Secondary School Drinking Habits and Academic Success:
A Study on the Correlation Between the Two

Jihyeon Lee (20646185)
Juhyeon Park (20581452)
Youseung Kim (20648937)
Yunso Kim (20553699)

Abstract. In a survey by Ho et al. (2019), more than one-fifth of the underaged students answered they recently consumed alcohol despite the drinking age regulation. To examine its severity, this paper investigates the relationship between workday alcohol consumption and academic performance among secondary school students in Portugal using a two-stage least squares (2SLS) regression analysis. The results of the study demonstrate a significant negative association between workday alcohol consumption and academic performance, with a stronger impact observed on Mathematics grades. This association was observed to be more pronounced among students whose mothers attained higher levels of education and those who are under 18 years old: the legal drinking age in Portugal. This study contributes valuable insights for policymakers and educators and highlights the future importance of addressing student drinking and implementing evidence-based policies.

1. Introduction

1.1. Background of the Study

Adolescent alcohol consumption is a significant issue due to its potential ramifications on various aspects of youth development including academic performance. Despite alcohol regulation by age in most countries, it is highly improbable to prevent all students from purchasing and consuming alcohol. In Hong Kong, Ho et al. (2019) found in a survey that 21.5% of the student respondents answered that they recently consumed alcohol in a month. Among these drinkers, more than one-third (7.5% of the whole participants) admitted that they were binge drinking in the time period. These figures indicate a substantial prevalence of adolescent drinking in Hong Kong.

It seems that resolving such issue is of great importance since many studies associate drinking with underperformance at school. A study by Balsa, Giuliano, and French (2011) reveals a strong negative correlation between increases in alcohol consumption and academic

performance. Their study indicates that higher levels of alcohol consumption result in small yet statistically significant reductions in Grade Point Average (GPA) for male students. In case of female students, significant number of them has self-reported academic difficulty for female students. While the changes in GPA for female students were statistically non-significant, the self-reported academic difficulty suggests an adverse impact, warranting further investigation.

A pitfall of the study is that the alcohol consumption and school grade must not be explained directly through simple regression. A multitude of other factors, including socioeconomic status of the family, peer influence, parental involvement, mental health, and relation status are often entwined with both drinking habits and academic performance. This presents a challenge in disentangling and extracting the direct impact of drinking habits on academic performance. The endogeneity issue requires careful examination. Thus, there is a crucial need to examine the actual impact that adolescent drinking habits have on and their academic performance.

1.2. Research Objectives

This study examines the relationship between secondary school students' drinking habits and their academic performance. The data of this study is collected in Portugal, which is geographically distant from Hong Kong. However, it is expected that most aspects of the result can be extrapolated into global level based on the universality of adolescence and the widespread accessibility to alcohol across different countries. This justification is further reinforced by the previously mentioned study conducted in Hong Kong (Ho et al., 2019), which demonstrated similar drinking behaviors among adolescents.

The primary objective is to ensure an accurate and unbiased understanding of the relationship. To accomplish the goal, two main approaches have been imposed to the Two-stage Least Square (2LS) linear regression model: using instrumental variables and controlling for other parameters. Instrumental variables which are only correlated with drinking habits but unrelated directly to academic performance will be utilized in this study. Through this approach, this study aims to minimize the effect of endogenous variables and assess the causal relation of the instrumental variables of drinking on academic outcomes. This effectively isolates the direct impact of drinking from the influence of factors that might be driving the drinking behavior. It is also crucial to control for other potential confounding factors that might be influencing both drinking habits and academic performance. Using control variables, this study aims to provide a clearer effect of drinking on academic performance.

This research topic is crucial given that the academic performance potentially has long-term impacts on the life trajectories and well-being each student. A plethora of studies have identified adverse effects of alcohol consumption in the long-term, such as cognitive impairment, brain structure alterations, and functional disruptions (Lees at al., 2020). However, this does not suggest that the academic performance of the students is likewise negatively affected by alcohol consumption in the short-term. Many studies do not sufficiently separate alcohol use from other influential factors. Consequently, their findings might be skewed by these confounding variables, obscuring the true nature of the problem. One of the concerns is the legality of adolescent alcohol use. In most countries of the world, drinking age is regulated. This could affect the population of underage drinkers since they may have other characteristics that negatively affect school performance. This adds further complexity to the research issue.

In recognizing these challenges, this study plans to deepen the understanding of the relationships through a comprehensive analysis. In essence, this study aims to give policymakers, educators, and parents valuable insights into the short-term effect of adolescent alcohol consumption and academic performance. If there is a causal link between drinking habits and school grade, this may lead to the development of an evidence-based policies aimed at curbing underaged alcohol consumption. If such relation is not identified, the result leads to further studies which examine the controlled variables in this study. With these, they can make informed decisions to help mitigate the harm of this issue, ultimately promoting healthier behaviors and better academic outcomes among secondary school students.

2. Literature Review

The research by Cortez and Silva (2008), aims to develop a prediction model of the academic performance in the field of data mining. This work is particularly significant as it contains the original research conducted on the dataset that is utilized in this research. Cortez and Silva's research focuses on the capability of data mining techniques. Specifically, their study aimed to predict student performance in core subjects, Mathematics and Portuguese language, leveraging real-world data acquired from two Portuguese secondary schools through mark reports and questionnaires. The authors deployed business intelligence and data mining techniques to extract high-level knowledge from raw data, intending to impact student achievement positively.

A significant aspect of the study's methodology was its use of various data mining models: Decision Trees, Random Forest, Neural Networks, and Support Vector Machines.

They concluded that while the students' achievement was highly influenced by past two evaluations, other factors such as the number of absences, parental occupation and education, and social variables also played a crucial role. It is notable that while alcohol consumption was among the parameters considered, it did not emerge as a significant variable in predicting the final grade. This may be true since academic performance of the past has a huge impact on future performances as well. Such effect may have concealed the effect of alcohol consumption. In this study, data of past academic performance is not of a concern. Thus, this study approaches to a clearer examination on the effect of alcohol.

Data mining approach aims for a different output from the methodology this study intends to adopt. While the predictive models excel in anticipating student grades based on a range of parameters, they do not explicitly allow for an understanding of the causal relationship between the variables and the final grade. This distinction leads to one of the limitations in their study. The optimal model falls short in explicating any causal linkage between the parameters under consideration and the final grade. Hence, although alcohol consumption was included in their model, its actual impact on academic performance remains unclear within their research context. Thus, this literature work underscores the need for an econometric approach that enables a deeper examination of causal relation.

DeSimone's research (2010) serves as a comprehensive examination of the link between alcohol consumption (binging) and academic achievement among high school students. DeSimone distinctively emphasize on the role of multiple variables and a precise methodology. The study adopted a regression analysis that controls for unobserved heterogeneity in evaluating the relationship between past-month alcohol use and previous-year grades. A particularly notable facet of DeSimone's study is the comprehensive set of variables incorporated into the regression analyses. The dependent variable is the GPA, whereas proxies for risk and time preference, mental health, self-esteem, and measures of other addictive substances are used as independent variables. Furthermore, the paper utilizes the instrumental variable approach. Specifically, state regulations of alcohol, which correspond with the independent variable (the number of binging incidents), are used as an instrumental variable. This methodological decision allows for a more rigorous control of unobserved heterogeneity, which is of particular importance in this research as well, given that instrumental variables are also used in this research as well.

His study provides significant insights, and it is important to replicate the research in different populations. It is possible that cultural, social, and regulatory differences in alcohol use and academic pressures might yield different relationships between alcohol use and academic performance, underscoring the necessity of locale-specific research. This research is different from DeSimone's work in terms of the control variables. While DeSimone

controls for internal values (e.g. risk and time preferences, mental health, and self-esteem) that can easily change, this study mostly controls external values that cannot be changed (e.g. age, relationship status, parent education status, and school support). This yields a similar method, but completely different results.

There is research to suggest that some parameters should be controlled in order to measure the real effect of alcohol consumption on academic performance. Honghao, Po, and Tianyu (2021) lends weight to the premise that factors such as romantic relationships, can significantly impact academic performance. According to their study, adolescents engaged in romantic relationships exhibited a 10.5% increase in poor academic performance compared to their peers not involved in such relationships. This finding illustrates the possible effects of romantic involvement on academic outcomes, an aspect this research plans to control. Such status had a greater impact on female adolescence than male which the authors predicted was due to differences in gender roles. Gender is another controllable variable regarding such outcome.

Simultaneously, a separate study by Idris, Hussain, and Ahmad (2020) emphasizes the role of parental education levels in influencing academic achievement. A robust correlation was observed between the academic achievements of students and their parents' educational status, suggesting it being a critical factor in students' academic performance. In this study, these findings suggest the need for the inclusion of control variables such as age, gender, romantic relationships, and parental education levels, alongside alcohol consumption, to provide a more comprehensive understanding of factors influencing academic performance. By controlling these variables, this study aim to discern the specific impact of alcohol consumption on academic outcomes more accurately.

3. Data

The data utilized for this research was sourced from the paper "Using Data Mining to Predict Secondary School Student Performance," written by Paulo Cortez and Alice Silva. They collected the data during the 2005-2006 academic year in the Alentejo region of Portugal with the aim of predicting the academic performance of secondary school students (Cortez & Silva, 2008). In this research, the same dataset is used to examine the impact of alcohol consumption on student academic performance.

Portugal's secondary education system comprises of a three-year curriculum that is generally based on public and free education. The curriculum covers essential subjects such as Mathematics and Portuguese. The grading for this education system employs a 20-point scale, with 0 representing the lowest score and 20 representing the highest score possible.

Students are evaluated three times over a course, with the final evaluation serving as their ultimate grade. In this study, the mean of the final evaluations of Mathematics and Portuguese is used as the dependent variable.

The original dataset was collected from two sources: school reports and questionnaires. School reports are based on paper sheets, including a few attributes such as absences, and questionnaires consisting of 37 closed questions related to several demographic, social, emotional, and school-related variables. The questionnaires were answered by 788 students, and 111 answers were discarded due to a lack of identification details. The dataset consists of two separate files containing information on Mathematics and Portuguese language classes, respectively. The files contain 395 and 649 records of information and answers to 33 variables. However, as 380 students belonged to both datasets, a single dataset was created by identifying the common attributes that characterized each student and merging the datasets from the different sources in this study. The data is further processed by eliminating unnecessary columns.

Table 1. Preprocessed variables

Name	Description
Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
G3_mean	mean of Portuguese and Mathematics final grade (numeric: from 0 to 20, output target)
G3_por	final grade in Portuguese
G3_mat	final grade in Mathematics
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
romantic	with a romantic relationship (binary: yes or no)
higher	higher - wants to take higher education (binary: yes or no)

failures	number of past class failure (numeric: n if n < 3, else 4)
schoolsup	extra educational support (binary: yes or no)
Medu	mother's education (0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
Fedu	father's education (0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
freetime	free time after school (numeric: from 1 - very low to 5 - very high)
goout	going out with friends (numeric: from 1 - very low to 5 - very high)
famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

Table 1 indicates the variables from the original dataset that are used in this research. It includes a range of variables that have been denoted as categorical, binary, or numeric. Most of the variables in the dataset are categorical with values ranging from 1 to 5. This represents the participant's levels of agreement or disagreement with questions or statements. Some variables, such as "G3" and "failures," are numeric, indicating the student's grade and the number of previous failures, respectively. "romantic" and "higher" are binary variables, with "Yes" or "No" values.

"G3_mean" is a newly created variable, which is the mean score obtained from both Mathematics and Portuguese classes. The control variables included "romantic," "higher," "failures," "schoolsup," "Medu," and "Fedu," while the instrumental variables comprised "freetime," "goout," and "famrel." The dependent variable was G3_mean and the independent variable was Dalc, which represents the student's weekday alcohol consumption levels. Additionally, some variables, such as "age" and "sex", were used to describe the demographics of the data.

4. Descriptive Statistics Analysis

Descriptive statistics were utilized to obtain a comprehensive understanding of the dataset. Bivariate analyses were conducted between the control variables and dependent variables, as well as between the instrumental variables and independent variables. These analyses provided valuable insights into how these variables impact the dependent and independent variables of interest. Through the use of univariate and bivariate analyses, any potential relationships and patterns between variables were identified.
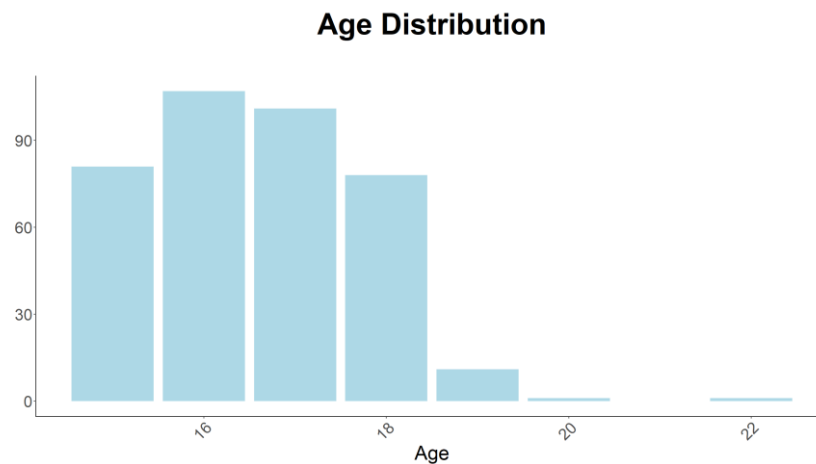
4.1 Univariate Analysis
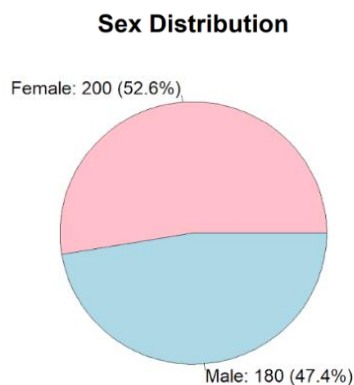

Figure 1. Bar graph for students' age

Figure 2. Pie chart for students' sex

The dataset consists of 380 secondary school students, with 52.6% (200) male and 47.4% (180) female. The age of the students ranges from 15 to 22 years old, with the majority falling in the 15-18 age group. The age distribution is skewed to the right, with about 50% of the students being either 15 or 16 years old. The median age of the students is 17 and the mean age is 16.58.
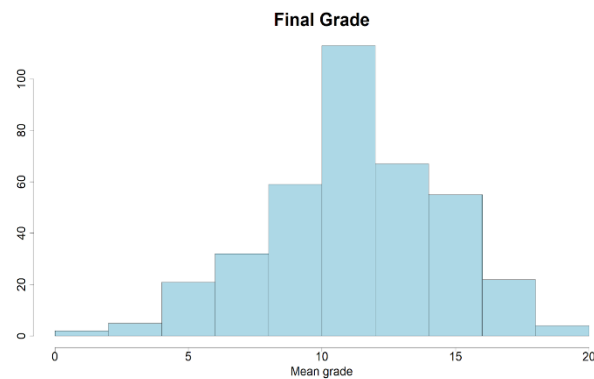
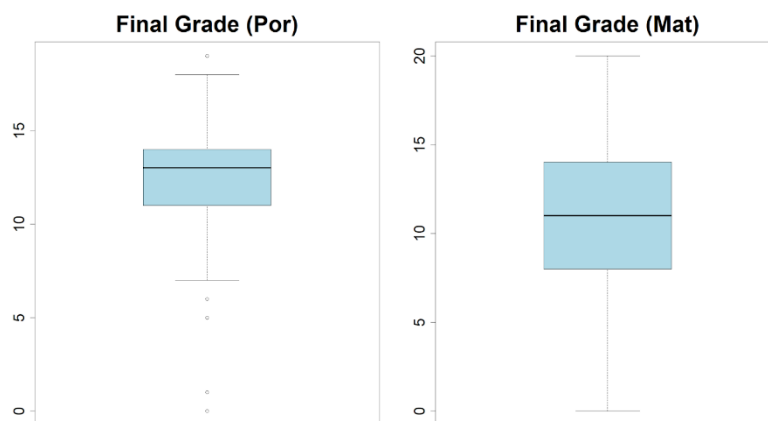**Figure 3. Histogram for students' final grade (G3_mean)**

Figure 4. Box plots for students' final grade on two subjects (Portuguese and Math)

The final grades in this dataset are based on a 21-point scale, with separate grades provided for two courses: Portuguese and Mathematics. The mean final grade for Portuguese is 12.89, with a median of 13, while the mean final grade for Mathematics is 11.08, with a median of 11. These central tendencies shows that the mean final grade for Portuguese is higher than that of Mathematics, indicating that students tend to perform better in Portuguese than in Mathematics. The distribution of Portuguese grade is skewed to the left with smaller variance while Mathematics grade is widely spread. Considerable number of students were found to score 0, the minimum grade in Mathematics compared to Portuguese grade. To combine the two scores, the G3_mean variable is used as a mean of these two scores, which has a mean of 11.89 and a median of 12. The distribution of G3_mean is approximately normal, with a standard deviation of 3.11, as shown in the graph.
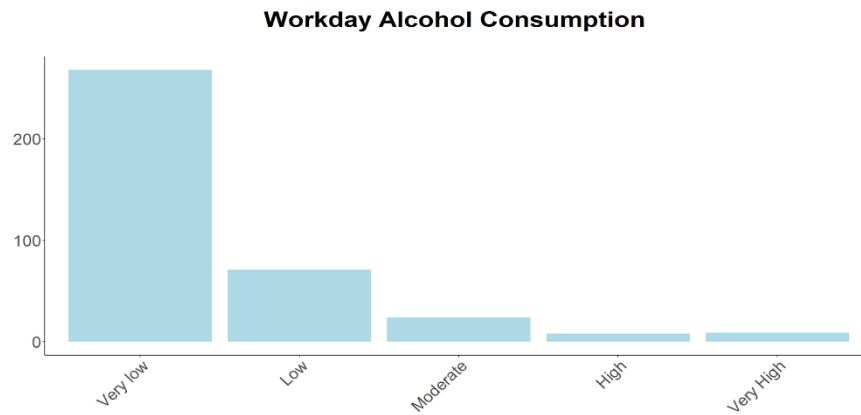
Workday Alcohol Consumption

Figure 5. Bar graph for workday alcohol consumption

The weekday alcohol consumption, the independent variable of the study, ranges from 1 to 5, with 1 indicating very low consumption and 5 indicating very high consumption. The bar graph displays the frequency of alcohol consumption levels among students. The most commonly reported level of consumption is very low, followed by low. There is a notable disparity between high and low levels of alcohol consumption, with the majority of students reporting very low levels of consumption during the day. These findings suggest that alcohol consumption among students on weekdays is generally low, with only a small proportion of students reporting high levels of consumption.
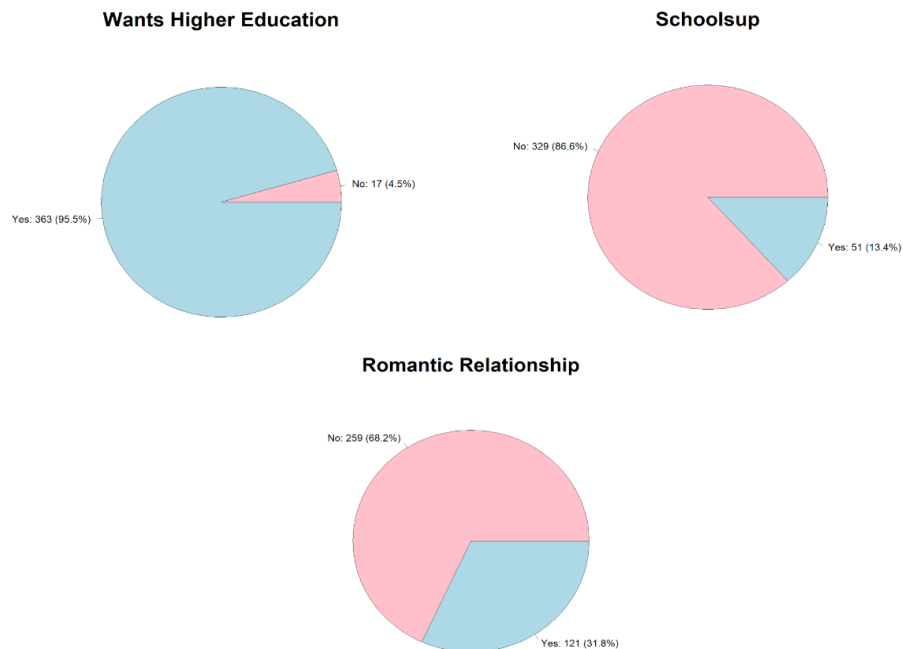

Figure 6. Pie charts for binary control variables

In the research, there are several control variables that could potentially impact the relationship between alcohol consumption and student academic performance. The control variables incorporated three binary variables: romantic, higher, and schoolsup. The variable romantic indicates whether a student is in a romantic relationship or not. The vast majority of the students (68.2%) reported that they were not in a romantic relationship. The variable "higher" reflects whether a student intends to pursue higher education. A large majority of the students (95.5%) expressed a desire to continue their studies beyond their current level, indicating a high level of motivation to pursue higher education. On the other hand, the variable "schoolsup" indicates whether a student receives additional educational support. Only a small proportion of students (13.6%) reported receiving such support, indicating that most students are able to manage their academic workload without additional assistance.
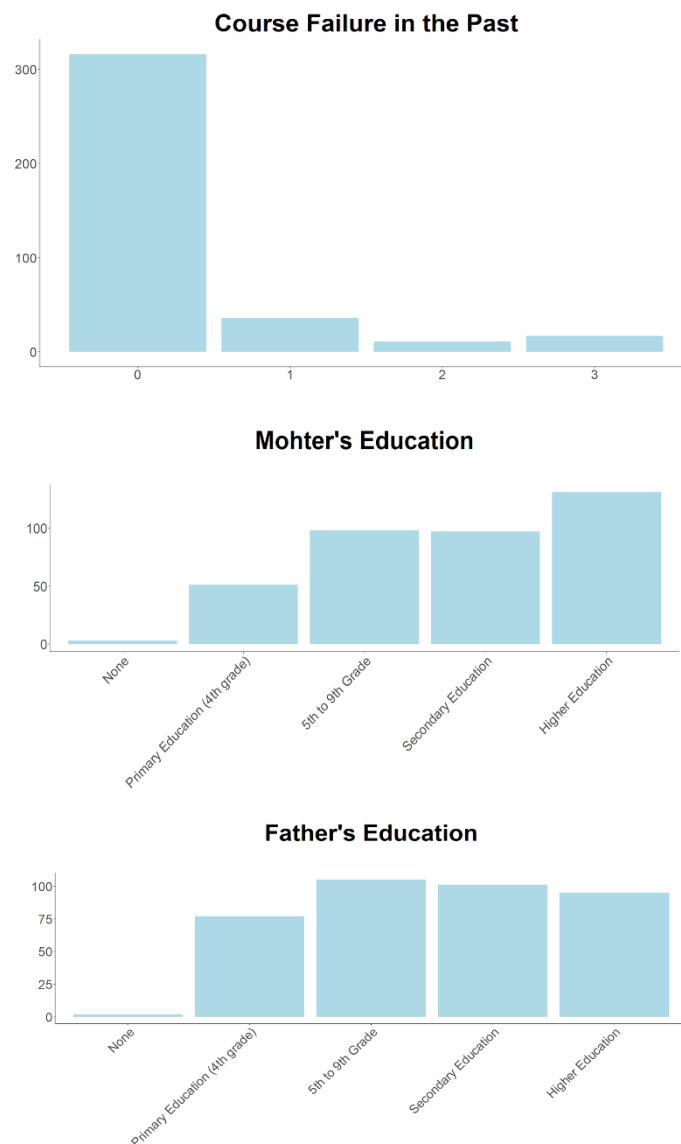
Figure 7. Bar graphs for ordinal control variables

In addition to the binary variables, three ordinal control variables were included in the analysis. These variables include the number of past failures (failures), as well as the education level of the student's mother (Medu) and father (Fedu).   The bar graph for the "failures" variable shows that most of the students did not receive any failures in the past. It also shows there is no big difference in the frequency of students who received 1, 2, or 3 failures.

The education level is measured on a scale from 0 to 4, with 0 indicating no education and 4 indicating an education level higher than secondary education. The most frequent level of education for the father was 5th to 9th grade, while the most frequent level for the mother was higher than secondary education. Also, the graphs are showing that the overall education level of the mother's education was higher than that of the father's education.
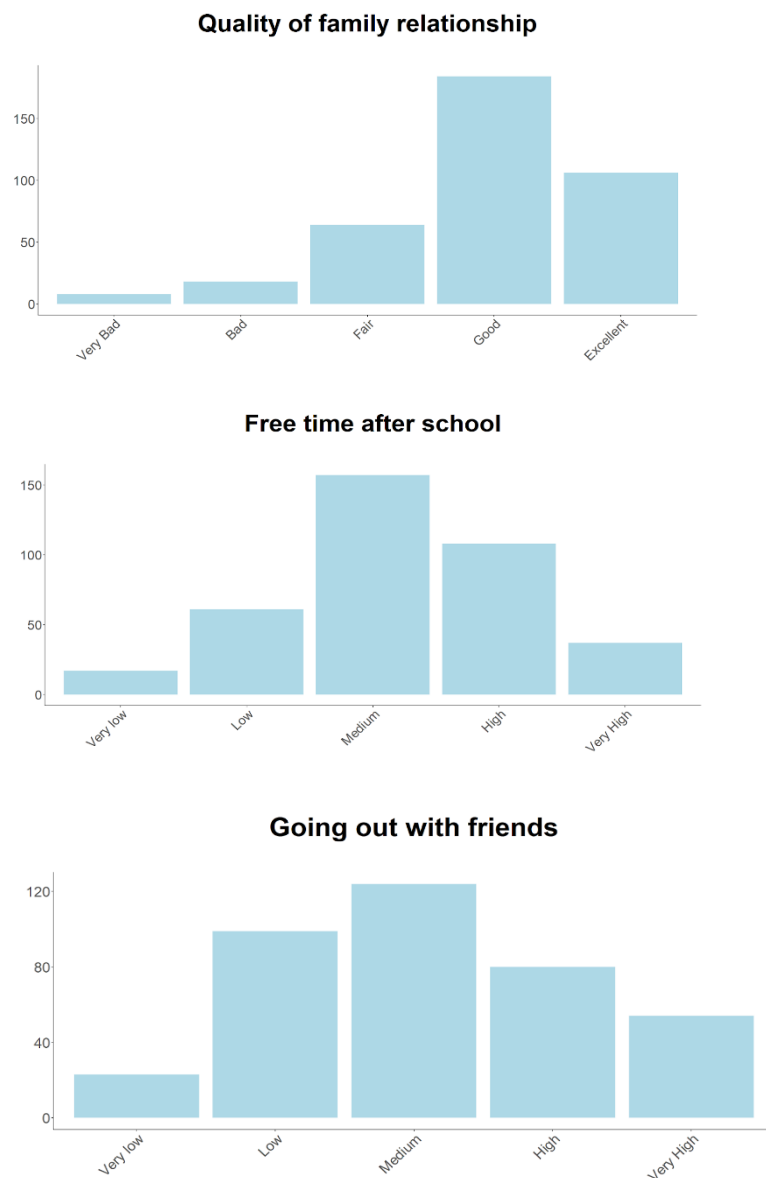

Figure 8. Bar graphs for instrumental variables

There are three instrumental variables to help identify causal relationships between other variables of interest: "freetime", "goout", and "famrel". "freetime" measures the amount of free time a student has after school, with a range from 1 (very low) to 5 (very high). The highest frequency observed was at a medium score of 3, with a slightly higher frequency observed on the higher score side, though the difference was not significant. "gout" measures how often a student goes out with friends, with a similar range of scores. The highest frequency was also observed at a medium score, and there was no significant difference between high and low scores. Finally, "famrel" measures the quality of a student's family relationships, with a range from 1 (very bad) to 5 (excellent). The highest frequency observed was at a score of 4 (good), while the second highest frequency was at a score of 5 (excellent). There was a significant difference between low and high scores, with most students reporting good family relationships and only a small number reporting bad relationships.

4.2. Bivariate Analysis
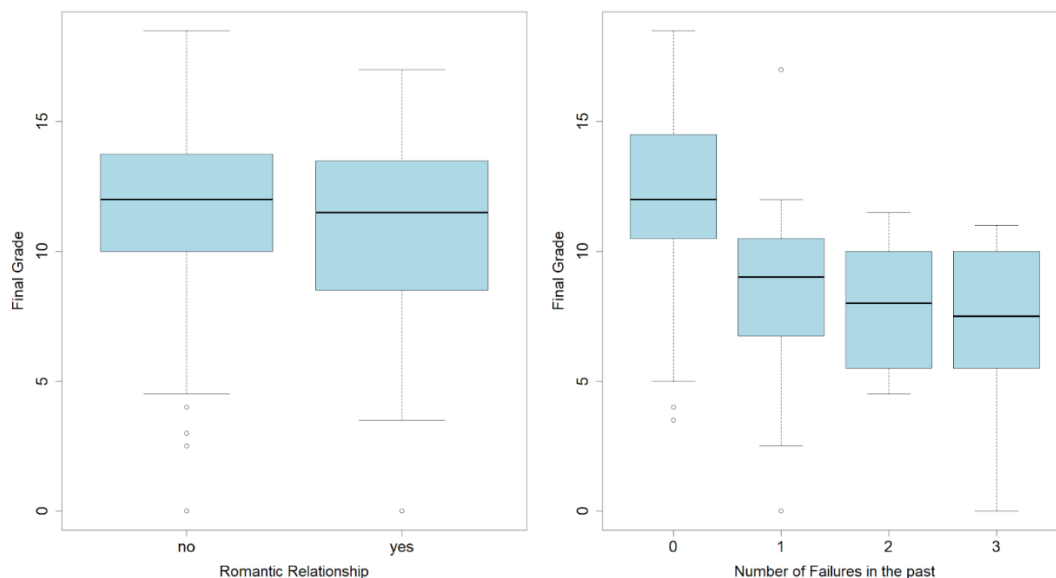
Figure 9. Box plots of control variables (romantic & failures) vs. dependent variable

Bivariate analysis with descriptive statistics also was employed to evaluate how the control variables affect the dependent variable. According to the graph depicting the relationship between romantic relationships and final grades, students who are not in a romantic relationship tend to have slightly higher final grades than those who are. The variable failures measure the number of times a student has failed in the past, and the analysis showed a clear negative relationship between failures and final grades. In other words, students who had failed more in the past were more likely to receive a lower final grade.

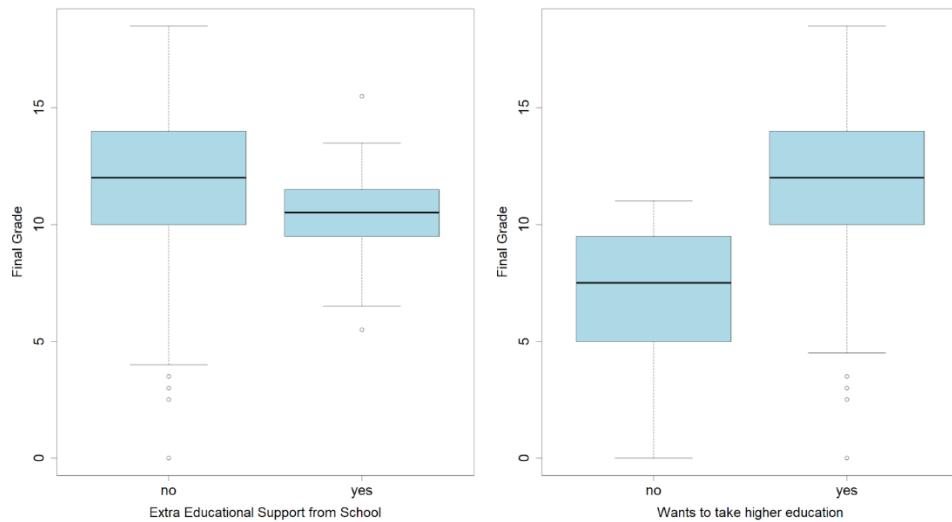**Figure 10. Box plots of control variables (schoolsup & higher) vs. dependent variable**

The variable schoolsup was used to determine whether a student received additional educational support, and it was expected that the final grade would be higher for students who received this support. However, the analysis showed a negative relationship between schoolsup and the final grade. This suggests that students who receive extra educational support from their school are more likely to receive a lower final grade, possibly because schools tend to provide support to students who are struggling academically. On the other hand, the variable higher measures whether a student wants to pursue higher education, and the analysis showed that students who want to pursue higher education have a significantly higher final grade than those who do not want to pursue higher education.
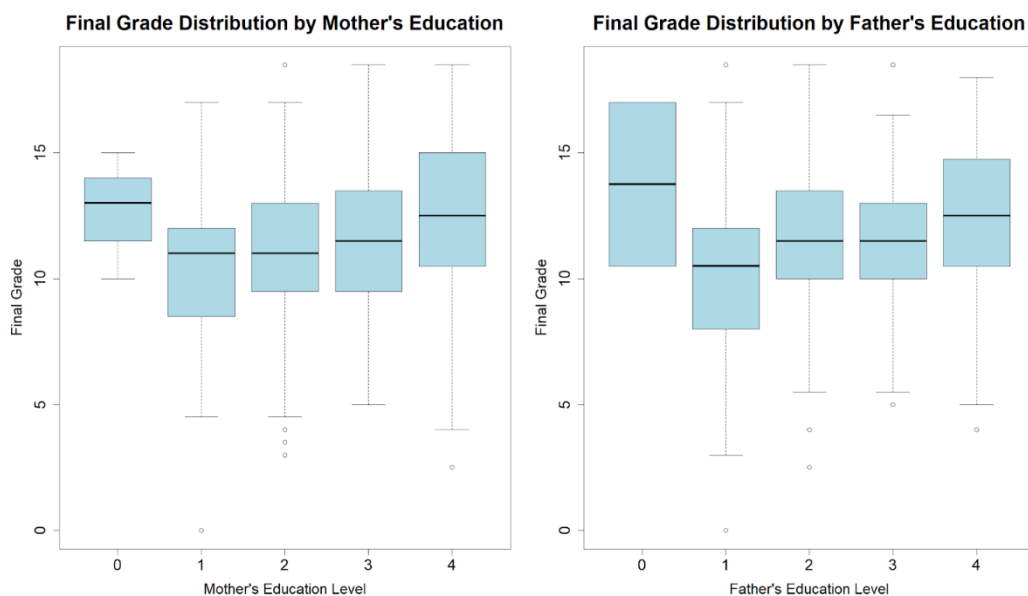
Figure 11. Box plots of control variables (Fedu & Medu) vs. dependent variable

Medu and Fedu measure the parents' education levels ranging from 0 (no education) to 4 (higher than secondary education). The findings of this study indicate a positive correlation between parents' education and final grades, with higher levels of education being associated with higher grades, except for the category of 0 education level. it is important to note that the graphs for both mothers' and fathers' education levels only had a small sample size of two individuals with a score of 0. This limited sample size may not be representative of the overall population. Therefore, this analysis suggests that parents' education positively affects a student's final grade.

Figure 12. Box plots of instrumental variables vs. independent variable

In addition, bivariate analysis was conducted to investigate the association between instrumental variables and independent variables. Results indicated a significant positive relationship between freetime and alcohol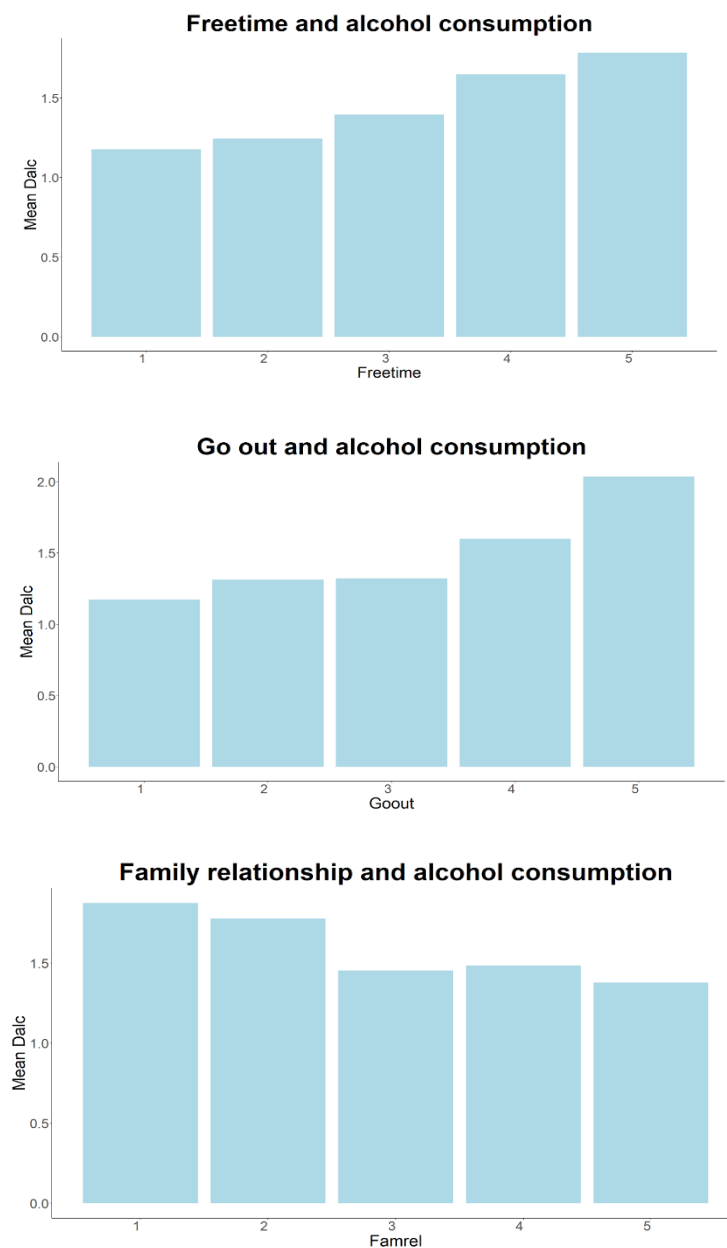 consumption, suggesting that students consume more alcohol as their free time increases. Additionally, students who go out frequently were found to consume more alcohol, indicating a positive relationship between goout and alcohol consumption. The study also revealed that family relationships play a significant role in alcohol consumption among students. Specifically, students with poor family relationships were found to consume more alcohol than those with better family relationships.

5. Methodology

$$y = \beta_0 + \beta_x \hat{x} + \beta_{romantic} W_{romantic} + \beta_{higher} W_{higher} + \beta_{failures} W_{failures}$$
$$+ \beta_{schoolsup} W_{schoolsup} + \beta_{Medu} W_{Medu} + \beta_{Fedu} W_{Fedu} + \varepsilon$$

$$\hat{x} = \gamma_0 + \gamma_{freetime} Z_{freetime} + \gamma_{goout} Z_{goout} + \gamma_{famrel} Z_{famrel} + v$$

In this study, the relationship between students' alcohol consumption during the workday and academic performance is primarily examined using two-stage least squares (2SLS) regression analysis. Even though there is a significant correlation between the two factors, this may not necessarily indicate causation because socioeconomic status or personality traits may also influence alcohol consumption and academic performance. Regression analysis may result in biased estimates as an output of this correlation, which might lead to incorrect conclusions. Instrumental variables must be used to isolate the causal relationship between alcohol consumption and academic performance and avoid skewed estimates. Endogeneity, which happens when an interesting variable is correlated with the model's error term, is dealt with using instrumental variables. This correlation may appear for a variety of reasons, including measurement error, simultaneity bias, or omitted variables. Endogeneity can occur in the discussion of student alcohol consumption and academic performance because students who drink more may also perform worse academically for other reasons, such as a lack of motivation, poor study habits, or family background. These factors can affect both academic performance and alcohol consumption even though they may not be obvious or measurable. Thus, in order to investigate the connection between student alcohol use and academic achievement, instrumental variables need to be utilized in this study as endogeneity can be handled and a method for calculating the causal impact of alcohol use on academic performance can be provided.

In the 2SLS regression model employed in this study, the mean final grade of math and Portuguese serves as the outcome variable, y ,while the alcohol consumption of students during workday is the endogenous variable, $\hat{x}$. To control for the potential endogeneity bias in

the relationship between alcohol consumption and academic performance, the regression model uses instrumental variables, Z_i, that have the potential to affect alcohol consumption, including free time after school, quality of family relationships, and frequency going out with friends. Each instrumental variable is associated with a coefficient, γ_i, and an error term v, which are estimated in the regression analysis. Also, to account for potential confounding factors, several control variables, W_i, are included in the 2SLS regression analysis. These control variables are willingness to take higher education, number of past class failures, extra educational support, and education level of parents. Willingness to take higher education is a binary variable that represents a student's interest in pursuing higher education after completing their current curriculum, and is included because students who are more interested in pursuing higher education may be more motivated to perform well academically; the number of past class failures is a numeric variable that represents the number of times a student has failed a class in the past, and is included because students who have failed classes in the past may have lower academic performance due to a lack of understanding of the material; extra educational support is a binary variable that represents whether a student is receiving extra educational support, and is included because students who are receiving extra educational support may have better academic performance due to the additional resources and guidance; the education level of parents is a numeric variable that represents the education level of a student's parents, and is included because students whose parents have a higher education level may have better academic performance due to the advantages and opportunities provided by their parents. Each control variable is associated with a coefficient, β_i, and an error term ε.

5.1. Feature selection

For selecting features, a statistical test is also utilized as a reference in this study. Using statistical tests is one efficient approach to identify variables that are strongly correlated with the endogenous variable while being uncorrelated with the error term. The Kendall correlation coefficient is one such statistical test that can be used for this purpose. The Kendall correlation coefficient measures the strength and direction of the association between two variables and is a nonparametric measure that is well-suited for identifying instrumental variables in situations where the underlying distribution of the variables may not be known or may deviate from normality. Kendall correlation coefficient ranges from -1 to +1, with a coefficient of 0 indicating no correlation, a coefficient of 1 indicating a perfect positive correlation, and a coefficient of -1 indicating a perfect negative correlation.
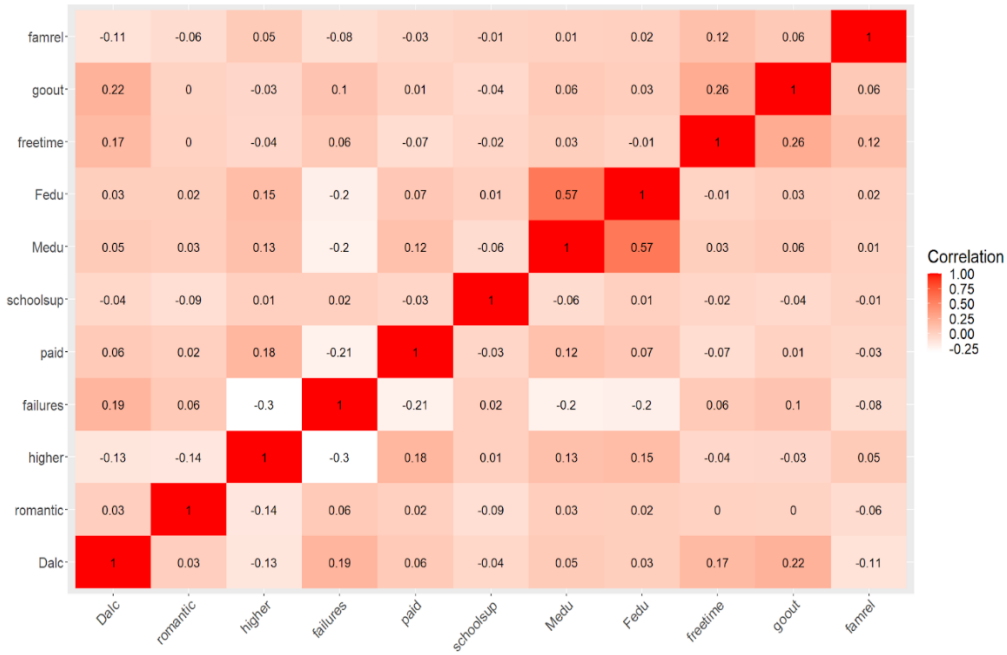
Figure 13. The Kendall correlation of features

The figure 13 indicates that the variables goout, freetime, and famrel have Kendall correlation coefficients of 0.22, 0.17, and -0.11, respectively, which suggests a relatively high level of correlation among available variables. These coefficients indicate a strong positive correlation among goout, freetime and alcohol consumption during workday; in contrast, famrel showed a moderate negative correlation with alcohol consumption during workday, indicating that higher levels of family relationship quality may be associated with lower levels of alcohol consumption during workday. Moreover, the coefficients between the three instruments and the six control variables included in the analysis are below 0.1, indicating a low level of correlation between these variables and the endogenous variable. This finding suggests that there is less possibility of endogeneity problems between these variables, which increases the validity of the regression analysis.

Table 2. IV diagnostic tests

IV Diagnostic tests				
		Dependent variable: Mean Grade		
IVs:	Freetime	Goout	Famrel	Freetime, Goout, Famrel
Weak instruments	8.31e-05 ***	2.37e-07 ***	0.0779 *	7.7e-09 ***
Wu-Hausman	0.903 *	0.0626 *	0.2754	0.0598 *
Sargan				0.4674
Note:				$^{*}p<0.01; ^{**}p<0.05; ^{***}p<0.01$

5.2. Robustness and validity check

2SLS regression analysis depends heavily on the selection of the instrumental variables. The instrumental variables must be pertinent, which means they must be highly correlated with the variable of interest, student's workday alcohol consumption. Additionally, they ought to be valid, that is, they shouldn't be correlated with the model's error term. The instrumental variables must also be exogenous, which means they must not be impacted by the outcome variable. Lastly, the relationship between the endogenous variable and the outcome variable is linear, as nonlinear relationships may lead to biased or inconsistent estimates. These presumptions can be broken, resulting in biased calculations and incorrect conclusions. To check robustness and validity of instrumental variables, IV diagnostic tests are used in this study.

The following 3 tests are the major tests used in the diagnostic tests.

- Weak instruments: This is an F-test on the instruments in the first stage. The null hypothesis is essentially that instruments are not jointly significant, so a rejection means that the instruments are significant, which indicates they are suitable to utilize.

- Wu-Hausman: This tests the consistency of the OLS estimates under the assumption that the IV is consistent. When the null gets rejected, it indicates that OLS is inconsistent, and endogeneity may be present. If the null is accepted, endogeneity may not have been a major issue because the OLS and IV estimates are essentially similar.

- Sargan: This is a test of instrument exogeneity using overidentifying restrictions, called the J-statistic in Stock and Watson. If the null is rejected, it means that at least one of selected instruments is invalid, and possibly all of them. (It can only be used if there are more instruments than endogenous regressors)

According to table 2, all the instrumental variables, freetime, famrel, goout have significant correlations on weak instruments which prove that the instrumental variables are strongly correlated with the endogenous variable of interest, and the values of Wu-hausman are accept the null hypothesis at a significance level of 0.05. Also, the Sargan value accepts the null hypothesis, it means that selected instruments are appropriate to use. These tests can help to improve the robustness and validity of the regression analysis, ultimately providing more accurate estimates of the causal effect of the endogenous variable on the outcome variable.

6. Results

6.1. Dependent and Control Variables

Table 3. Multiple Linear Regressions on Academic Performance

Multiple Linear Regression Analysis on Academic Performance

	Dependent variable: Mean Grade		
	Workday Alcohol Consumption (1)	Control Variables (2)	Workday Alcohol Consumption with Control Variables (3)
Dalc	-0.546***		-0.362**
	(0.188)		(0.167)
Romantic		-0.655**	-0.632**
		(0.320)	(0.319)
Higher		2.054***	1.984**
		(0.781)	(0.778)
Failures		-1.553***	-1.496***
		(0.224)	(0.224)
Schoolsup		-1.109**	-1.121**
		(0.436)	(0.434)
Medu		0.421**	0.442**
		(0.180)	(0.179)
Fedu		0.009	0.012
		(0.179)	(0.179)
Constant	12.288***	9.125***	9.636***
	(0.323)	(0.875)	(0.902)
Observations	380	380	380
R^2	0.022	0.248	0.257
Adjusted R^2	0.019	0.236	0.243
Residual Std. Error	3.250 (df = 378)	2.869 (df = 373)	2.855 (df = 372)
F Statistic	8.438*** (df = 1; 378)	20.471*** (df = 6; 373)	18.389*** (df = 7; 372)

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3 illustrates the results of three linear regression analyses which investigated the effects of the workday alcohol consumption and other control variables, without accounting for instrumental variables.

The first simple linear regression only considered alcohol consumption as an independent variable. It suggests that the mean score of academic performance tends to decrease by 0.546 as the level of alcohol consumption increases by 1, which is found to be statistically significant at 0.01 level.

In the second regression analysis, all control variables were examined as independent variables. As a result, all variables, except for the father's education, are significantly associated with academic performance, which indicates that the inclusion of these control variables in the model is crucial for a more accurate understanding of the impact of alcohol consumption on academic performance.

The third regression analysis was conducted using both the workday alcohol consumption and controls as independent variables. The negative impact of alcohol consumption decreased to negative 0.36, compared to the result in the first regression. Although the size of the coefficient in the third regression is smaller than the first one, it is still statistically significant and suggests that alcohol consumption continues to have a negative impact on academic performance even after controlling for other factors.

6.2. Mean, Math, and Portuguese Grades

Table 4. Two-Stage Least Square Regression on Mean, Math, and Portuguese

Two-Stage Least Squares (2SLS) Regression Analysis: Mean, Math, and Portuguese

	Dependent variable:		
	Mean	*Math*	*Portuguese*
	(1)	(2)	(3)
Dalc	-1.135**	-1.490*	-1.101**
	(0.571)	(0.896)	(0.528)
Romantic	-0.582*	-0.955*	-0.039
	(0.330)	(0.496)	(0.292)
Higher	1.834**	1.807	2.313***
	(0.807)	(1.189)	(0.701)
Failures	-1.376***	-1.801***	-0.927***
	(0.246)	(0.361)	(0.213)
Schoolsup	-1.148**	-0.686	-1.362***
	(0.446)	(0.664)	(0.391)
Medu	0.488***	0.565**	0.357**
	(0.187)	(0.273)	(0.161)
Fedu	0.018	-0.003	0.044
	(0.184)	(0.270)	(0.159)
Sex		1.917***	-0.675*
		(0.604)	(0.356)
Constant	10.728***	9.339***	11.611***
	(1.205)	(1.668)	(0.983)
Observations	380	380	380
R^2	0.214	0.142	0.265
Adjusted R^2	0.200	0.123	0.249
Residual Std. Error	2.936 (df = 372)	4.311 (df = 371)	2.542 (df = 371)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4 presents the findings from the Two-Stage Least Squares (2SLS) regression analyses, which include all instrumental variables discussed in the previous section.

This study conducted three separate regression analyses with the mean grade, math grade, and Portuguese grade as the dependent variables to investigate the effects of alcohol consumption on academic performance in different subjects. The sex variable was included as an additional control variable to the last two regression models in order to address the gender differences frequently observed in academic performance of students, which tend to associate STEM subjects with male domains and non-STEM subjects such as languages, social studies, and humanities with female domains. The results shown in the table estimate a positive association (β=1.917, $p < 0.01$) between the sex variable and math grades, but a negative association (β=-0.675, $p < 0.1$) with Portuguese grades. These findings are consistent with gender differences in student performance, suggesting that gender should be taken into account when comparing the grades of STEM and non-STEM subjects.

The results reveal that workday alcohol consumption has a statistically significant negative impact on academic performance regardless of subjects, with coefficients of -1.135 ($p < 0.05$), -1.490 ($p < 0.1$), and -1.101 ($p < 0.05$) for mean, math, and Portuguese grade, respectively. The strongest negative correlation found between alcohol consumption and math grades implies that alcohol consumption may have a stronger negative effect on students' math-related performance compared to other subjects.

Moreover, while the mother education variable is consistently significant across all three models, the father education variable shows no significance in any of the models. This suggests that the impact of daily alcohol consumption on students' academic achievement may vary depending on the education levels of their parents, with maternal education having a stronger influence than paternal education.

6.3. Parental Education Level

Table 5. Two-Stage Least Square Regression on Parental Education Level

Two-Stage Least Squares (2SLS) Regression Analysis: Parental Education Level

	Dependent variable: Mean Grade			
	Mother with Low Education	Mother with High Education	Father with Low Education	Father with High Education
	(1)	(2)	(3)	(4)
Dalc	-0.449	-1.055*	-0.813	-0.902
	(1.002)	(0.627)	(0.718)	(0.734)
Romantic	-0.451	-0.547	-1.026*	-0.010
	(0.531)	(0.391)	(0.523)	(0.412)
Higher	1.769*	2.722**	1.689*	4.788**
	(1.028)	(1.307)	(0.955)	(1.937)
Failures	-1.300***	-1.766***	-1.401***	-2.204***
	(0.303)	(0.379)	(0.298)	(0.427)
Schoolsup	0.420	-2.606***	-0.329	-2.045***
	(0.659)	(0.569)	(0.717)	(0.541)
Constant	10.237***	11.830***	11.582***	9.166***
	(1.909)	(1.700)	(1.495)	(2.356)
Observations	152	228	184	196
R^2	0.180	0.299	0.206	0.253
Adjusted R^2	0.152	0.283	0.184	0.233
Residual Std. Error	2.969 (df = 146)	2.711 (df = 222)	3.131 (df = 178)	2.656 (df = 190)

Note: *p<0.1; **p<0.05; ***p<0.01

To investigate the impact of workday alcohol consumption on academic performance across different parental education levels, the dataset was divided into four groups based on the educational attainment of the mother and father as presented in Table 5. Parents who received secondary education or above were classified as highly educated in this study.

Among the four regression analyses conducted in each group, a statistically significant negative impact of alcohol consumption on academic performance was observed only in the group of students whose mothers had a high education level, with a coefficient of -1.055.

6.4. Legal Age of Drinking

Table 6. Two-Stage Least Square Regression on Legal Age of Drinking

Two-Stage Least Squares (2SLS) Regression Analysis: Legal Age of Drinking

	Dependent variable: Final Grade	
	Age Above 18 (1)	Age Below 18 (2)
Dalc	-1.976	-1.241**
	(1.778)	(0.586)
Romantic	-0.974	-0.445
	(0.995)	(0.367)
Higher	1.544	1.361
	(1.589)	(1.192)
Failures	-0.802	-1.473***
	(0.859)	(0.253)
Schoolsup	-0.307	-1.281***
	(1.857)	(0.442)
Medu	0.369	0.544***
	(0.639)	(0.192)
Fedu	0.500	-0.105
	(0.565)	(0.189)
Constant	10.987***	11.623***
	(2.990)	(1.457)
Observations	91	289
R^2	0.108	0.190
Adjusted R^2	0.033	0.170
Residual Std. Error	3.948 (df = 83)	2.711 (df = 281)
Note:		*p<0.1; **p<0.05; ***p<0.01

To account for the legal age limit for alcohol consumption in Portugal, the two-stage least square regressions were conducted separately for two groups of students: those aged above 18 and those aged below 18. The results from Table 6 indicate that the coefficient for workday alcohol consumption among students aged below 18 is statistically significant at the 0.05 level, with a value of -1.241, while the impact of alcohol consumption among students aged above 18 is statistically insignificant.

7. Findings & Discussion

The present study aimed to investigate the relationship between workday alcohol consumption and academic achievement among secondary school students. To achieve this, the two-stage least square method was mainly used to analyze survey data collected from Portugal students.

There are four main findings regarding the research question of this study. Firstly, workday alcohol consumption generally has a significant negative association with students' academic performance, which implies that students who consume alcohol during the week may be at risk of lower academic achievement. Secondly, the negative impact of alcohol consumption is found to be most pronounced on math grades, potentially because math requires more cognitive focus and attention to detail than other subjects, making it more sensitive to the negative effects of alcohol consumption. Thirdly, students with mothers who have a higher level of education may be more susceptible to the negative effects of alcohol consumption on academic performance. One possible explanation for this result is that highly educated mothers may be more likely to pursue their own careers, which limits the time they can spend with their children. Since mothers are typically assumed to be the primary caregivers for their children's education, this ultimately could increase the likelihood of negative academic outcomes associated with alcohol consumption among their children. Finally, alcohol consumption among students who are below the legal drinking age of 18 has a significant negative impact on academic performance, underscoring the importance of addressing underage drinking as a serious issue that may hinder students' academic achievements.

However, the study also has some limitations that should be considered when interpreting the results. Firstly, the data used in this research was obtained from the survey, which may introduce selection bias and limit the sample size. For instance, the number of observations for certain subgroups, such as students aged above 18, was relatively small. This small sample size may lead to less accurate regression coefficient estimates and larger standard errors, which could reduce the statistical significance and make it more difficult to identify meaningful relationships. Secondly, some of the variables, such as the number of past failures, exhibit skewness in their distributions. This can lead to potential errors in the statistical analyses and interpretations of the results. Finally, the analysis may not be able to capture all relevant variables that may impact the outcome of interest. Therefore, to obtain a more comprehensive understanding of this topic, future research can benefit from incorporating additional features that may interact with alcohol consumption to impact academic performance.

Nevertheless, the research provides invaluable insights for policymakers and educators on the potential impact of workday alcohol consumption on academic performance and highlights the need for strict regulations on student drinking to enhance their academic achievements.

Reference

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, 30*(1), 1–15

Cortez, P., & Silva, A. (2008, January). *Using data mining to predict secondary school student performance*. ResearchGate. https://www.researchgate.net/publication/228780408_Using_data_mining_to_predict_secondary_school_student_performance

DeSimone, J. (2010). Drinking and academic performance in high school. *Applied Economics, 42*(12), 1481–1497. https://doi.org/10.1080/00036840701721554

Ho, D. S. Y., Huang, R., Wang, M. P., Lo, W. S., & Lam, T. H. (2019). Adolescent alcohol drinking in Hong Kong: a school-based survey. *Hong Kong Medical Journal, 25 Suppl 3*(1), 13–15.

Honghao, J., Po, Y., & Tianyu, Y. (2021). The influence of adolescents' romantic relationship on individual development: Evidence from China. *International Journal of Chinese Education, 10*(3), 221258682110700–. https://doi.org/10.1177/22125868211070036

Idris, M., Hussain, S., & Ahmad, N. (2020). Relationship between Parents' Education and their children's academic achievement: Parent's education and their children academic achievement. *Journal of Arts and Social Sciences (Lahore, Pakistan), 7*(2), 82–92. https://doi.org/10.46662/jass-vol7-iss2-2020(82-92)

Lees, B., Meredith, L. R., Kirkland, A. E., Bryant, B. E., & Squeglia, L. M. (2020). Effect of alcohol use on the adolescent brain and behavior. *Pharmacology, Biochemistry and Behavior, 192*, 172906. https://doi.org/10.1016/j.pbb.2020.172906