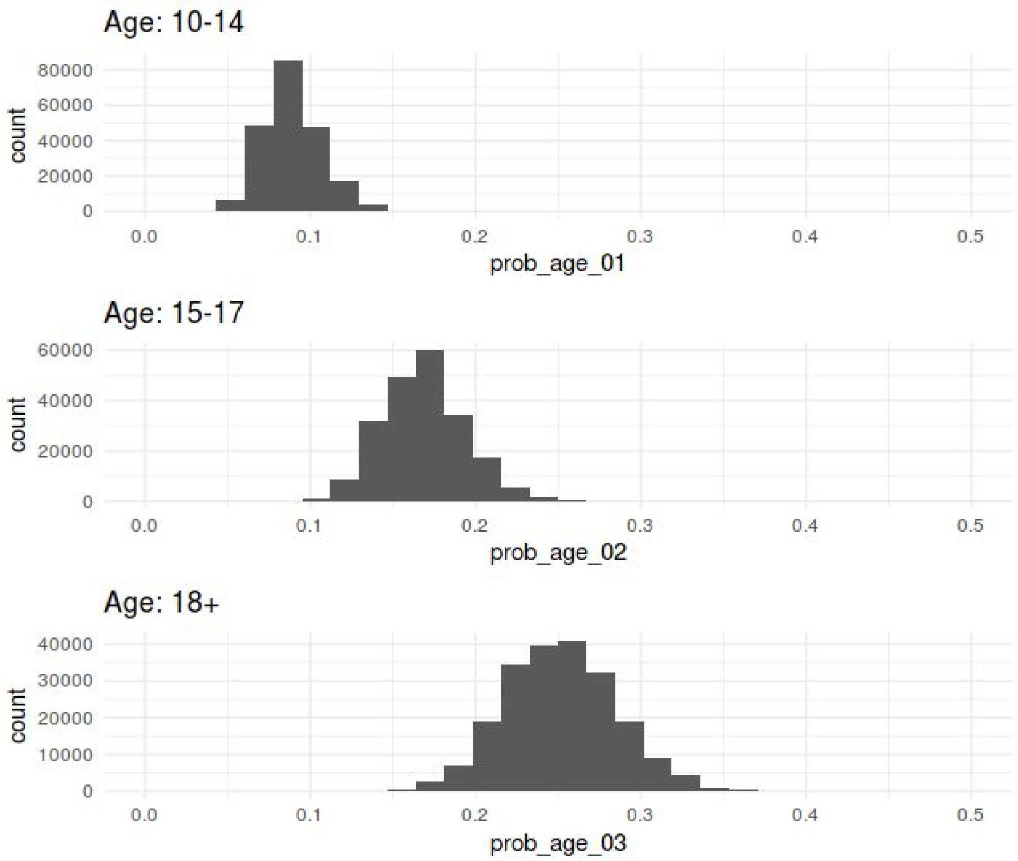
- Intercept
- Age Category (two dummy variables)
  - Older age groups more likely to be risky drinker
- Parent’s Education (only one of the dummy variables)
- Student Teacher Relationship (four dummy variables)
  - Interestingly, having a good relationship with teacher increased odds of being risky drinker
- Sex
  - Females less likely to be risky drinker

Probability of Risky Drinker by Age Group

This shows how the distribution of the pi's change for the three different age groups

Used reference values for other variables:
white males; aged 10-14 years old; parents education less than high school, living in rural area, income $0-20K

Interpretation:
For 10-14 year olds, on average, the probability of being a risky drinker is <0.1



Discussion

Limitations

- Missing Values
- Co-linearity?
    - School Curriculum

Further Analysis

- Put in continuous variables
- Sweep across all possible combinations of the model

Thanks
Applause

https://github.com/chendaniely/stat_5444g-bayesian_project