What makes students to drink? An Analysis on Student Alcohol Consumption data

Rishabh Jain (rj19aao@herts.ac.uk), Neha Barve (nb19aap@herts.ac.uk),

Atik Jain (aj19abe@herts.ac.uk), Harshavardhan Vidap (hv19aai@herts.ac.uk)

and Akhil Kishore Dundi (ad19acv@herts.ac.uk)

University of Hertfordshire, United Kingdom

Abstract: Nowadays early age alcohol consumption has become very common amongst students, starting from school life. There can be multiple reasons for alcohol consumption of students, which is also seen in the data set. This report describes a detailed analysis on the factors affecting the alcohol consumption among the students of two different schools. The samples are collected from the students of Math and Portuguese language courses in secondary school. After supervising a series of tests on the data, we have found astonishing outcomes. We lead our studies to find out various factors responsible, which affects student's life in different aspects such as grades, failures, etc. Our findings revealed a lot more than anticipated. Our analysis gave rise to striking results.

Keywords: walc, dalc, Pstatus, freetime, gout, Pearson's R method, T-test.

1. Introduction:

In this report a dataset known as "Student Alcohol Consumption" has been studied. The variables in the following dataset are interesting which will produce an unexpected outcome when combined. There are two Independent variables (Walc and Dalc) which are representing the data about the level of alcohol that is consumed by the students. Walc values has been considered during the data visualization and analysis. The dataset will be further analyzed to answer few questions which were raised during the research. Initially, the entire sample is considered as a single population, as a part of experimenting with the given data. New condition has been introduced to analyze the data by dividing the entire population into two groups, based on their sex. First group is for male and second one is for females. T- test has been performed for statistical analysis of data and the variables were correlated using Pearson's R correlation method. During data visualization different plotting techniques has been generated such as Box Plots, Scatter Plots, Histogram and Bar Plots.

2. Background:

2.1 About data set

The drinking issues start from early secondary level education of any individual and it has become an icon of social standard. But we must understand that this has adverse effects on student's career and social life rather than being the source of enjoyment and fun. Consuming alcohol at teenage reduces the child's mental and physical abilities as well as affects coordination which can lead to trouble.[1] Children who start drinking by the age of 13 are more likely to go on to have worse grades, skip school and, in the worst case scenario, be excluded from school[3]. But to understand the problem of such mass deviation of students toward alcohol we need to understand the reason and the aspects behind it. To understand these reasons, we have taken the data from the surveys of Math and Portuguese language students in secondary school. The dataset used in our research was uploaded on Kaggle. The data set for this study is obtained from the survey conducted in two different secondary schools namely Gabriel Pereira or Mousinho da Silveira. It contains a lot of interesting aspects and factors about students. The data obtained in survey contains 649 data points compared with 33 different entities mainly on: [4]

- school - student's school ('GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
- sex - student's gender ('F' - female or 'M' - male)
- age - student's age (15 - 22)
- Pstatus – Parent's status ('T' - living together or 'A' - apart)
- studytime - weekly study time (1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
- failures - number of past class failures (n if 1<=n<3, else 4)
- freetime - free time after school (1 - very low to 5 - very high)
- Walc - weekend alcohol consumption (1 - very low to 5 - very high)
- health - current health status (1 - very bad to 5 - very good)
- G3 - final grade (0 to 20)

2.2 Research Questions:

When we introduced different cases and parameters, our research expanded its bounds and we had come up with two major research questions to answer.

1. What are the major factors causing alcohol consumption in students? (Refer section 3 for explanation on method studied and used)
    1.1. Can it be determined which gender consumes higher level of alcohol?
    1.2. Does Study Time affect the drinking habit of student?
    1.3. Does going out affect the alcohol consumption of students?
2. Is alcohol consumption effecting the grades of students? (See section 4 for explanation and reviews)

3. Method:

To answer our research questions, we used two different methods to analyze the data. We have supervised 'T-method' for statistical analysis of the data and 'Pearson's R method' to obtain the correlation between the data. We even conducted various computations using the available data and discovered impressive results. We have led our computation process using "R", "SPSS". To find the rate of alcohol consumption with respect to different factors, we have applied different conditions and observed the subsequent outcome. The IBM SPSS Statistics is used to visualize data based on 33 different entities of the dataset and we have obtained scattered plots, bar graph and pie chart.

Correlations

		studytime	goout	G3	Walc
studytime	Pearson Correlation	1	-.075	.250**	-.215**
	Sig. (2-tailed)		.055	.000	.000
	N	649	649	649	649
goout	Pearson Correlation	-.075	1	-.088*	.389**
	Sig. (2-tailed)	.055		.026	.000
	N	649	649	649	649
G3	Pearson Correlation	.250**	-.088*	1	-.177**
	Sig. (2-tailed)	.000	.026		.000
	N	649	649	649	649
Walc	Pearson Correlation	-.215**	.389**	-.177**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	649	649	649	649

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 1. Correlations by SPSS

4. Result:

Figure 1 shows the correlation values of different entities in the data set performed using SPSS. We can see that tables shows which entity has better correlation with others**.**

Our research started by questioning the relations associated between the alcohol consumption and the grades of the students. We have analyzed the data using 'SPSS' and 'R'. The histogram below was used to investigate the percentage of alcohol consumption in a week, to help in understanding the alcohol consumption of students in percentage.

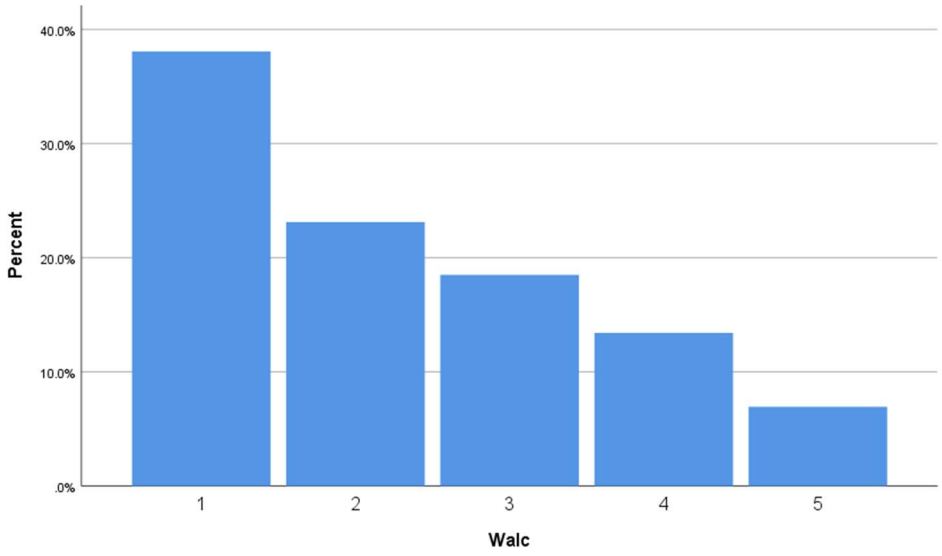


Figure 2. Weekly Alcohol Consumption Histogram

We plotted graphs using the data and separated the results according to the sexuality of students to find out the rate of alcohol consumption between them. By observing our computations, we found out that male students are more likely to consume higher amount of alcohol than the female students.

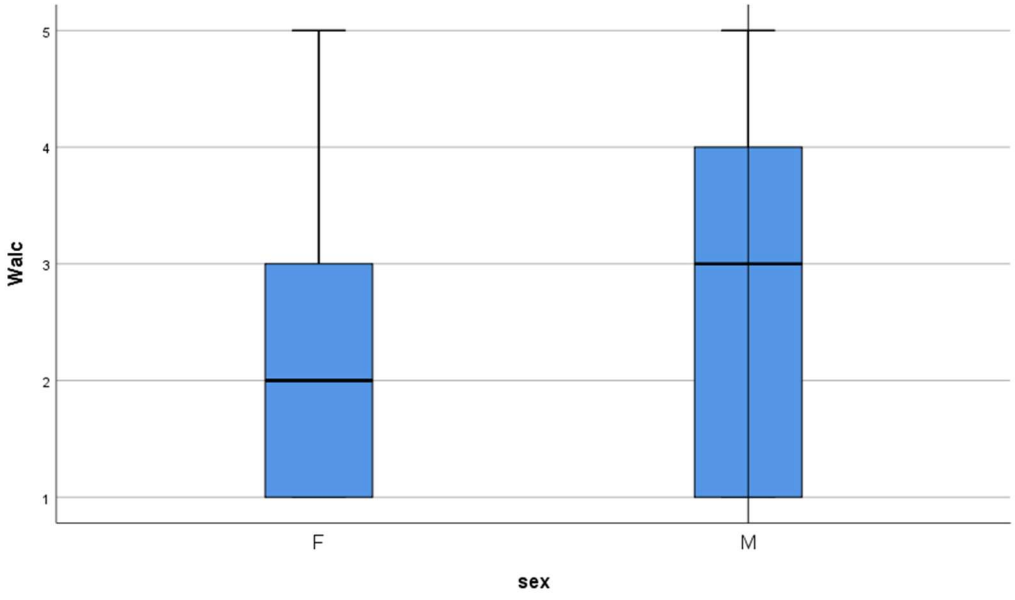
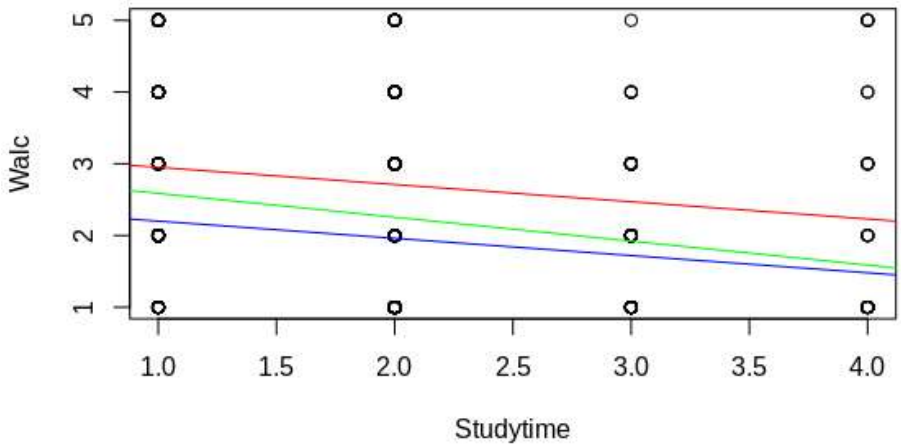


Figure 3. Boxplot between male and female for walc

We can clearly observe that the median value of male students is higher than female students, which leads us to conclude that consumption of alcohol is higher in male students. The above concluded statement has been also proved using T-test.

So, if alcohol plays a major role in influencing the grades of the students. Then the male students should be the most affected one. To illustrate this concept, we plotted a graph between the alcohol consumption and the study-time of the students



.

Figure 4. Scatter plot between walc and study-time

The plot discusses the prominence of alcohol consumption level of students with respect to the time utilized for studying. The colors ‘Red’, ‘Blue’ and ‘Green’ represents male, female and total strength of the student's respectively. The plot clearly explains that the consumption of alcohol affects the study-time. Rise in alcohol consumption leads to a fall in study-time. So, we found out that sexuality affect the consumption of alcohol as female students do not tend to consume much alcohol as per our studies.

Our research led to another major factor affecting the consumption of alcohol. By analyzing the samples, we found that ‘going out’ also plays a crucial role in alcohol consumption. It all depends on the frequency of the act. If a student leaves out frequently, higher are the chances that he/she consumes alcohol and vice-versa.

The plot drawn between alcohol consumption and ‘go-out' is shown above. The data is described by male (red line) and female (blue line) perspectives. We can clearly observe that as a student tends to go out, greater are the possibilities that he ends up getting drunk. This phenomenon is higher in the male students according to the studies performed using Pearson’s R method. Female students are less prone to this phenomenon.

We've approached “Pearson’s R” method to correlate the data and we found a strong correlation between these following factors. The final question emphasizes the way on how alcohol consumption is affecting the grades of the students. Achieving higher grades is the whole point in attending the school/university, we cannot just judge a student by his grade. But in this case, grades are the only thing that matters to analyze the data.

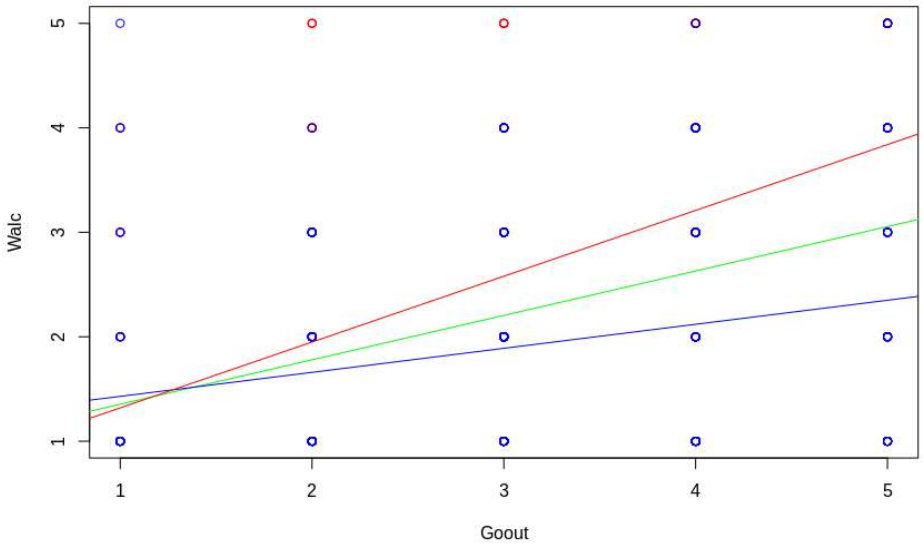


Figure 5. scatter plot between walc and goout

Grades totally depend on the extent of time spent on the studying; the parameter is termed 'study-time' in the data. The more a student spends time to study, higher would be the chances to get a better grade. So, by the studies we got to know that, alcohol consuming students mostly spend time on getting drunk rather than performing academic work. The higher the intake of alcohol, lower are the chances to get an exceptional grade. They are inversely proportional to each other. If a student increases the intake of alcohol, there will be a significant decrease in the study-time of the student and that reflects on his/her result and vice-versa. The following observations prove that alcohol can affect the grade of a student.

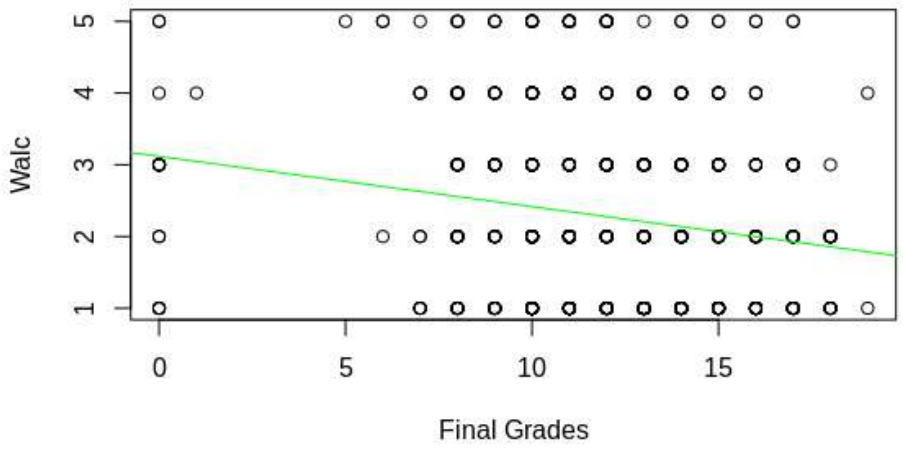


Figure 6. scatter plot between walc and final grades

The following scatter plot compares the alcohol consumption with the final grades of the students. We can observe a significant degradation in the final grades as the alcohol consumption rise. So, we can clearly state that alcohol shows a direct impact on the grades of the students. By the studies we've found that among all, male students were the most affected by alcohol consumption. So, most males achieved lower grades than the females. The computations and observations led to a conclusion that alcohol consumption affects the students and among all, male students were the highly influenced population by this circumstance. Our studies also found that, the students who are irregular to the school/university are most likely to be alcoholic. So, alcohol is absolutely a distraction and an essential reason for failure of students.

5. Discussion:

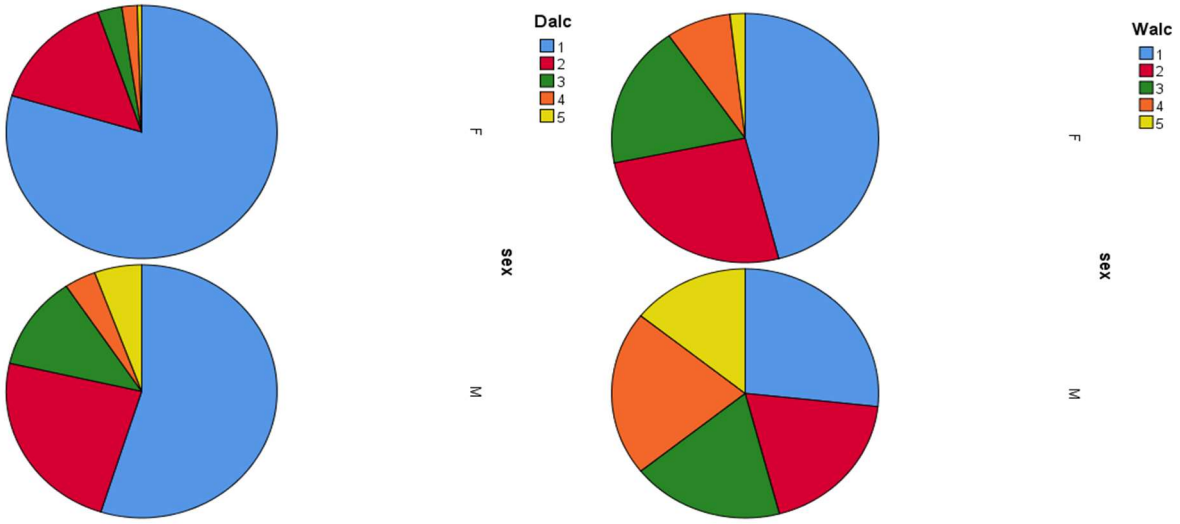


Figure 7. comparison of alcohol consumption observed in female and male students during daily and weekly basis

Above figure shows the comparison of alcohol consumption between male and female students during weekdays and weekends. We can observe that in the weekdays the alcohol consumption of the Females is low, the highest rate of drinking level is also very low and can be considered as negligible, but if we compare it with the male it is more than female and the amount of consumption of more alcohol is high, i.e. the quantity of the alcohol consumed by an individual in high amount is also observatory. Comparing this whole weekday figure with respect to the weekend we can clearly find that the difference between the alcohol consumption is high, people are drinking more alcohol, the graph rises in both male and female. We can see the level of alcohol consumption is increased and the level of less consumption is decreased, similarly in the females we can see that the high level of alcohol consumption is increased, and the low level of the alcohol consumption is decreased.

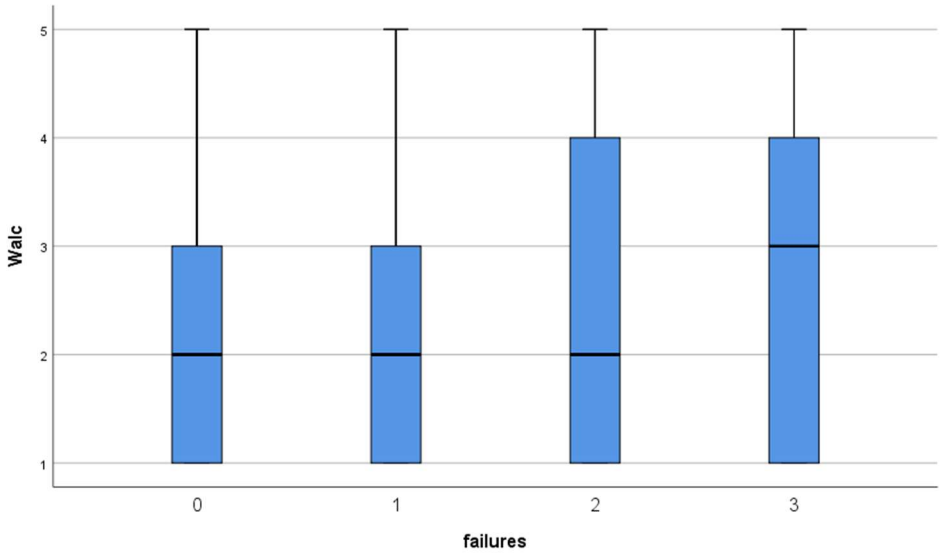


Figure 8. Boxplot for failures with walc

Boxplot above shows the number of failures of students consuming alcohols on weekends. The graph shows even if the student has zero failure or three failures, he/she is drinking alcohol but we can also see that student having zero or one failure drinks less as compared to students having more than one failure which concludes that number of failures is also one of the reasons for alcohol consumption by students.

We are discussing about the aspects influencing the alcohol consumption and the grades of the students, to answer our research question. During our research, we came across many interesting questions that, we thought may have affected the consumption of alcohol. They remained unanswered due to the lack of data to prove our point. In our view the alcoholic students possess ability to manipulate the non-alcoholic students and that would have been a major point for the course of our study, which is missing from data. But according to the given data, we have worked our best to provide most accurate results to our conclusions. This report emphasizes the abnormal condition of the school students by analyzing their alcohol consumption data. Most of the sample are too young to consume alcohol, this created an instability in the academics of the student.

The figure 9 shows the result for T test carried out on the dataset using R programming. The test had a p-value of 2.623e-15 which is less thus, the null hypothesis was rejected. An alternative hypothesis was also carried out for means.

T- test was performed based on the following hypothesis:

H0: The average alcohol consumption is same amongst male and female students of both schools.

H1: The average alcohol consumption differs amongst male and female students of both schools.

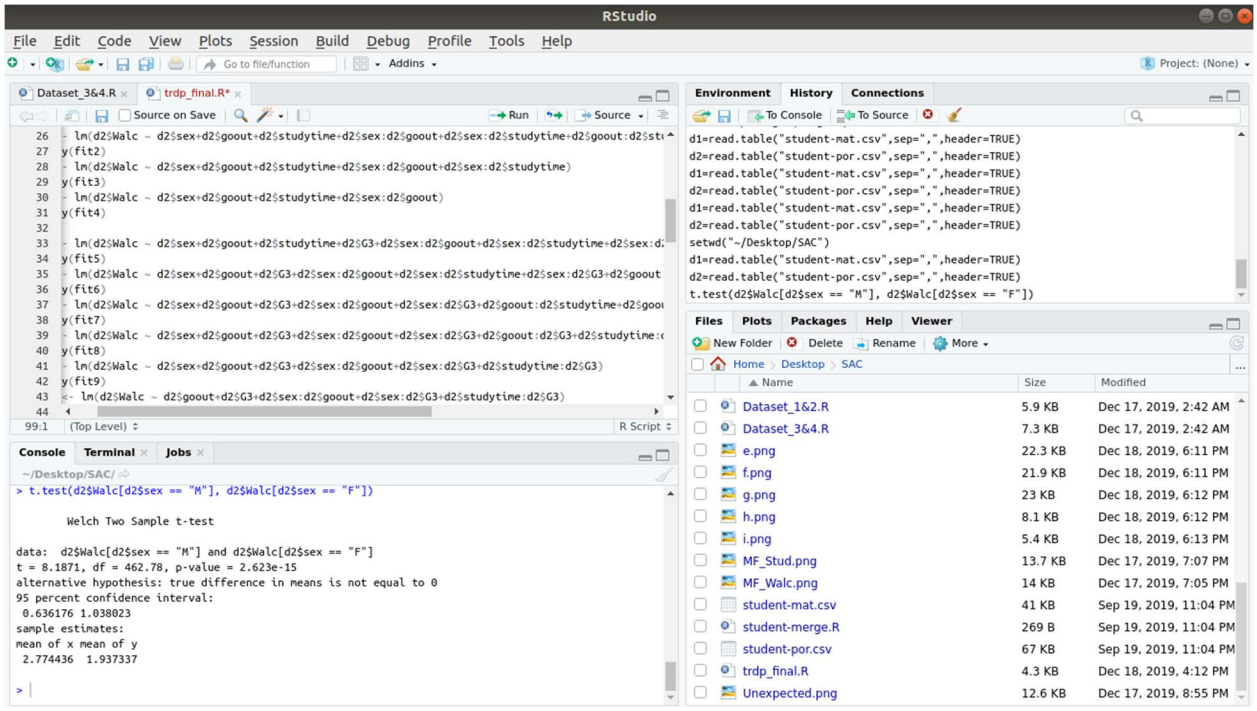


Figure 9. T Test Result

6. Conclusion:

In this paper, we analyzed alcohol consumption among students and established the factors that have a significant impact on alcohol consumption among students. It is found that men tend to consume more alcohol than women. The higher frequency of going out leads to higher alcohol consumption and vice-versa. The grades are significantly dependent on the alcohol consumption, as alcohol consumption of a student is influencing his/her study-time which directly affects the grades. The higher the consumption, the lower the grades to be expected. So, after all this research, we finally conclude that alcohol affects the grades and male students are highly affected due of their high consumption of alcohol, which is seen through various tests. Our research also shows that the male students drinks more on weekends as compared to weekdays than female students.

7. References:

[1] https://ieeexplore.ieee.org/document/8376297

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3843305/

[3] https://www.kaggle.com/uciml/student-alcohol-consumption

[4] Fabio Pagnotta, Hossain Amran, *Using Data Mining To Predict Secondary School Student Alcohol Consumption*, 2016.

[5] P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April 2008, EUROSIS, ISBN 978-9077381-39-7.

Appendix:

[1] Bit Bucket: https://bitbucket.org/cwgroup35/project-report/src/master/

[2] Trello: https://trello.com/b/cmaQfgGt/team-research-group-35