**Effect of Alcohol Use on Academic Performance of School Students**

Prashant Mittal

**Abstract**

The research is a part of an observational study that examines the effects of several socio-economic factors on the grades of secondary school students. The primary area of interest for this paper is to see the effect of alcohol on student’s final grade and to select other significant factors affecting the grades. The data was collected from Portuguese class of two secondary schools as part of a survey. The result shows that alcohol consumption on workdays affects students’ final grades whereas alcohol consumption on weekend has no significant effect on final grades. Further research, including a larger and more representative sample of a true population, would improve the findings.

**Introduction**

In United States, one in four individuals between the ages of 12 and 20 drinks alcohol on a monthly basis, and the ratio changes to five or more drinks every two weeks for a similar proportion of 12th graders. Alcohol consumption among students is generally viewed as a socially important topic and is a reason of concern among many parents and school administrators. In the past some of the researches have been conducted that has shown alcohol consumption negatively affects students as they are not able to balance out their academic work and social life causing potential dropouts and failing classes. On the contrary some researches have shown that drinking alcohol has no significant effect on the student’s performance. So, the objective of this research is to successfully learn if drinking alcohol has any effect on student’s final grades or not and to find out other significant factors that helps in explaining grades of a student.

**Related Work**

Walid El Ansari, Christiane Stock and Claire Mills (2012) examined the association between alcohol consumption and academic achievement among university students. The author employed five alcohol consumption measures (length of time of and amount consumed during most recent drinking occasion, frequency of alcohol consumption, heavy episodic drinking, problem drinking); and an educational achievement indicator i.e. students’ actual module mark in their study. They concluded that actual module mark was not associated with any alcohol consumption measure.

Butler et al. (2011) examined the relationship between drinking and academic demands. “National surveys indicate that 40% of college students binge drink regularly, and the rate of binge drinking appears to be increasing”. The results of this study suggest that students drank 38% less when they had an exam or paper due the next day. Results also indicated that a student’s current day academic demands are related to a 30% decrease in their current day alcohol consumption. This research demonstrates that some students are not reducing their alcohol intake even when they have academic demands to meet the next day.

Ana I. Balsa, Laura M. Giuliano and Michael T. French (2011) examined the effects of alcohol use on academic achievement in high school students’ quality of learning. Their study concluded that increase in alcohol consumption results in small yet statistically significant reductions in GPA for male students and statistically insignificant changes for female students.

Singleton and Wolfston (2009) examined alcohol consumption, amount of sleep, and academic performance. The authors showed the relationship between alcohol use and sleep, alcohol use and academic performance, and sleep and academic performance. The students from North-eastern university were 89% white, 98% ranged between the ages of 18-22 years, and 82% lived on campus. They concluded that students who drink more alcohol maintain poor sleep patterns, which negatively affects academic performance.

Aertgeerts and Buntinx (2002) examined the relationship of alcohol abuse among first-year students and their academic performance. Irresponsible drinking patterns among college students in the United States range between 7% and 17%. This study was conducted using a questionnaire given to 3518 first-year students who attended a specific college. They found that 14.2% of the students were identified as having alcohol abuse or dependence, and nearly one-third of the students passed their exams on the first attempt. Results suggest that 49.7% of male students and 48.9% of female students went on to continue their second year of school. It is potentially concerning that only about half of the freshman class succeeded and moved onto their sophomore year. This may indicate that alcohol consumption is a major social issue among college students.

Most of the research articles cited above found a negative relationship between alcohol use and academic performance. But they all had a major limitation of considering alcohol as a single entity without observing its pattern of consumption i.e. whether it is being consumed casually over the weekends for a party or is consumed as a daily habit.

**Problem Statement**

The main aim of this research is to find out if drinking alcohol (whether in weekdays or weekend) has any impact on secondary school students’ final grades or not. And to find out the best set of significant explanatory variables that can help in explaining student’s final grades.

**Research Methodology**

The primary research on the subject is done by Fabio Pagnotta, Hossain Mohammad Amran from University of Camerino by conducting a survey on Portuguese class students of two different schools. Around 650 students participated in the survey and students were asked to provide response against twenty-three questions without telling them the exact purpose of the research. Most of the questions asked in the survey were in the form of yes or no response for the ease of the students. The results were consolidated and stored in a csv file at the UCI database to cater future researches. The major benefit of using this type of data selection technique is low cost and time efficiency.

**Hypotheses**

As per the socio-economic theory alcohol consumption has a negative effect on the performance of the school students i.e. their GPA is lower than the students who don’t consume alcohol. So, in the current research null hypothesis is that drinking alcohol whether on weekday or weekend has no effect on the student’s grades. The alternative hypothesis is that drinking alcohol either on weekdays or weekends or both has negative effect on the grades. In the model, prominent independent variables are workday alcohol consumption and weekend alcohol consumption. The dependent variable is student’s final grades.

H0: workday and weekend alcohol consumption has no effect on grades.

H1: at least one of them has negative effect on grades.

**Solution**

The solution to this problem could be broken into two parts. First is to select list of significant explanatory variables that helps in estimating final grades. Multiple linear regression can be used for this purpose to create an initial model with all the independent variables and then forward selection or backward elimination could be used to find out most significant list of explanatory variables using both AIC & BIC penalties. Interaction terms can be added to this model as per the economic theory and final model could be tested against AIC, BIC & Cp statistics to see which model performs better. A lower value of these statistics suggests better model. The end model created in this step is termed as reduced model.

Secondly, weekday and weekend alcohol consumption variables can be added to the reduced model from part one and an ANOVA test could be used on the full model and the reduced model to see if any of the two or both alcohol consumption variables are significant for the model or not. This will provide us an insight if alcohol consumption has any effect on student’s grades or not.

**Analysis**

The sample includes 22 independent variables and 1 dependent variable. Variables with their description is present in Appendix table 1.

**Sample Characteristics**

*Independent Variables*

Descriptive data exploration shows that the sample contains approx. 41% male and 59% female students, out of which about 37% are involved in romantic relationship and 63% are not. About 70% of the students lives in Urban area and 30% in rural area. Age variable shows a peculiar trend as about 99% of the data lies in the range of 15 to 19 years and only about 1% students lies in the age group of 20 to 22 years. So, the variable requires further investigation. Table 1 provides a tabular description of some of the prominent explanatory variables.

*Table 1*

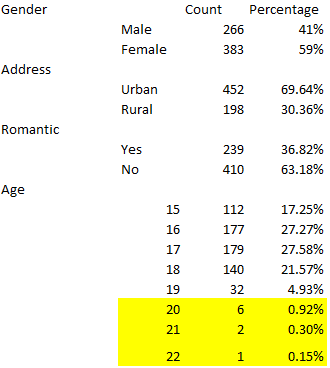
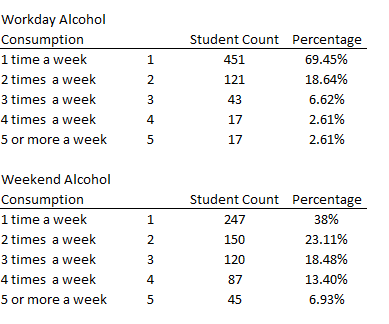
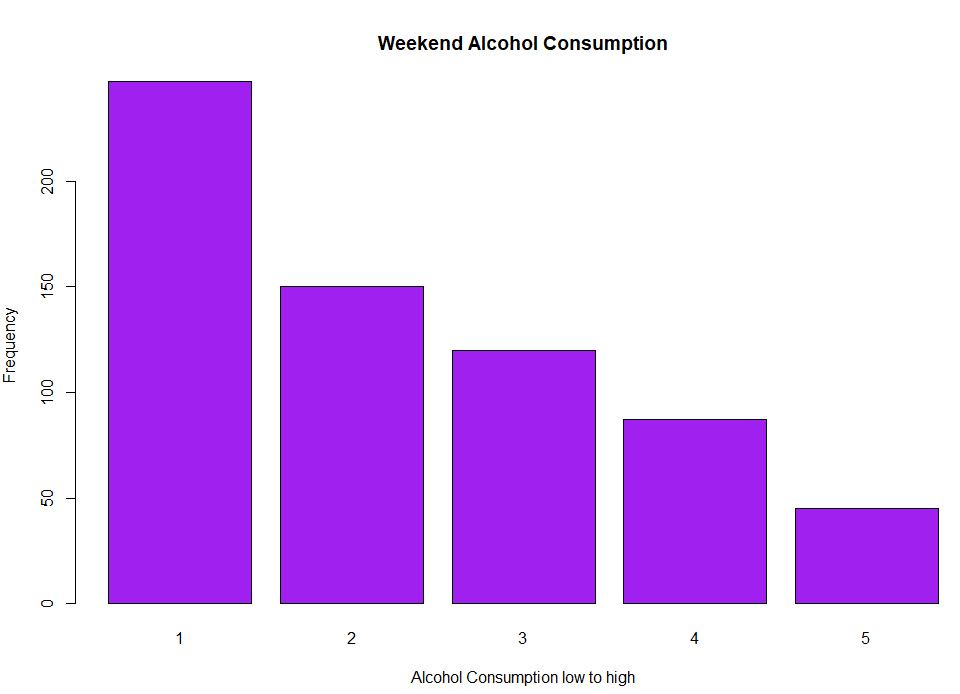
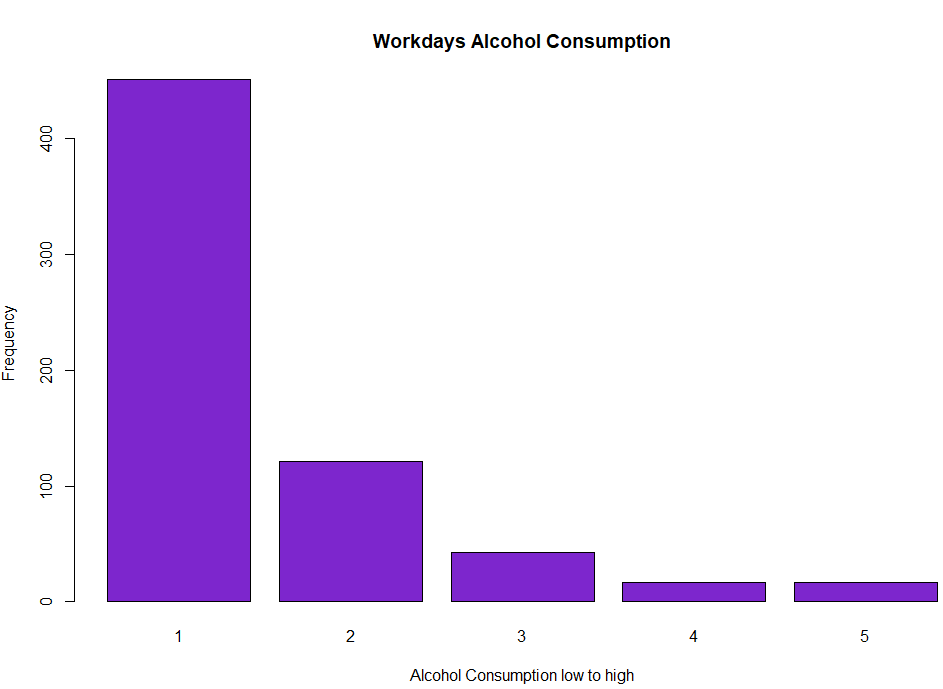


Table 2 and figure 1 provides a tabular and graphical description of the workday and weekend alcohol consumption respectively. It can be seen from the both that nearly 70% of the students consume alcohol only once in the weekdays and about 61% students consume alcohol once or twice on a weekend.

Table 2





Figure

*Dependent Variable*

Figure 2 provides a graphical view of frequency distribution of the dependent variable Final Grades and it can be observed that maximum students have scored in the range of 50% to 60% marks and the graph appears to be normally distributed.

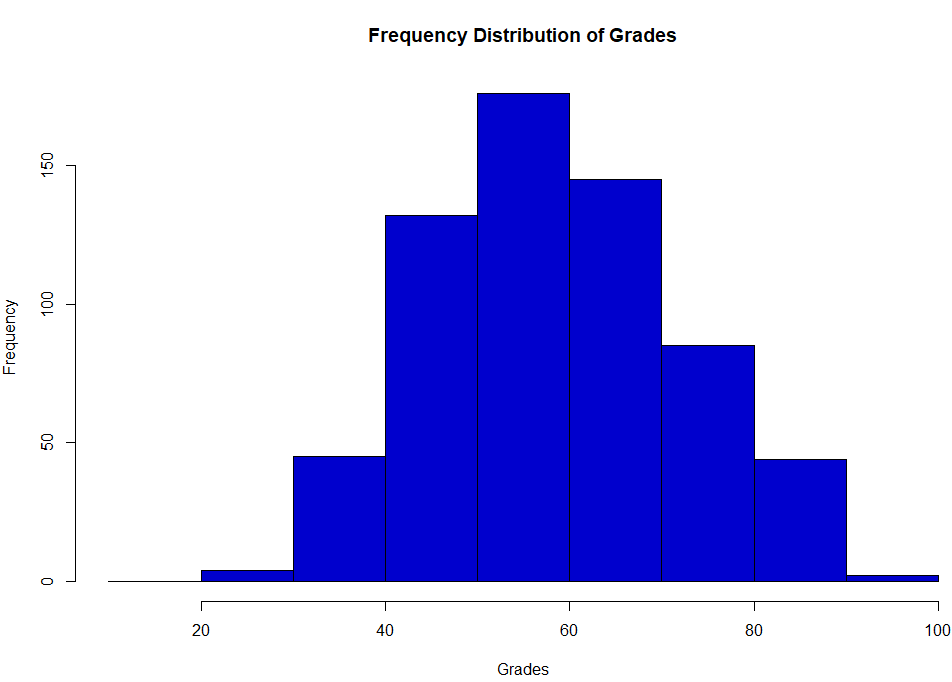


Figure 2

**Variable Transformation**

The dataset presented here is mostly cleaned with no missing values. Most of the variables including dependent variable is normally distributed however there are few independent variables in the dataset that needs transformation in the form of outlier removal or skewness removal treatment.

*“Absence” Variable*

From the visual inspection of the variable it can be seen that the variable is positively skewed as the number of absences of most of the students were very less. So, in order to eradicate skewness lograthimic transformation of the dependent variable is used. Figure 3 provides a before transformation frequency distribution (red graph) and after transformation frequency distribution of absence variable (green graph). After transformation data appears to be approximately normal now.

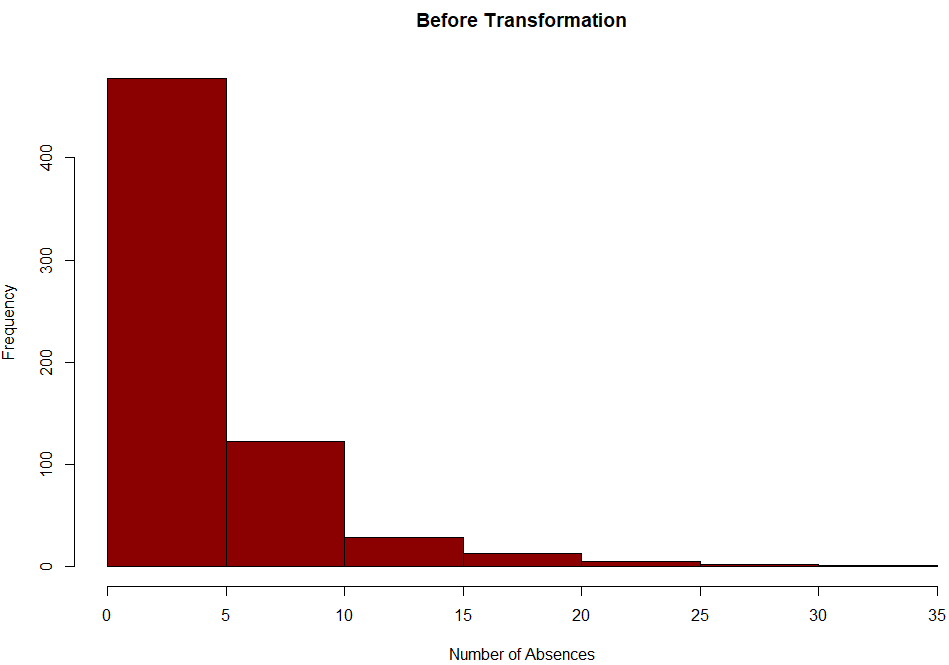
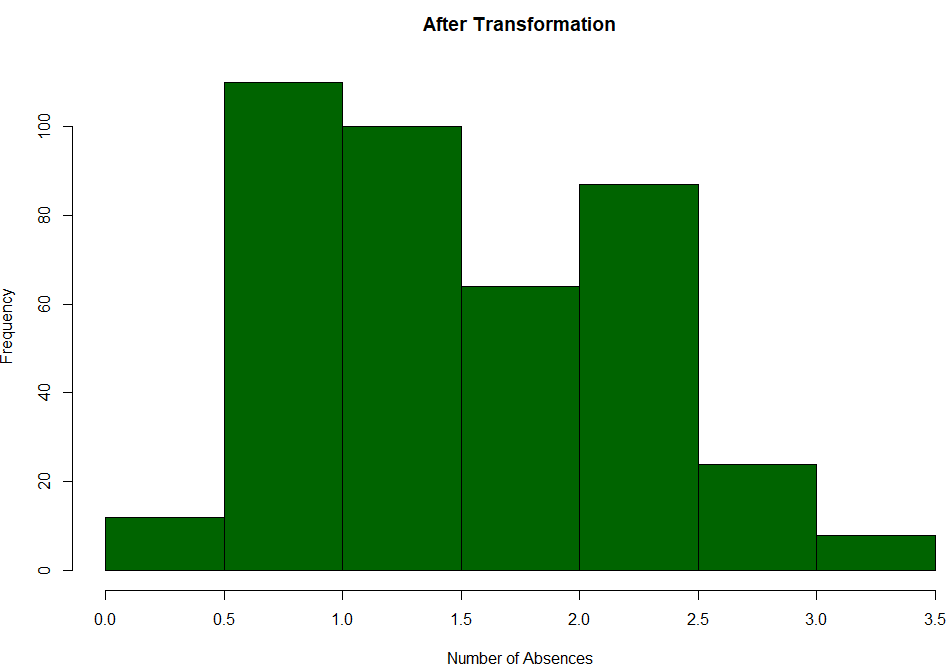
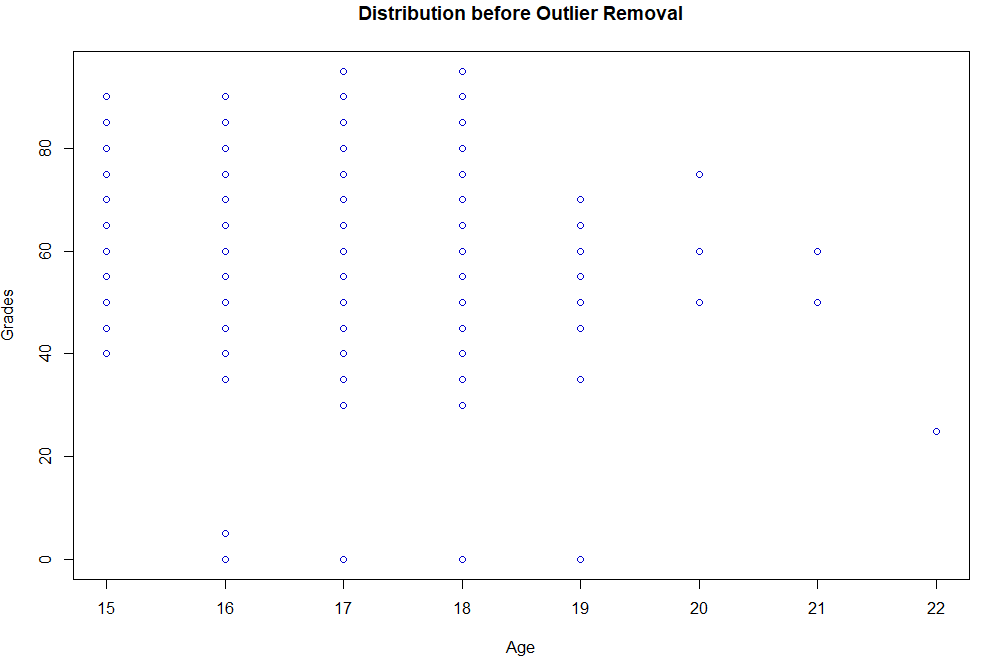
 

Figure 3

*“Age” Variable*

Figure 4 shows the distribution of final grades as per the age variable and it can be observed that only one student out of 649 has an age of 22 years, two students have an age of 21 years and only three students have an age of 20 years. As per economic theory ideally students in secondary school (up to 12th standard) have a maximum age of 19 years and the data below even suggests the same as only 1% students fall in the band of 20 to 22 years. There is a possibility that some of the students might have provided incorrect information about their age on the survey. So, treating these observations as outlier and removing them from the data set would be an ideal choice. Figure 5 shows the outlier treated final variable.

Figure 4



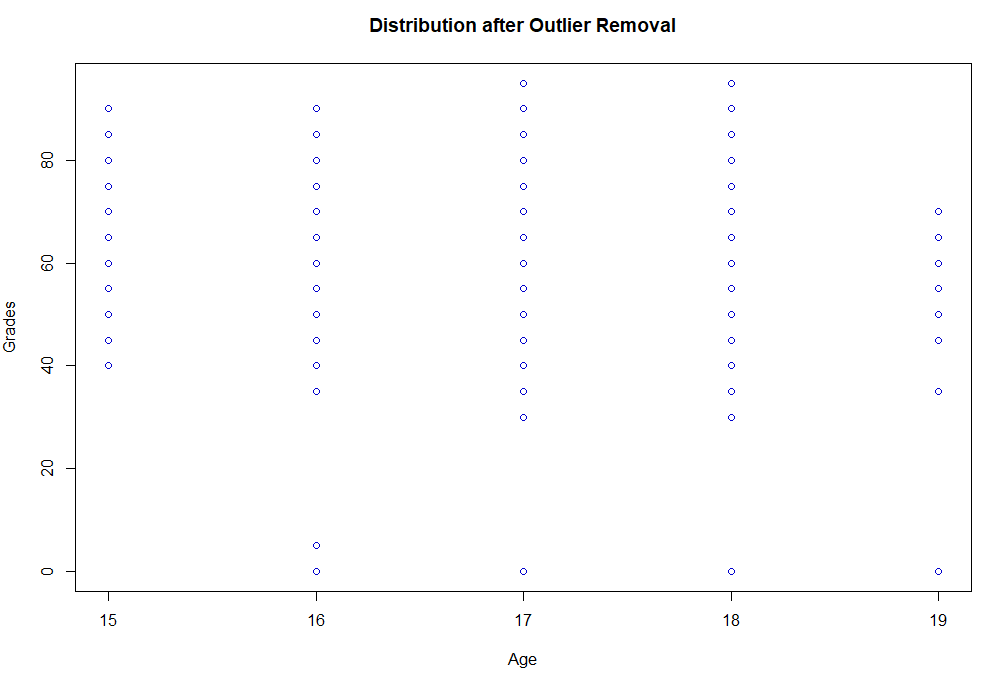


Figure 5

*Alcohol Vs Grades*

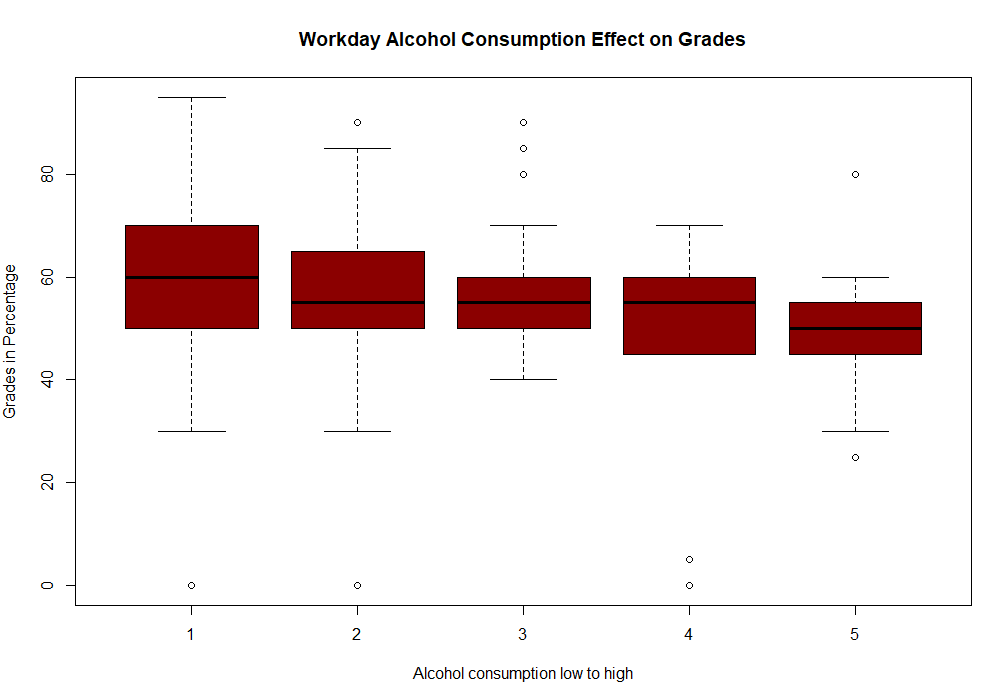
Figure 6 shows the box plots of final grades against the explanatory variable workday alcohol consumption and weekend alcohol consumption on a scale of 1 to 5 where one means very low alcohol consumption and five means very high consumption. Visual inspection from both the graphs tells that median grades of students decreases slightly when consumption increases from one to five drinks in workdays whereas median grades doesn’t seem to be affected by increase in consumption during the weekend as median grades are almost constant. 



Figure 6

**Model Building**

*Variable Selection*

In order to select significant explanatory variables for the model, forward selection and backward elimination processes are being used with AIC and BIC penalties. In case of a conflict in selection, BIC penalty would be preferred as it penalizes the model to the highest.

Figure 7 shows the AIC, BIC & Cp statistics of final models selected by forward selection and backward elimination processes using AIC & BIC penalties. Backward elimination with AIC penalty selected a model with 15 explanatory variables with R^2 as 29% whereas backward elimination with BIC penalty selected model with only 6 significant variables with R^2 as about 26%. Also, forward selection and backward elimination with BIC penalties selected the same model and the selected model has the lowest BIC value amongst all the models. So, as per Occam’s razor principle selecting the model with least number of explanatory variables as the final model for analysis as addition of 9 explanatory variables is increasing the R^2 of the model by only 3%.

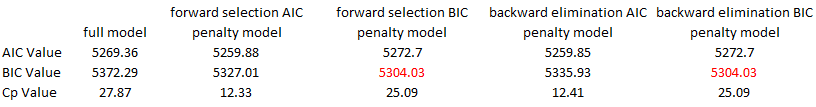


Figure 7

*Interaction Term Addition*

As per economic theory it is believed that addition of some of the interaction terms might help in explaining the model better like a student who goes out a lot and has a romantic relationship might have lower grades compared to a student who just goes out more or has a romantic relationship. So, repeating the backward elimination process with all the possible interaction terms to obtain the final model. Figure 8 shows the R output of a linear regression executed against the final list of selected variables and the BIC statistics. For future references this model will be termed as reduced model. From BIC statistics we can see that BIC value of the model has not increased much but R^2 of the model has increased approximately by 2%. Both the added interaction terms are significant for the model at alpha equal 1%.

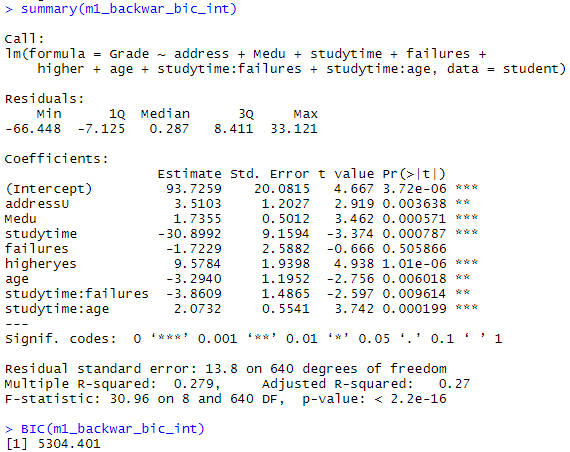


Figure 8

*Full Model*

Full model is prepared by running multiple linear regression on selected set of input variables present in reduced model along with alcohol consumption variables (*Walc*, *Dalc*). Figure 9 shows the R output of the full model run and it can be seen that R^2 of the full model increased from 27.8% to 29.53% when compared with the reduced model.

T-test on *Dalc* & *Walc* suggests the coefficient of Dalc is significantly different from zero at alpha equal 5% whereas coefficient of Walc may not be significantly different from zero even at alpha equal 10%.

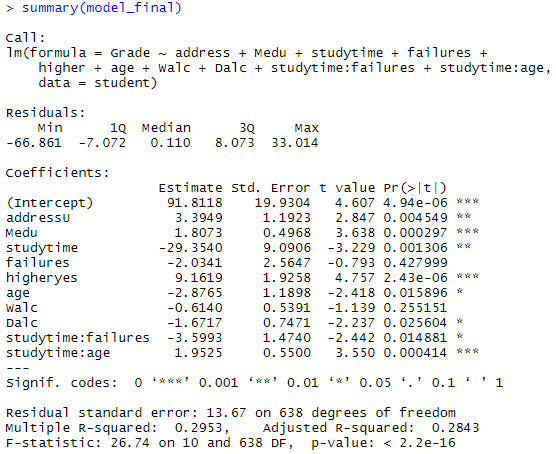


Figure 9

*Model Comparison: ANOVA*

In order to see if alcohol consumption variables (*Walc* & *Dalc*) has any significant effect on the student grades or not, an ANOVA test is conducted. Null hypothesis for the test would be both variables are insignificant for the model and alternate hypothesis would be at least one of them or both are significant for the model i.e.

H0: coefficients of *Walc* and *Dalc* are equal to zero.

H1: coefficient of at least of them not equal to zero.

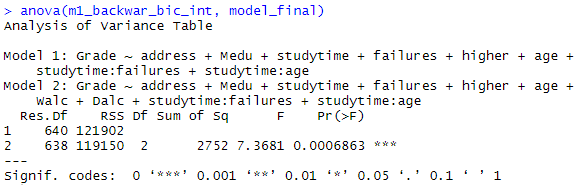


Figure 10

Figure 10 shows the R output of an ANOVA test. It can be observed that p-value is very less i.e. 0.00068 so we reject the null hypothesis and conclude that at least one of the alcohol variables or both are significant for the model.

**Conclusions**

The analysis provides overwhelming evidence that alcohol plays a significant effect in determining the final grades of secondary school students (F-value 7.36). Weekend alcohol consumption has no significant effect on grades (one sided p-value 0.2551) whereas weekday alcohol consumption has statistically significant negative effect on the final grades (one sided p-value 0.0256) i.e. 1 unit increase in alcohol consumption decreases the grades by 1.67%.

**Scope of Inference**

Since this is an observational study a causal inference cannot be established. The sample here is not the representative of the true population, as the participants were selected only from the Portuguese language class of two schools. Hence inference to population cannot be established here. However, findings of this research are consistent with the economic theory that alcohol has negative effect on the students grades but more research is required in this area to prove that it’s consumption in limited quantity as a recreational activity may not affect the grades significantly.

**References**

1. Claire Mills. 2013. “Is Alcohol Consumption Associated with Poor Academic Achievement in University Students?” *International Journal of Preventive Medicine*.
2. Ana I. Balsa. 2011. “The effects of alcohol use on academic achievement in high school”. *HHS Author Manuscripts*
3. Jill Coyman. 2002. “The Effects of Alcohol Use on Academic Performance Among College Students”. <https://www.semanticscholar.org/>

**Appendix**

*Variable definitions*

Table 1 Variable Overview

