Exploring the Impact of Parental Education on High School Students' Academic Success: An Integrated Analysis of Familial Influences

Ethan Alexander

Fall 2023

Abstract

This report investigates the relationship between parental education levels and the academic achievements of high school students. The study, conducted with a dataset encompassing 372 students from two schools in Portugal, utilizes a MANOVA model with parental education as the independent variable and students' grades in Portuguese and math classes as dependent variables. The overarching aim is to provide meaningful insights that contribute to the development of strategies fostering academic success.

Our analysis yields a significant correlation (p-value = 0.02332 < alpha = 0.05), offering robust evidence to conclude that parental education does impact academic success. This finding underscores the relevance of parental education as a critical factor influencing students' academic trajectories. The report emphasizes the importance of considering this association when crafting programs and policies aimed at fostering academic success. The observed link between higher levels of parental education and increased academic achievement among their children accentuates the pivotal role of familial influences in shaping educational outcomes. This insight is crucial for educators, policymakers, and stakeholders working towards the holistic development of students.

Introduction

High school is a critical period in a student's academic journey, marked by various factors that can significantly impact their success. This comprehensive statistical report seeks to explore the relationships between the academic achievements of high school students and various aspects of their familial dynamics. This report aims to provide insights that contribute to the development of strategies to foster the academic success of students. Understanding the relationship between family life and academic success can contribute to the formulation of support programs, optimal allocation of educational resources, and identification of at-risk students.

Data was collected from school reports and questionnaires involving 372 students across two high schools in Portugal. Level of parental education is categorized by number according to the following guidelines:

Rating	Description
0	No education
1	Up to 4th grade
2	Up to 9th grade
3	Secondary education (12th grade)
4	Higher Education

Academic success is quantified using the final grades of students in both their Portuguese and math classes. The grading system is as follows:

Scale	Grade Description	English Translation	USA Equivalent
20	Muito bom con distincao e louvor	Very good with distinction and honors	A+
18 - 19	Excelente	Excellent	A+
16 - 17	Muito Bom	Very Good	A
14 - 15	Bom	Good	B
10 - 13	Suficiente	Sufficient	C
7 - 9	Mediocre	Poor	F
0 - 6	Mau	Poor	F

Grade Distribution

Math
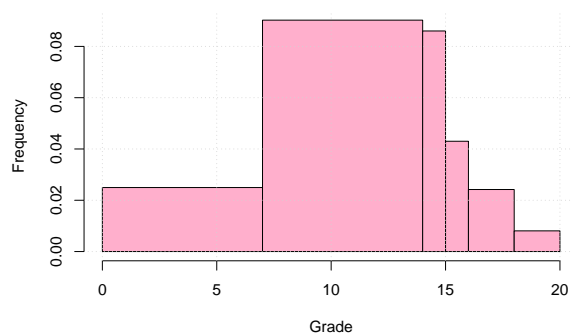
Figure 1: Math Grades Histogram

Grade	Frequency	Proportion
0	35	0.094
1	0	0.000
2	0	0.000
3	0	0.000
4	1	0.003
5	7	0.019
6	15	0.040
7	7	0.019
8	31	0.083
9	26	0.070
10	54	0.145
11	43	0.116
12	29	0.078
13	25	0.067
14	27	0.073
15	32	0.086
16	16	0.043
17	6	0.016
18	12	0.032
19	5	0.013
20	1	0.003

Table 1: Math Grades Frequency Table

Statistics	Values
Min.	0.000
1st Qu.	8.000
Median	11.000
Mean	10.465
3rd Qu.	14.000
Max.	20.000

Table 2: Math Grades Summary Statistics

Grade Distribution

Portuguese
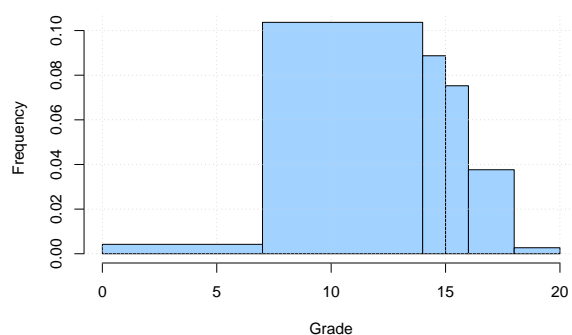
Figure 2: Portuguese Grades Histogram

Grade	Frequency	Proportion
0	5	0.013
1	1	0.003
2	0	0.000
3	0	0.000
4	0	0.000
5	1	0.003
6	2	0.005
7	2	0.005
8	7	0.019
9	11	0.030
10	45	0.121
11	54	0.145
12	54	0.145
13	61	0.164
14	38	0.102
15	33	0.089
16	28	0.075
17	19	0.051
18	9	0.024
19	2	0.005
20	0	0.000

Table 3: Portuguese Grades Frequency Table

Statistics	Values
Min.	0.000
1st Qu.	11.000
Median	13.000
Mean	12.554
3rd Qu.	14.000
Max.	19.000

Table 4: Portuguese Grades Summary Statistics

Factor Reduction and Variable Analysis

The dataset I based this observational study off of contains information about parental education in two statistics: mother's education and father's education. Due to the nature of my hypothesis viewing parental education as a collective, I considered the option to consolidate the two variables.

To make this decision, I wanted to analyze the relationship between mother education and father education to make sure consolidating data wouldn't mask any important information. First, I generated bar charts illustrating the distribution of education levels for mothers, fathers, and both parents combined. All three charts exhibited similarities in their patterns.
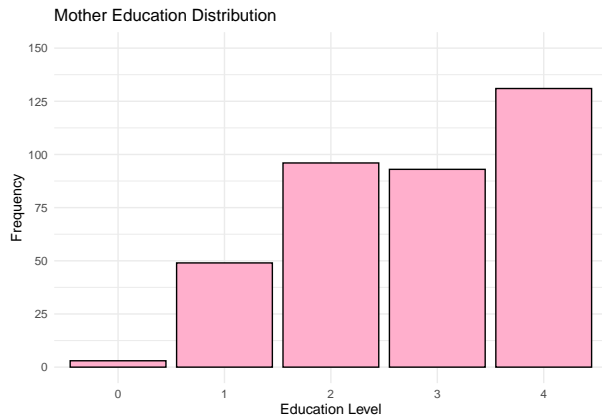

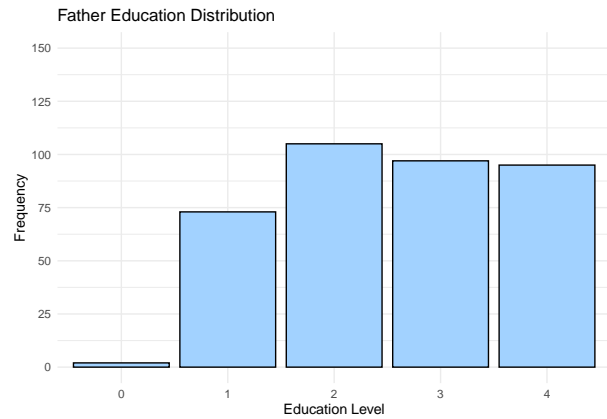Figure 3: Mother Education Distribution
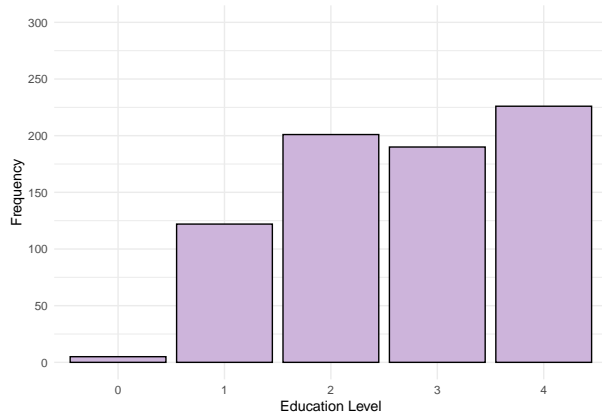

Figure 4: Father Education Distribution

Figure 5: Combined Parent Education Distribution

Rating	Description
0	No education
1	Up to 4th grade
2	Up to 9th grade
3	Secondary education (12th grade)
4	Higher Education

Subsequently, a contingency table was constructed, and a test for independence between mother's education and father's education was conducted. The resulting low p-value strongly suggested a lack of independence between the two variables.

	0	1	2	3	4
0	0	1	0	0	1
1	1	32	25	14	1
2	2	12	49	24	18
3	0	4	16	37	40
4	0	0	6	18	71

Table 5: Mother and Father Education Contingency Table (Row = Father, Column = Mother)

H_0: Mother's education and father's education are independent of one another.

H_1: Mother's education and father's education are not independent of one another.

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 200.38, df = 16, p-value < 2.2e-16
```

$p < 2.2 * 10^{-16} < \alpha = 0.05 \implies$ We have sufficient evidence to reject the H_0 and conclude that mother's education and father's education are not independent.

Additionally, I computed the means and standard deviations for both factors and employed the Kolmogorov-Smirnov test to assess the similarity of their distributions. Remarkably, the means and standard deviations were highly comparable, and the test failed to provide significant evidence rejecting the hypothesis that the distributions did not differ significantly.

Category	Mean	Standard Deviation
Mother	3.806452	1.084148
Father	3.564516	1.088350

Table 6: Mean and Standard Deviation for Mother and Father Education Levels

H_0: The distributions of mother's education and father's education do not differ significantly.

H_1: There is a significant difference between the distribution of mother's education and the distribution of father's education.

```
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(students$Medu) and as.numeric(students$Fedu)
## D = 0.096774, p-value = 0.06138
## alternative hypothesis: two-sided
```

$p = 0.06138 > \alpha = 0.05 \implies$ We do not have sufficient evidence to reject the H_0 and conclude that the distributions of mother's education and father's education are significantly different.

Consequently, the decision was made to amalgamate the two factors. This decision aligns with both statistical findings and practical considerations, emphasizing a collective parental influence on academic success rather than isolating the impact of mothers or fathers individually. This approach is particularly pertinent considering the diversity of family structures, acknowledging that not every set of parents consists of a mother and a father.

To create a composite factor that synthesized the information provided in both factors, I considered the structure of the factors in the original data set, which used numerical categorical variables with larger numbers representing higher levels of education, as follows:

Rating	Description
0	No education
1	Up to 4th grade
2	Up to 9th grade
3	Secondary education (12th grade)
4	Higher Education

As such, I summed the numbers for each parent, leading to a scale from 0 to 9 that similarly represents higher education with larger numbers and weights both parents' educations equally.

Diagnostics

To assess the hypothesis, I used the following one-way Multivariate Analysis of Variance (MANOVA) model:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} X + \epsilon$$

The matrix Y represents the grades in Portuguese and math classes for high school students:
$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$

The matrix μ represents the means of grades in Portuguese and math:
$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

The matrix B represents the regression coefficients for the effect of parental education on grades in Portuguese and math:
$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$

The matrix X represents the parental education levels.

The vector ϵ represents the residual errors.

Assumptions of the MANOVA Model

Independent Random Sampling

Students' responses are independent of each other and the 372 students were selected randomly.

Level and Measurement of the Variables

MANOVA assumes that the independent variables are categorical and the dependent variables are continuous. Parental education level is categorical and student final grades is continuous.
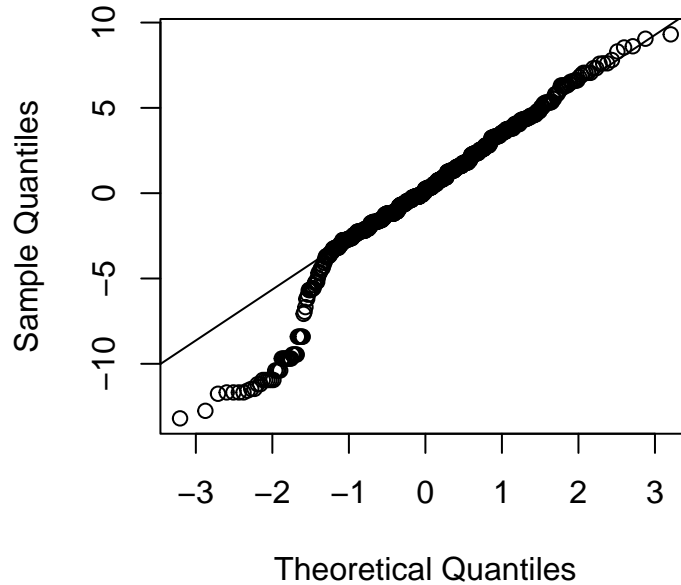
Absence of Multicollinearity:

```
cor(students$G3.m, students$G3.p)
```

```
## [1] 0.4932525
```

Our correlation of r = 0.4932525 does not suggest a high correlation between the Portuguese and math grades.

Normal Q–Q Plot

According to our Q-Q plot, we can assume normality. We could also assume this due to the sufficiently large sample size of 372 students.

Homogeneity of Variance

We assume population covariance matrices of each dependent variable are equal.

H_0: All group variances for the grades being tested are equal.

H_1: Not all group variances for the grades being tested are equal.

```
## [1] "Math Grades:"

## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   7  0.6519 0.7127
##       364

## [1] "Portuguese Grades:"

## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   7  1.1046 0.3595
##       364
```

Our large p-values of 0.7127 and 0.3595 suggest that this assumption is not violated.

H_0: Parental education has no impact on student grades.

H_1: Higher parental education corresponds with increased student grades.

```
## Analysis of Variance Table
##
##              Df  Pillai approx F num Df den Df  Pr(>F)
## (Intercept)   1 0.95096   3519.5      2    363 < 2e-16 ***
## Pedu          7 0.07054      1.9     14    728 0.02332 *
## Residuals   364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since $p = 0.02332 < \alpha = 0.05$, we have significant evidence to reject H_0 and conclude that higher levels of parental education correspond with increased academic success among their children.

Parental Education	Mean	SD	Count
3	12.500000	4.949747	2
4	8.411765	4.619574	34
5	9.459459	4.622377	37
6	9.691176	4.591548	68
7	10.390244	4.810811	41
8	11.196721	3.722990	61
9	10.948276	5.117642	58
10	11.676056	4.312025	71

Table 7: Math Grades by Parental Education Level

Parental Education	Mean	SD	Count
3	14.50000	4.949747	2
4	11.47059	3.727338	34
5	11.56757	3.329502	37
6	12.07353	2.339257	68
7	12.63415	2.817411	41
8	12.75410	3.284589	61
9	13.25862	2.380921	58
10	13.19718	2.675923	71

Table 8: Portuguese Grades by Parental Education Level

Conclusion

In conclusion, our statistical exploration of high school students' academic achievements in relation to familial dynamics has revealed significant associations between parental education levels and academic success. The observed correlation highlights the pivotal role of parental education as a critical factor influencing students' academic trajectories. However, it is important to acknowledge the limitations of our study.

One notable area for improvement is the scope of our sample. The study focused exclusively on students from two high schools in Portugal, and a more comprehensive study involving a wider range of schools, regions, and cultural backgrounds would contribute to a richer understanding of the nuanced factors at play.

Ultimately, this association underscores the importance of considering parental education as a critical factor influencing students' academic trajectories when developing programs and policies aimed at fostering academic success.

References

UCI Machine Learning Repository. (2014). Student Performance Data Set. Retrieved from https://archive.ics.uci.edu/dataset/320/student+performance

Scholaro. Grading System in Portugal. Retrieved from https://www.scholaro.com/db/Countries/Portugal/Grading-System