# Final Report

Contents

[Final Report 1](#_Toc61039524)

[Introduction 2](#_Toc61039525)

[Visualization 2](#_Toc61039526)

[Analysis 5](#_Toc61039527)

[Conclusion 8](#_Toc61039528)

[References 9](#_Toc61039529)

## Introduction

The research question was developed from a Kaggle dataset we selected to use titled ‘Student Alcohol Consumption Social, gender, and study data from secondary students’ (UCI-Machine-Learning, 2016). The data used in this report is collected from a class of math students in secondary school. The data collected varies from a wide range of attributes, specifically 33 attributes. The attributes chosen to be used in this research were ‘absences’ and ‘famrel’. ‘Absences’ is the number of absences reported and ‘famrel’ is the quality of the family relationship a student reports to have ranged from very bad to excellent. The research question extrapolated from this dataset was ‘Is there a difference in the mean number of school absences between students who report different quality of family relationships (very bad - excellent)?’.

The null hypothesis of the research question was that ‘there is no difference in the mean number of school absences between students who report different quality of family relationships (1 very bad - 5 excellent)’. The alternative hypothesis was that ‘there is a difference in the mean number of school absences between students who report different quality of family relationships (1 very bad - 5 excellent)’.

This is a particularly interesting topic as the current situation in the world has made it so that the pandemic has led to the majority of schools being closed to stop the spread of covid-19, as of writing this reports schools are expected to remain closed until mid-February with the possibility of that date being extended (Penna & Bird, 2021). Some schools in order to not miss out on learning have moved education online. Therefore, in this paper we strive to study weather potential family relationships at home could contribute to absences and so would overall likely affect education seeing how education is an incredibly hot topic right now it is a very interesting topic area, in a time where there are concerns of children falling behind in education.

Some studies showed that families involved with their children result in improved attendance more specifically that school, family, and community partnerships result in 0.5% improved attendance whereas if that was not the case attendance declined year by year (Sheldon B, 2010).

## Visualization

The boxplots are shown in figure 1 shows how data is distributed. The line through each box shows the median. Those who reported very bad and bad family relationships showed the highest measurement of the median both being the same as 5.5 number of absences. Those who reported excellent or average family relationships both had a median of 4.0 as shown on the boxplot and those who reported good family relationships are shown to have had a median of 3.0 as shown on the boxplot.

The lower quartile of the boxplot can be seen as decreasing as students self-report better quality relationships. Those who reported having very bad have a boxplot with a lower or 1st quartile of 3.5, those who reported bad have a boxplot with a lower quartile of 2.0, and those who reported average relations showed a lower or 1st quartile of 1.75 as shown in the plot in figure 1 and in table 1. Those reporting good and excellent relations 4 and 5 respectively have boxplots with a lower or 1st quartile of 0.

As with the lower quartile, a similar pattern is seen with the upper or 3rd quartile whereas students report better quality relations there is then a decrease in the upper or 3rd quartile of the box. Those reporting very bad and bad have an upper or 3rd quartile of 11 and 9.5, respectively. Those who reported average and good are shown to have the same upper or 3rd quartile of 8. Finally, those who reported excellent family relations had an upper or 3rd quartile of 7.75.

The final aspect of the boxplot is the whiskers that show the lowest and highest values that are still within 1.5 times of the interquartile range. For the lower whisker, all reported relationships have a lower whisker of 0. Those who reported good and excellent do not appear to have whiskers as the lower hinge of the 1st quartile is also 0, so there appear to be no whiskers. However, there are variations in the upper whisker where the value of the higher whisker increases as individuals report better quality family relationships. Those who report very bad have a higher whisker of 14, those who reported bad have a higher whisker of 15, those who report average relations have an upper whisker of 16 followed by an upper whisker of 20 in both those students who report good, and excellent relationship.

The black points on plots 1 and 2 show the outliers. Those reporting very bad and bad relationships have no outliers. Those reporting average relationships have 5 visible outliers close to the upper whisker. Those reporting good relationships had the most outliers spread the most above the upper whiskers mostly ranging from 20 to 60 with one outlier above 60 number of recorded absences. Those reporting excellent family relationships have two visible outliers in the *Boxplot*, one slightly above 20 the other above 50. On table 1 the highest data point can be shown as max on table 1.

In figure 1 the mean is indicated with a red ‘+’. In the means, there also seems to be a pattern where, as the quality of relations is increased the mean number of absences decreases with those reporting good break the pattern having a higher mean of 6.010256 where those with excellent have a mean of 4.952830 and those reporting average have 5.808824 mean. This may be as a result of those reporting good relations with their family have the highest number of outliers that vary the most along the y-axis. The values of the mean, and median from the boxplot and the values of the different quartiles can be interpreted from both the boxplot in figure 1 and the table in table 1 and 2.

Upon inspection of this boxplot those who reported poor family relationships (very bad and bad) had a less varied boxplot including the box and the whisker. Whereas the rest varied across the y-axis resulting in various outliers. This suggests those reporting poorer relationships have more consistent absences being lower.

|  |
| --- |
| Very bad Bad Average Good Excellent  Min. : 0.00 Min. : 0.0 Min. : 0.00 Min. : 0.00 Min. : 0.00  1st Qu.: 3.50 1st Qu.: 2.0 1st Qu.: 1.75 1st Qu.: 0.00 1st Qu.: 0.00  Median : 5.50 Median : 5.5 Median : 4.00 Median : 3.00 Median : 4.00  Mean : 6.88 Mean : 6.0 Mean : 5.81 Mean : 6.01 Mean : 4.95  3rd Qu.:11.00 3rd Qu.: 9.5 3rd Qu.: 8.00 3rd Qu.: 8.00 3rd Qu.: 7.75  Max. :14.00 Max. :15.0 Max. :25.00 Max. :75.00 Max. :56.00  NA's :387 NA's :377 NA's :327 NA's :200 NA's :289 |

Table : Descriptive statistics of the box plot

|  |
| --- |
| [,1] [,2] [,3] [,4] [,5]  [1,] 0.0 0.0 0.0 0 0  [2,] 3.0 2.0 1.5 0 0  [3,] 5.5 5.5 4.0 3 4  [4,] 12.0 10.0 8.0 8 8  [5,] 14.0 15.0 16.0 20 20 |

Table :stats of the boxplot raw data, the lower whisker, lower hinge of the boxplot, the median, upper hinge of the boxplot, and the upper whisker. Retrieved from the boxplot with the code ‘boxplot$stats’

The table above table 2 shows the raw stats with [,1], [,2], [,3], [,4], [,5] is each independent variable 1 very bad working up to 5 excellent. [1,], [2,], [3,], [4,], [5,] are the different data relating to the parts of the boxplot one is the lower whisker, two being lower hinge of the boxplot, three being the median, four upper hinge of the boxplot, five is the higher whisker.

Chart, box and whisker chart

Description automatically generated

Figure : Box plot showing the number of absences of students reporting different quality of family relationships

## Analysis

Those students reporting very bad and bad family relationships seem to have lower absences than those who reported average to an excellent relationship. Another noticeable point is that those reporting having better quality relationships had much more extremes and variations in absences leading to greater outliers.

As can be seen with the histograms in figures 2 through to 6 below none of the independent variables had a normal distribution, in fact, in every reported quality relation, the data is skewed towards the left. As a result of there not being a normal distribution, a Mann-Whitney test was used.

After carrying out the Mann-Whitney/Wilcoxon test using various independent variables it is revealed that the lowest p-value is the test carried out for those reporting bad and excellent relationship with the p-value being 0.1856. This does not fall under the 0.05 or 0.01 p-value meaning there is a higher chance of our result being caused by an error specifically an 18.56% chance of being caused by chance. In total 10 Mann-Whitney/Wilcoxon tests were carried out.

Chart, histogram

Description automatically generated

Figure : Histogram of student reporting very bad

Chart, histogram

Description automatically generated

Figure : Histogram of student reporting bad

Chart, histogram

Description automatically generated

Figure : Histogram of student reporting Average

Chart, histogram

Description automatically generated

Figure : Histogram of student reporting good

Chart, histogram

Description automatically generated

Figure : Histogram of student reporting excellent

## Conclusion

Because the p-value is larger than 0.05 the null hypothesis cannot be rejected, specifically from the Mann-Whitney test the lowest value was 0.1856 meaning there is an 18.5% chance the results are due to chance. Therefore, there is no difference in the mean number of school absences between students who report different quality of family relationships from very bad to excellent as a result it cannot be suggested that student number of absences is affected by the quality of relations students have with their family from this research and analysis.

From stereotypes, it would be expected that children who have poorer relationships with their family would likely have a higher number of absences however the visualisation shows the opposite where children with poorer relations had lower absences to those with relations with their family. Some research found that children of parents who abused alcohol had higher levels of parentification (Godsall, Jurkovic, Emshoff, Anderson, & Stanwyck, 2004) this could be used to support this opposing notion (assuming that due to the alcohol they had poorer relations) that those kids from their poor situation can do well.

One thing to note about the dataset is that there is no criterion provided by the dataset of how ‘famrel’ or reported family relationship is reported making it very subjective to each student.

# References

Godsall, R. E., Jurkovic, G. J., Emshoff, J., Anderson, L., & Stanwyck, D. (2004). Why Some Kids Do Well in Bad Situations: Relation of Parental Alcohol Misuse and Parentification to Children's Self-Concept. *Substance Use & Misuse*, 789-809.

Penna, D., & Bird, S. (2021, January 8). *School closures in Covid lockdown: latest news on when they could reopen*. Retrieved from Telegraph: https://www.telegraph.co.uk/news/2021/01/08/schools-closed-new-lockdown-when-reopen-uk-areas-covid/

Sheldon B, S. (2010). Improving Student Attendance With School,. *The Journal of Educational Research*, 267-265.

UCI-Machine-Learning. (2016, October 19). Student Alcohol Consumption.

[Table 1: Descriptive statistics of the box plot 3](#_Toc61039543)

[Table 2:stats of the boxplot raw data, the lower whisker, lower hinge of the boxplot, the median, upper hinge of the boxplot, and the upper whisker. Retrieved from the boxplot with the code ‘boxplot$stats’ 4](#_Toc61039544)

[Figure 1: Box plot showing the number of absences of students reporting different quality of family relationships 5](#_Toc61039537)

[Figure 2: Histogram of student reporting very bad 6](#_Toc61039538)

[Figure 3: Histogram of student reporting bad 6](#_Toc61039539)

[Figure 4: Histogram of student reporting Average 7](#_Toc61039540)

[Figure 5: Histogram of student reporting good 7](#_Toc61039541)

[Figure 6: Histogram of student reporting excellent 8](#_Toc61039542)