

Family, Gender, Alcohol, and Academic Performance: What're the Connections?

by Yen Nguyen & Nguyen Pham

Denison University, Professor Wang Zhe

Introduction

This project use statistical analysis to explore the stereotypes surrounding high school students' academic performance in Portuguese language courses. The dataset was obtained from Kaggle containing detailed information on students at Gabriel Pereira and Mousinho da Silveira secondary school. It has a total of 33 variables and 649 observations.

In this research, our independent variables are Sex, parents' education(parent_edu), weekly alcohol consumption (alc), and our dependent variable is Final grade (G3).

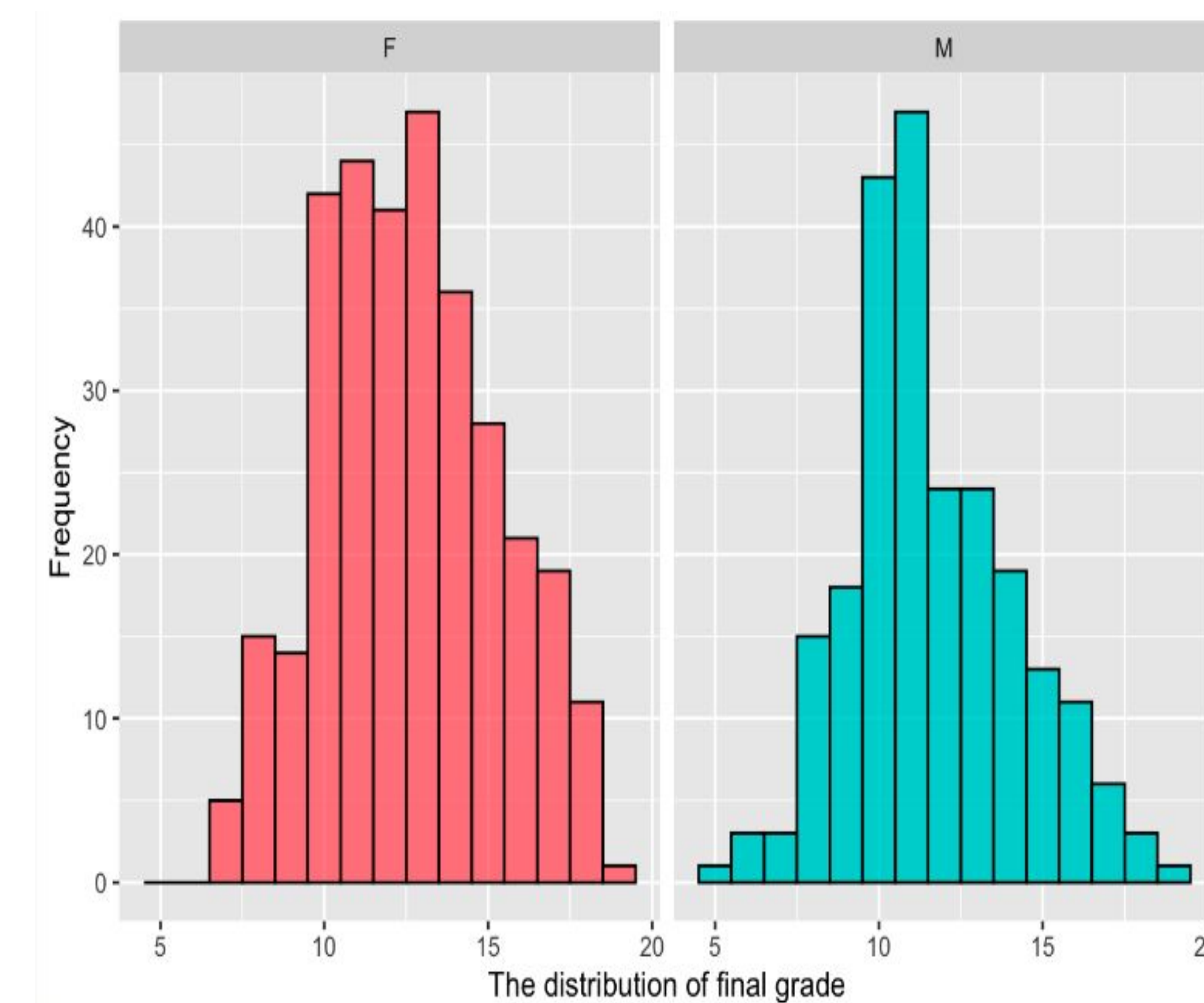


Objectives

- ❖ Explore the relationship between gender, family background, alcohol consumption, and academic performance.
- ❖ Examine how these factors impact students' final grades in Portuguese language courses.
- ❖ Identify risk elements for poor academic performance.
- ❖ Inform interventions, policies, and prevention efforts aimed at supporting students' academic achievement.

Methodology

Step 1: Check assumptions of t.test on sex groups



- Both groups have normally distributed data on the Portuguese final grades
- The assumption of normality in the data is met. We can conduct t-test on these groups.

Step 2: Use t-test to compare the mean final grade of two sex groups.

Welch Two Sample t-test

data: G3 by sex
t = 4.4278, df = 504.99, **p-value = 1.168e-05**
alternative hypothesis: true difference in means between group F and group M is not equal to 0
95 percent confidence interval:
0.5529005 1.4349142
sample estimates:
mean in group F mean in group M
12.61728 11.62338

Step 3: Perform multiple linear regression model to find the relationship among sex, family background, alcohol consumption level and students' final grade.

Call:
lm(formula = G3 ~ sex + alc + parent_edu, data = studentData)

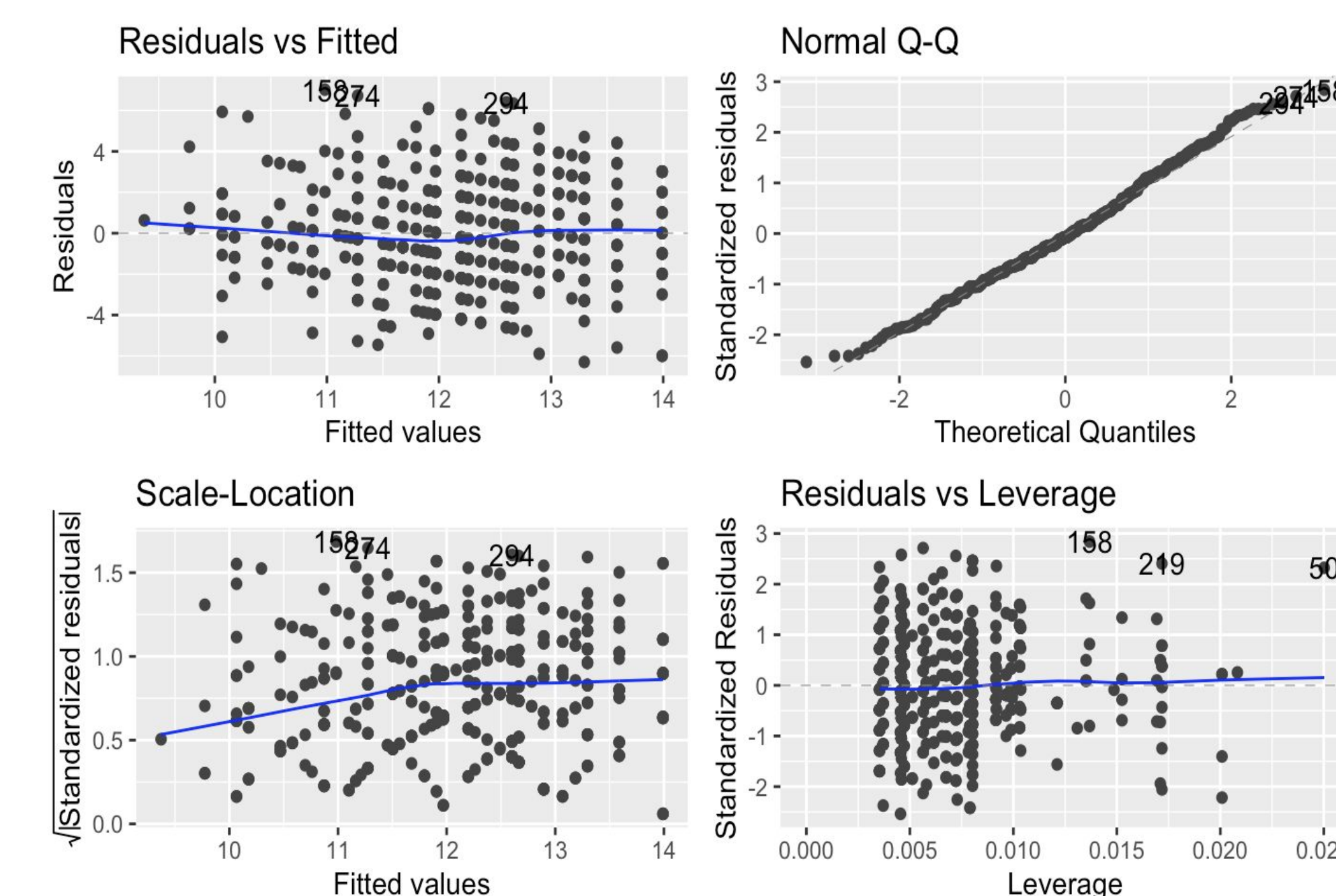
Residuals:
Min 1Q Median 3Q Max
-6.2963 -1.6643 -0.1855 1.5685 7.0173

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) **10.6907** 0.4251 25.148 < 2e-16 ***
sex **0.9239** 0.2274 4.063 **5.56e-05 *****
alc **-0.4028** 0.1067 -3.776 **0.000177 *****
parent_edu **0.6948** 0.1025 6.781 **3.07e-11 *****

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.486 on 551 degrees of freedom
Multiple R-squared: **0.134**, Adjusted R-squared: 0.1292
F-statistic: 28.41 on 3 and 551 DF, p-value: < 2.2e-16

Step 4: Check assumptions of the model



- The QQ plot shows a relatively straight line
- The Residuals vs Fitted plot shows a random scatter of points with no discernible pattern.
- The assumptions of homoscedasticity and normality for this multiple linear regression model are met.

Results

From the t.test:

- p-value < 0.05: The difference between the mean final grade of two groups is statistically significant
- 95% confidence interval does not include 0: 95% the true population mean difference is non-zero.
- Mean final grade: Female group > Male group

From the multiple linear regression model:

- p-value < 0.05: The relationship between students' final grade and sex, weekly alcohol consumption, and parents' education is statistically significant
- Grades have a negative correlation with alcohol consumption and a positive correlation with parents' education and sex (see the coefficients).
- Multiple R² = 0.134: The model can only account for 13.4% of the variance in final grades.
- Regression formula:

Final grade = 10.6907 + 0.9239 x sex - 0.4028 x alcohol consumption + 0.6948 x parents' education

Conclusion

- ❖ Female students are better at Portuguese than male counterparts, and this difference is statistically significant.
- ❖ Parents' education positively affects students' academic success in Portuguese language courses.
- ❖ Alcohol consumption is a risk factor for poor academic performance.
- ❖ Highlights the need for parental involvement, timely acts to buffer the negative effects of early alcohol on academic success.