Alcohol Consumption in Schools

Hypothesis Heroes - Drew Davison, Lisa Zhang, Ellie Culman, Austin Chang

2023-11-14

Introduction and data

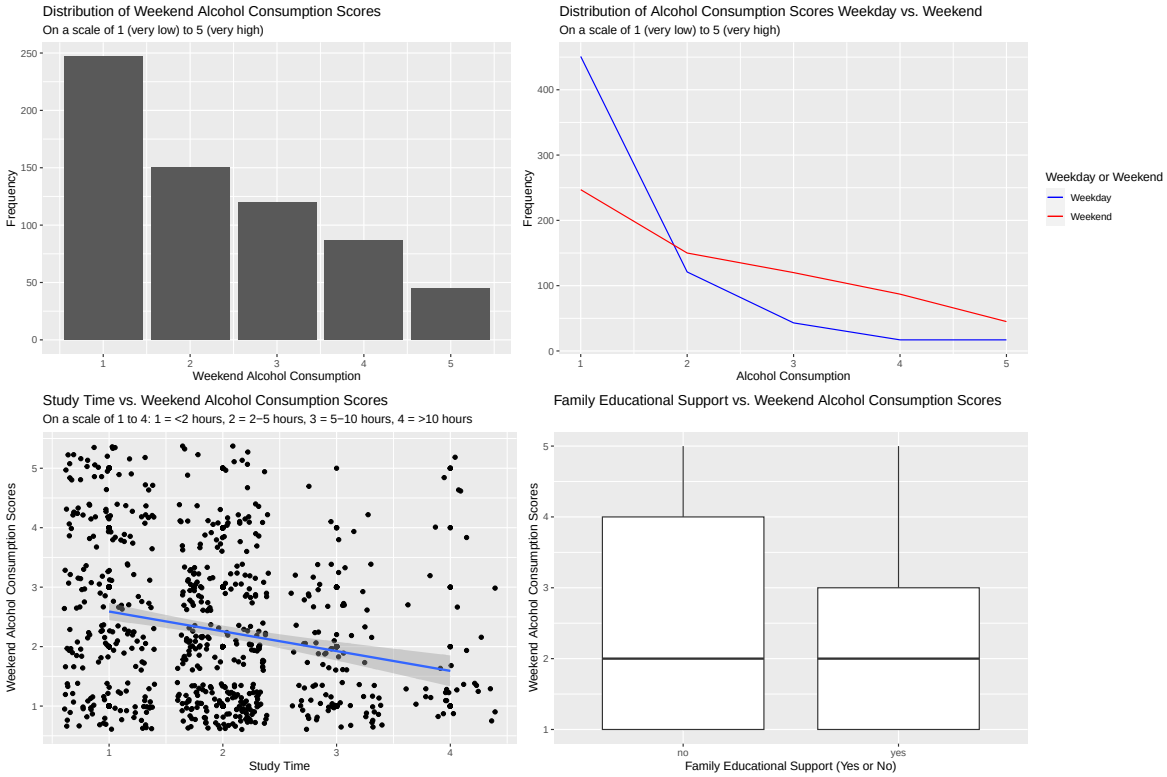
All around the world, underage drinking among students is a significant public health issue and takes a huge toll on the quality of students' lives and education. There is plenty of existing documentation on the effects of socioeconomic status on risky drinking behavior amongst college students. For example, a paper by Susan E. Collins, PhD, professor and licensed clinical psychologist, found that individuals with lower socioeconomic status as well as people of racial and ethnic minorities and homelessness experience greater alcohol-related consequences. Additionally, studies from the National Institutes of Health have found that "aspects of college life—such as unstructured time, widespread availability of alcohol, inconsistent enforcement of underage drinking laws, and limited interactions with parents and other adults" drives up rates of underage drinking, and "college students have higher binge-drinking rates and a higher incidence of driving under the influence of alcohol than their noncollege peers." Since most existing literature on underage student drinking focuses on college students, we wanted to examine the factors contributing to drinking amongst secondary school students. Our research question is as follows:

How do social indicators affect student alcohol consumption in secondary schools?

We hypothesize that factors like gender, familial status, family and school support, and other social and economic indicators will strongly influence the rates that secondary school students consume alcohol and that increased alcohol consumption is correlated with their school performance.

This data is from a Kaggle public dataset, originally sourced from the UC Irvine Machine Learning Repository. The data consists of information collected in 2008 on secondary school students from two schools in Portugal: Gabriel Pereira and Mousinho da Silveira. There are 649 observations, each one being a student, and 33 variables which cover a range of characteristics about each student's family, education, social situation, alcohol consumption, and grades in their Portuguese language class. Some key variables we will be examining are sex, male or female, Pstatus, whether their parents are together or apart, schoolsup, whether or not their school provides them extra academic support, famsup, whether or not their family provides

them extra academic support, Mjob and Fjob, the categories that their parents' jobs fall under, freetime and studytime, the amount of free time after school and amount of time spent studying on a scale of 1-5, number of absences, and amount of class failures, on a scale of 1-4. Our response variable is Walc, which is the students' average weekend alcohol consumption, on a scale of 1-5, with 5 being very high.



Methodology

We are using a multivariable linear regression to examine the effects of the predictor variables sex, Pstatus, schoolsup, famsup, Mjob, Fjob, freetime, studytime, absences, and failures on the response variable Walc, or weekend alcohol consumption. We are using a linear model rather than logistic since Walc is not binary. In our model we are also including interactions between freetime and studytime, failures and studytime, failures and absences, as well as famsup and Mjob and famsup and Fjob. In our model we used step_dummy() to create dummy variables for all categorical variables, step_interact() to create all of our interaction variables, and step_zv() to remove all variables with only one value.

Results

```
Linear Regression Model Specification (regression)

Computational engine: lm
```

StepAIC Model

```
Start:  AIC=184.95
Walc ~ sex + Pstatus + schoolsup + famsup + Mjob + Fjob + freetime +
    studytime + absences + failures

            Df Sum of Sq    RSS    AIC
- Mjob       4     6.830 669.85 181.93
- failures   1     0.027 663.05 182.97
- schoolsup  1     0.874 663.90 183.59
- famsup     1     1.341 664.37 183.94
- Pstatus    1     1.646 664.67 184.16
- freetime   1     1.923 664.95 184.36
<none>                   663.02 184.95
- Fjob       4    13.202 676.23 186.53
- absences   1    19.232 682.26 196.85
- studytime  1    21.440 684.46 198.42
- sex        1    41.513 704.54 212.47

Step:  AIC=181.93
Walc ~ sex + Pstatus + schoolsup + famsup + Fjob + freetime +
    studytime + absences + failures

            Df Sum of Sq    RSS    AIC
- failures   1     0.032 669.89 179.96
- schoolsup  1     0.558 670.41 180.34
- famsup     1     0.832 670.69 180.54
- Pstatus    1     1.544 671.40 181.05
- freetime   1     2.191 672.04 181.52
<none>                   669.85 181.93
- Fjob       4    13.307 683.16 183.49
- absences   1    19.376 689.23 193.79
- studytime  1    20.916 690.77 194.88
- sex        1    45.151 715.00 211.63

Step:  AIC=179.96
```

```
Walc ~ sex + Pstatus + schoolsup + famsup + Fjob + freetime +
    studytime + absences

            Df Sum of Sq    RSS    AIC
- schoolsup  1     0.553 670.44 178.36
- famsup     1     0.837 670.72 178.56
- Pstatus    1     1.564 671.45 179.09
- freetime   1     2.159 672.04 179.52
<none>                  669.89 179.96
- Fjob       4    13.279 683.16 181.50
- absences   1    19.355 689.24 191.80
- studytime  1    20.974 690.86 192.94
- sex        1    45.126 715.01 209.64

Step:  AIC=178.36
Walc ~ sex + Pstatus + famsup + Fjob + freetime + studytime +
    absences

            Df Sum of Sq    RSS    AIC
- famsup     1     0.747 671.19 176.90
- Pstatus    1     1.575 672.01 177.50
- freetime   1     2.243 672.68 177.98
<none>                  670.44 178.36
- Fjob       4    13.004 683.44 179.69
- absences   1    19.067 689.51 189.99
- studytime  1    20.552 690.99 191.03
- sex        1    44.635 715.07 207.68

Step:  AIC=176.9
Walc ~ sex + Pstatus + Fjob + freetime + studytime + absences

            Df Sum of Sq    RSS    AIC
- Pstatus    1     1.464 672.65 175.96
- freetime   1     2.156 673.34 176.46
<none>                  671.19 176.90
- Fjob       4    13.377 684.56 178.49
- absences   1    18.660 689.85 188.23
- studytime  1    21.734 692.92 190.39
- sex        1    45.597 716.78 206.84

Step:  AIC=175.96
Walc ~ sex + Fjob + freetime + studytime + absences
```

```
            Df Sum of Sq    RSS    AIC
- freetime   1     2.193 674.84 175.54
<none>                   672.65 175.96
- Fjob       4    13.916 686.57 177.91
- absences   1    17.605 690.25 186.51
- studytime  1    21.518 694.17 189.26
- sex        1    48.428 721.08 207.75

Step:  AIC=175.54
Walc ~ sex + Fjob + studytime + absences

            Df Sum of Sq    RSS    AIC
<none>                   674.84 175.54
- Fjob       4    13.363 688.20 177.07
- absences   1    17.354 692.20 185.88
- studytime  1    23.041 697.88 189.86
- sex        1    51.864 726.71 209.52


# A tibble: 8 x 5
  term         estimate std.error statistic  p.value
  <chr>           <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    2.15      0.249      8.65  7.67e-17
2 sexM           0.688     0.114      6.06  2.75e- 9
3 Fjobhealth     0.0785    0.338      0.232 8.17e- 1
4 Fjobother      0.287     0.211      1.36  1.76e- 1
5 Fjobservices   0.473     0.224      2.11  3.50e- 2
6 Fjobteacher   -0.164     0.314     -0.520 6.03e- 1
7 studytime     -0.274     0.0678    -4.04  6.23e- 5
8 absences       0.0400    0.0114     3.51  4.98e- 4
```

Model

term	estimate	std.error	statistic	p.value
(Intercept)	2.211	0.265	8.354	0.000
studytime	-0.307	0.085	-3.622	0.000
absences	0.022	0.030	0.732	0.464
sex_M	0.695	0.114	6.092	0.000
Fjob_health	0.066	0.339	0.194	0.847
Fjob_other	0.285	0.212	1.347	0.179
Fjob_services	0.468	0.224	2.089	0.037
Fjob_teacher	-0.153	0.315	-0.484	0.629
studytime_x_absences	0.011	0.017	0.649	0.517

> **!** Important
>
> Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.