How Does Alcohol Affect Students' Performance?

Serena Yao, Sen Feng, David Lee, Isabel Lawrence

December 16, 2022

Summary

The purpose of this report is to study the relationship between high school students' personal factors, such as their average test grades, and their overall alcohol consumption. Our analysis was carried out through cross-examination of qualitative variables such as sex, age, parent occupation/education of Portuguese high school students, and the student's tendency to consume alcohol. As a result, the impact of student alcohol consumption on math examination performance was observed for students who drank during the week rather than on weekends. Since alcohol is a depressant, consuming too much of it has been shown to slow down brain activity and slow down our reflexes. Higher alcohol levels can lead to slower reaction times, decreased attention span, and impaired judgment, all of which can affect school performance and other quality-of-life matters as well. However, if used in moderation (1-2 drinks), a few drinks of alcohol can help relieve stress and tension which could indirectly benefit how well we perform in school or at work. From our study, with limited sample size and from limited variable knowledge, we found that overall higher levels of alcohol decreased average grades for those students.

1. Introduction

Underage drinking has been a serious public health problem in many countries including the United States and Portugal. As alcohol is the most widely used substance among youth, it generally creates concerns among parents and school administrators. There has been a debate on whether drinking alcohol has an effect on students' performance. It is still unclear if the student's performance will be affected by alcohol consumption. Some research has shown alcohol consumption negatively affects students as they are more likely to skip classes, get lower grades and even drop out of school. Other studies have reported that drinking alcohol has no significant effect on the student's performance.

According to the 2018 US National Institute on Alcohol Abuse and Alcoholism, 29.8 percent of 15-year-olds report that they have had at least 1 drink in their lives. In the past month, about 7.1 million people ages 12–20 (18.8 percent of this age group) reported drinking alcohol, 11.4 percent reported binge drinking, and 2.3 percent reported heavy alcohol use.

To examine the relationship between alcohol and academic performance, analysis was conducted on a dataset collected from two Portuguese secondary schools (Gabriel Pereira and Mousinho da Silveira). We intend to investigate and understand which variables are associated with alcohol consumption behavior.

2. Data

2.1. Data Sources and Scope of Analysis

Dataset was obtained from Kaggle but originally comes from The University of Minho which is a public university in Portugal. The data approach student achievement in two Portuguese secondary schools (Gabriel Pereira and Mousinho da Silveira) and was collected by using school reports and questionnaires.

While the dataset consisted of two separate sets for math and Portuguese courses, we chose to focus on the math set to narrow the scope and avoid overlapping observations. Students' math grades is more valuable as mathematics promotes mental discipline and encourages logical reasoning and mental rigor.

The dataset used for the analysis is student-mat. This dataset provides information on each 395 students 15-22 years old regarding their grades, demographic, parental, social, and school-related features. This dataset contains 33 variables, we only select 18 which we consider to be the most important for our analysis, as shown in the table below.

Table 1: Student-related Attributes

Name	Description
sex	student's sex (binary: 0 - female and 1 - male)
age	student's age (numeric: from 15 to 22)
address	student's home address type (binary: 0 - urban and 1 - rural)
studytime	weekly study time (numeric: 1 – <2 hours to 4 – >10 hours)
failures	number of past class failures (numeric: n if 1 <= n <3, else 4)
Walc	weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)
Dalc	workday alcohol consumption (numeric: from 1 – very low to 5 – very high)
health	current health status (numeric: from 1 – very bad to 5 – very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20)

Table 2: Parent-related Attributes

Name	Description
Pstatus	parent's cohabitation status (binary: 0 - apart or 1 - living together)
Medu	mother's education (numeric: from 0 to 4)
Mjob	mother's job (nominal)
Fedu	father's education (numeric: from 0 to 4)
Fjob	father's job (nominal)
famsize	family size (binary: <= 3 or >3)
famrel	quality of family relationships (numeric: from 1 – very bad to 5 – excellent)

Education Level: 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education.
Occupation: teacher, health care related, civil services (e.g. administrative or police), at home or other.

2.2. Summary statistics

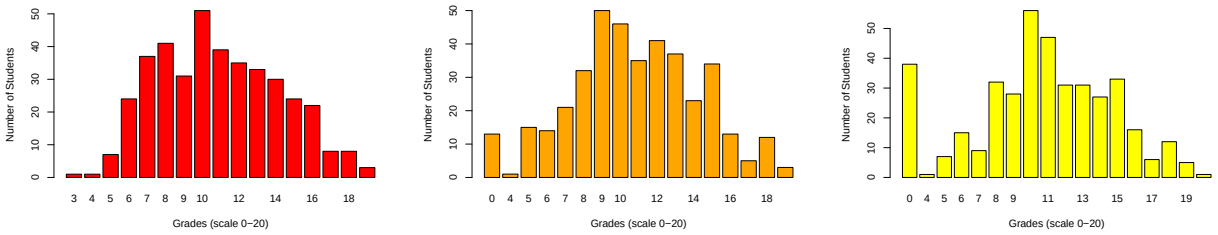
Table 3: Summary Statistics

Statistic	Mean	St. Dev.	Min	Max
Sex	0.47	0.50	0	1
Age	16.70	1.28	15	22
Parent's Cohabitation Status	0.90	0.31	0	1
Mother's Education	2.75	1.09	0	4
Father's Education	2.52	1.09	0	4
Quality of Family Relationships	3.94	0.90	1	5
Weekly Study Time	2.04	0.84	1	4
Number of Past Class Failures	0.33	0.74	0	3
Weekend Alcohol Consumption	2.29	1.29	1	5
Workday Alcohol Consumption	1.48	0.89	1	5
Current Health Status	3.55	1.39	1	5
Number of School Absences	5.71	8.00	0	75
First Period Grade	10.91	3.32	3	19
Second Period Grade	10.71	3.76	0	19
Final Period Grade	10.42	4.58	0	20
Average Grade from Three Period	10.68	3.70	1.33	19.33
Notes	The grades are based on a 20-point scale which is the most common grading system in Portugal.			

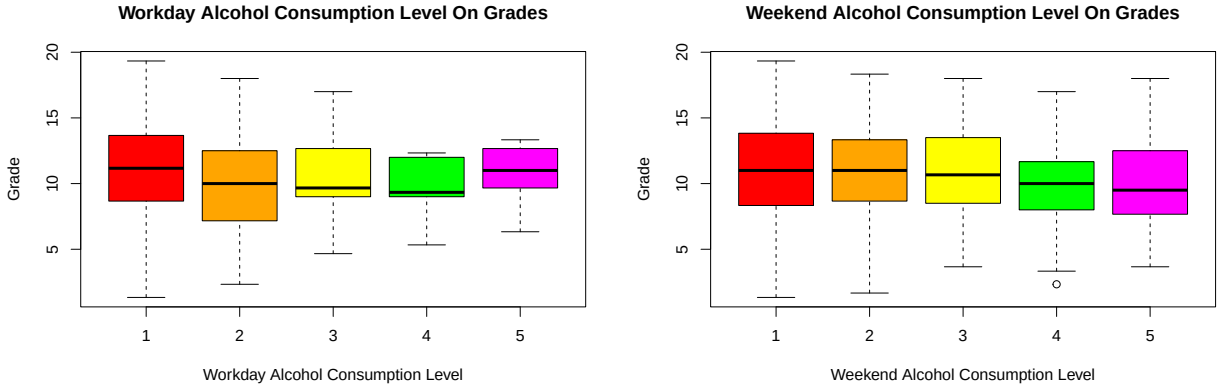
3. Analysis

3.1 Data Exploration

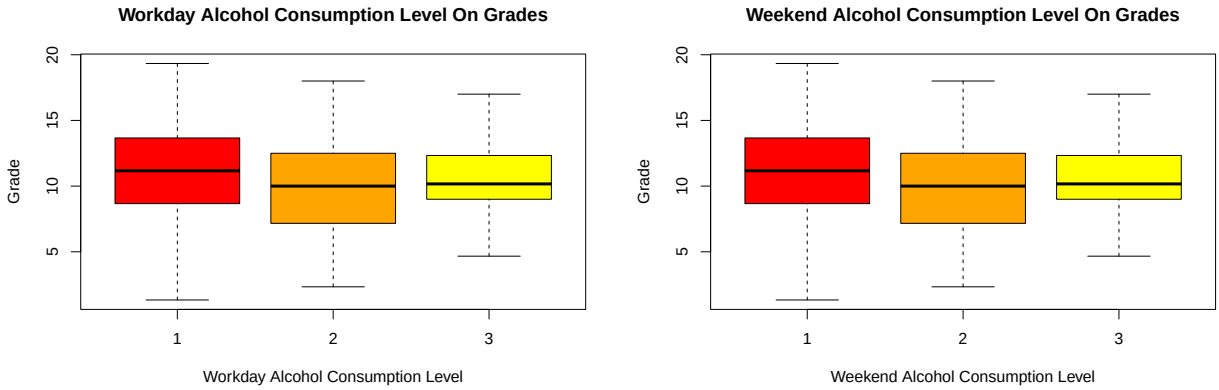
We first look at the distribution of grades across all there marking periods. The grades are mostly allocated over a normal distribution beside a peak with students receiving zeros on the final period.



Then we used boxplots to visualize the level of alcohol consumption both on workday and weekend. There seems to be a trend that the math grades decreases as alcohol intake increases from levels 1 to 4. However, we observed unexpected outcome for level 5. The small sample size made our data less significant than other levels.



Given the small number of observations for alcohol consumption levels 4 and 5, we combined them into level 3 to observe trends with higher sample sizes.



The updated boxplots shows that a higher level of alcohol consumption among students results in a lower average grade. Although the mean for the highest alcohol consumption is a little higher than the previous level, we can observe that overall range decreased.

3.2 Linear Regressions

3.2.1 Alcohol Consumption

$$meanGrade = \alpha + \beta_1 alcoholConsumption + \epsilon_i,$$

Where i is equal to each individual student, Y_i is equal to the math grade, and X_{1i} is equal to their alcohol consumption level

Table 4: Alcohol Consumption

	Dependent variable:
	meanGrade
as.factor(Dalc)2	-1.088^{**}
	(0.479)
as.factor(Dalc)3	-0.537
	(0.598)
Constant	10.946^{***}
	(0.222)
Observations	395
R^2	0.014
Adjusted R^2	0.009
Residual Std. Error	3.681 (df = 392)
F Statistic	2.709^{*} (df = 2; 392)
Note:	*p<0.1; **p<0.05; ***p<0.01

This regression table shows that consuming alcohol decreases the average mean math score. To interpretate the regression table, we observed that with every level increase in alcohol consumption level, the mean math score is predicted to decrease. From the regression residuals, we see that our residuals were relative symmetrical and that our model was predicting evenly at both the high and low ends of our dataset. Our coefficients show that consuming alcohol lowers the average math scores, but there are possible inconsistencies due to smaller sample sizes for higher alcohol consumption levels (Dalc having less sample size than others), leading to higher error values and higher p-values. Regardless of accurate values, we can conclude from this regression that consuming alcohol at any level will decrease your average math score.

3.2.2 Demographics (gender, age, family size, urban)

$$meanGrade = \alpha + \beta_1 Alcohol + \beta_2 Gender + \beta_3 Age + \beta_4 FamSize + \beta_5 Urban + \epsilon_i,$$

Where i is equal to each individual student, Y_i is equal to the math grade, X_{1i} is equal to their alcohol consumption level, X_{2i} is gender, X_{3i} is age, X_{4i} is family size (Less than or equal to 3 members), and X_{5i} is whether or not the household is located in an urban area.

Table 5: Demographics

	Dependent variable:
	meanGrade
as.factor(Dalc)2	-1.245^{***}
	(0.479)
as.factor(Dalc)3	-0.739
	(0.621)
sex	0.876^{**}
	(0.380)
age	-0.332^{**}
	(0.146)
famsizeLE3	0.723^{*}
	(0.408)
address	-0.660
	(0.446)
Constant	16.066^{***}
	(2.436)
Observations	395
R^2	0.059
Adjusted R^2	0.044
Residual Std. Error	3.614 (df = 388)
F Statistic	4.025^{***} (df = 6; 388)
Note:	*p<0.1; **p<0.05; ***p<0.01

This regression table gives us a better understanding of the data than the first regression, also adding in demographic information such as gender, age, whether the family size is less than or equal to 3, and home address is locate at urban or rural area. Adding a relevant variable can prevent bias in the estimate of the other regression coefficient but can also increase the variance of other regression coefficients.

Once again looking at the residuals we see that there is clear symmetry in predicting evenly at both the high and low ends of our dataset. Observing our coefficients we see that there are some statistically significant variables, such as gender and age.

Overall we see that: higher alcohol levels result in less average mean score, being male increases your mean score, being older typically results in lower mean scores, a family size of fewer than 3 results in higher mean scores, and being in an urban household increases mean scores. With some of these variables having high p-values and error values, we can conclude that this regression gives us a good idea of trends, but not totally accurate values.

3.2.3 Parental information (education)

$$meanGrade = \alpha + \beta_1 Alcoho + \beta_2 motherEducation + \beta_3 fatherEducation + \epsilon_i,$$

Where i is equal to each individual student, Y_i is equal to the math grade, X_{1i} is equal to their alcohol consumption level, X_{2i} is the level of their mother's education , X_{3i} is the level of their father's education.

Table 6: Parental information

	Dependent variable:
	meanGrade
as.factor(Dalc)2	-1.110^{**}
	(0.467)
as.factor(Dalc)3	-0.569
	(0.583)
Medu	0.634^{***}
	(0.211)
Fedu	0.205
	(0.213)
Constant	8.692^{***}
	(0.531)
Observations	395
R^2	0.067
Adjusted R^2	0.057
Residual Std. Error	3.590 (df = 390)
F Statistic	6.970^{***} (df = 4; 390)
Note:	*p<0.1; **p<0.05; ***p<0.01

Compared to the other regression models, this table shows slightly different coefficients for alcohol consumption's effect on average math scores. With the addition of two other variables, we can observe that the mother's education level is a statistically significant variable, and higher levels of education for both mother and father result in higher average math scores.

4. Conclusion

From our study, we can conclude that higher levels of alcohol consumption end up in a decrease in exam grades. Besides alcohol consumption, there were no significant correlations between student academic performance and other possible variables, such as age or family size. Some of these non-significant variables can be explained by our limited sample size relative to the true population, as our study participants were selected through only two high schools in Portugal. For our study's shortcomings and possible future directions, we believe that a larger subject size of students from different populations would've allowed for clearer data trends and possibly new significant correlations. Working with a broader range of potential confounding variables would help us understand this initial relationship (between alcohol consumption and average test score) further as well. Although there won't be any novel policy implications that arise from this study, this study does help signal the harms and effects that come with alcohol consumption for young students.

References

Balsa, A. I., Giuliano, L. M., & French, M. T. 2011. "The effects of alcohol use on academic achievement in high school." *Economics of Education Review*

Bonnie, R. J., Mary Ellen O'Connell, & National Research Council (US) and Institute of Medicine (US) Committee on Developing a Strategy to Reduce and Prevent Underage Drinking. 2013. "Drinking and Coming of Age in a Cross-Cultural Perspective." *NIH*

Cortland Area Communities That Care. 2020. "Teen Drinking in Europe Vs. the United States." *CACTC*

Cortez, P., & Silva, A. 2008. "USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE." *University of Minho*

Devaux, M., & Vuik, S. 2021. "The relationship between alcohol consumption and educational outcomes in children" *OECD*

El Ansari, W., Stock, C., & Mills, C. 2013. "Is alcohol consumption associated with poor academic achievement in university students?" *International Journal of Preventive Medicine*

National Institute on Alcohol Abuse and Alcoholism. 2022. "Alcohol Facts and Statistics." *NIAAA*

Schmengler, H., Peeters, M., Kunst, A. E., Oldehinkel, A. J., & Vollebergh, W. A. M. 2022. "Educational level and alcohol use in adolescence and early adulthood—The role of social causation and health-related selection—The TRAILS Study" *PLOS*