The impact of family dynamics in high school students' academic performance

Alejandra Savage Briz

Data Analysis 2, Coding 1

Abstract

This study draws upon one clean dataset from two Portuguese high schools, capturing comprehensive student achievement data. Sourced from school reports and questionnaires, this rich dataset includes a spectrum of student-related features, including grades, demographic information, social indicators, parental details, and school-related attributes.

Notably, the focus of this analysis centers on one of the datasets, meticulously cleaned and refined to enhance accessibility and usability, specifically describing students' performance in Mathematics. This dataset provides a detailed lens into the interplay between family dynamics and academic performance, offering a focused exploration of the intricate relationship between familial factors and students' mathematical achievements within the high school landscape.

Introduction

The early days of high school often create a complex interweaving of academic challenges and evolving family dynamics. This study focuses on one aspect of this intricate tapestry: the relationship between family dynamics and academic performance, particularly emphasizing first-period math grades. Employing regression analysis as a tool, the research aims to dissect how the quality of family relationships influences students' mathematical abilities.

Beyond this primary inquiry, the study acknowledges the multifaceted nature of the family environment. It considers various factors such as family size, parental education levels, family educational support, and the complexities of parental cohabitation status. Even a student's dedication to studying contributes to this textured weave. By accounting for these elements, the analysis seeks to unravel the nuanced impact of family dynamics on academic outcomes.

The overarching goal of this research is to shed light on the diverse influences of family dynamics on the academic journey. By identifying specific dimensions of family dynamics that significantly correlate with, or even predict, math grades, the study aims to establish foundations for impactful interventions and support systems. Imagine educational initiatives aimed at repairing strained relationships, reinforcing supportive elements, and untangling obstacles hindering academic success.

The ultimate ambition of this study is to harness the potential of positive family relationships while mitigating the effects of unfavorable family environments. By comprehending the intricate dance between family dynamics and mathematical performance, we can empower students to navigate their academic paths with confidence, guided by the supportive framework of a well-woven family tapestry.

Data

The dataset employed in this research, sourced from **Kaggle**, comprises student achievement data from two separate Portuguese high schools. Gathered via school reports and questionnaires, this

extensive dataset encompasses a diverse array of aspects, including student grades, demographic specifics, social indicators, parental details, and school-related attributes.

The dataset was originally thoroughly cleaned with no missing values, focusing on the domain of mathematics. With a total of 396 observations, this study centers on analyzing the ***first-period grade*** as the ***dependent variable (y)***, while the ***independent variable*** is the numeric representation of ***Family relationship (ranging from 1 - very bad to 5 - excellent)***.

Additionally, key factors considered in this analysis encompass **Family Size**, **Parental Education** levels, **Family Educational Support**, **Parental Cohabitation Status**, and **Study Time**. Here's a breakdown of the main variables utilized in this study:

- ***First Period Grade (y)***: Represents the initial grade achieved in the mathematics course (numeric: from 0 to 20).
- ***Family Relationship (Independent Variable)***: Numeric scale assessing the quality of family relationships, ranging from 1 (very bad) to 5 (excellent). It acts as the independent variable, scrutinized for its impact on first-period math grades.
- **Family Size**: Categorized as "Less or equal to 3" or "Greater than 3," indicating the size of the student's family.
- **Parental Education**: Ordinal categorization of both mother's and father's education levels, ranging from "none" to "higher education."
- **Family Educational Support**: Binary assessment (yes or no) of whether the family provides educational support to the student.
- **Parental Cohabitation Status**: Binary categorization indicating whether parents live together or apart.
- **Study Time**: Weekly study time categorized on an ordinal scale from "¡2 hours" to "¿10 hours," reflecting the amount of time spent studying per week.

These variables were selected due to their potential influence on a student's academic performance in mathematics, aiming to understand how family dynamics and various related factors interrelate with the first-period math grades in high school.

Models

In this study, the objective is to construct various regression models to explore the relationships between different variables.

0.1 Model 1: Simple Regression between Final Grade and Family Relationship

One of the proposed models, termed as **Model 1**, involves a simple linear regression analysis. The model is represented as:

$$FinalGrade = \beta_0 + \beta_1 \times FamilyRelationship$$

This particular simple linear regression model aims to investigate the potential association between the quality of family relationships and students' final grades. It seeks to discern the impact or correlation between these variables within the context of academic performance.

While the equation $FinalGrade = 9.3800 + 0.2624 \times FamilyRelationship$ implies a positive relationship between family relationship quality and final grades, it offers an incomplete perspective. Three critical aspects raise doubts about its reliability:

- Lack of statistical significance: The p-value for the *family_relationship* coefficient exceeds 0.05, indicating an absence of strong evidence to confidently assert this association within the broader population.
- Minimal explanatory power: The R-squared value of 0.003 suggests this model explains only a negligible fraction of the variability in final grades. Solely considering family relationship quality fails to capture the comprehensive picture.
- Unrealistic intercept: The intercept of 9.38 proposes an average final grade even for a student experiencing the lowest possible family relationship quality, which seems implausible. This underscores the model's limitations in depicting real-world scenarios.

Consequently, despite the positive coefficient hinting at a potential correlation, the lack of significance and the model's overall inadequacy prevent a definitive conclusion regarding the significant impact of family relationships on student success. The development of more comprehensive models incorporating pertinent factors such as study habits, individual characteristics, and the school environment is indispensable to unraveling the intricate web of influences on academic achievement.

0.2 Model 2: Final grade on parents education

This model aims to predict a student's final grade based on their family relationships and parents' education levels.

While this model identifies some factors influencing student final grades, its limited explanatory power suggests caution in drawing strong conclusions. Mother's higher education emerges as a statistically significant predictor, with students whose mothers hold a degree scoring an average of 2.67 points higher on final grades ($p = 0.002$). This suggests a potentially substantial positive impact. However, the model only explains 5.7% of the variance in final grades, with the adjusted R-squared dipping further to 4.0%. Notably, the "family_relationship" coefficient, despite suggesting a positive association, lacks statistical significance due to its p-value exceeding 0.05. Similarly, other parental education variables fail to show reliable relationships with final grades. This underscores the complexity of factors influencing student success and suggests that mother's higher education, while significant, likely interacts with other crucial, yet uncaptured elements.

0.3 Model 3: Final grade on parents education

The model is represented as:

$$FinalGrade = \beta_0 + \beta_1 \times FamilyRelationship + \beta_2 \times FamilySize(Greaterthan3) + \beta_3 \times ParentStatus(Apart)$$

The model aims to uncover potential connections between family-related factors and final grades, but ultimately fell short in establishing statistically significant relationships. None of the variables—such as family relationship quality, family size, or parental cohabitation status—showed substantial links with final grades, as their respective p-values exceeded the accepted threshold of 0.05.

Moreover, the model's explanatory power was notably limited, explaining only a small fraction of the variance in final grades, suggesting a considerable portion of factors impacting student achievement remained unexplored. Additionally, the overall fit of the model, as indicated by the F-statistic's p-value of 0.202, raises doubts about its statistical relevance. Thus, based on this analysis, definitive conclusions regarding the significant impact of these family-related factors on student performance cannot be drawn.

Insights and Conclusions

0.4 Causal Analysis

Causal analysis, a statistical methodology employed in this study, endeavors to unravel the intricate web of causality between family dynamics and high school students' math grades. Through regression

models, researchers seek to discern whether variables like family relationships, parental education, and study habits exert a causal impact on academic performance while accounting for potential confounding factors.

However, establishing causality proves intricate due to various challenges. The mere correlation between variables, such as family dynamics and math grades, doesn't inherently imply a cause-and-effect relationship. Omitted variable bias, where essential variables are left out, and the potential for reverse causality further complicate the endeavor.

To confront these challenges, researchers have diligently crafted multiple regression models, endeavoring to isolate genuine causal associations amidst the complexity of influences. They've scrutinized statistical significance using p-values to distinguish true effects from chance occurrences and have remained vigilant regarding alternative explanations and potential confounding factors.

Despite these meticulous efforts, conclusive evidence elucidating a direct causal link between family dynamics and math grades remains elusive within this study's scope. While certain variables, notably a mother's higher education, hint at a plausible connection, further research is imperative to validate these initial findings. Moreover, this pursuit necessitates an exploration of additional potential causal factors contributing to high school students' academic performance.

0.5 Conclusions

The first hypothesis proposed a clear relationship between family relationships and math performance. However, the intricate results revealed by the analysis describe another reality. While the threads of family life are undeniably vital, our quest for a straightforward, significant impact yielded surprising results.

Across four models, statistically significant relationships between family relationships and math scores remained elusive. Whether family relationships were good or bad, the predicted impact on final grades remained negligible. This suggests that while family is undoubtedly important, other factors likely play a more prominent role in shaping student outcomes.

Yet, one main conclusion of the analysis revealed that students who have a mother with higher education emerged as a potential thread linking family to academic success. Students with mothers who had higher education were predicted to score higher on final exams, hinting at a complex interplay between parental academic attainment and student achievement.

Moving beyond family dynamics, the research also explored the textured thread of study habits. Here, a glimmer of a potential connection between dedicated study time and improved grades emerged. Students who invested 5 to 10 hours per week in studying were predicted to see a modest boost in their final scores compared to those who studied less. However, the lack of strong statistical significance calls for further weaving, with a focus on strengthening this potential link.

This underscores the need for a more comprehensive exploration, incorporating diverse variables like individual learning styles and socioeconomic backgrounds. Integrating qualitative research methods could enrich our understanding of students' experiences and contextual influences. Though our initial hypothesis wasn't confirmed, this study deepens our understanding of family dynamics and academic performance.

Appendices

0.6 Appendix 1. Comparison table across all models

0.7 Appendix 2. Regression model comparison

0.8 Appendix 3. Final grade vs Family relationship

	Dependent variable: final_grade			
	(1)	(2)	(3)	(4)
Constant	9.380***	7.576***	9.788***	9.175***
	(1.104)	(1.210)	(1.168)	(1.167)
Quality of family relationship	0.262	0.298	0.277	0.225
	(0.272)	(0.259)	(0.269)	(0.270)
Mother with high education		2.668***		
		(0.835)		
Mother with 5th-9th grade education		0.646		
		(0.771)		
Mother with secondary education		1.218		
		(0.803)		
Father with high education		0.407		
		(0.860)		
Father with 5th-9th grade education		0.498		
		(0.710)		
Father with secondary education		0.243		
		(0.731)		
Family size >3			-0.762	
			(0.495)	
Parents living apart			0.721	
			(0.719)	
2-5 hr study time				0.111
				(0.570)
5-10 hr study time				1.290*
				(0.751)
>10 hr study time				1.234
				(1.104)
Observations	395	395	395	395
R^2	0.003	0.057	0.012	0.015
Adjusted R^2	0.000	0.040	0.004	0.005
Residual Std. Error	4.581 (df=393)	4.489 (df=387)	4.572 (df=391)	4.570 (df=390)
F Statistic	0.930 (df=1; 393)	3.653*** (df=7; 387)	1.749 (df=3; 391)	1.394 (df=4; 390)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 1: Comparison table across all models

Model	Model 1	Model 2	Model 3	Model 4
R-squared	0.00	0.08	0.01	0.02
Adjusted R-squared	0.00	0.04	0.00	0.00
F-statistic	0.93	3.85	1.75	1.39

Figure 2: Regression model comparison

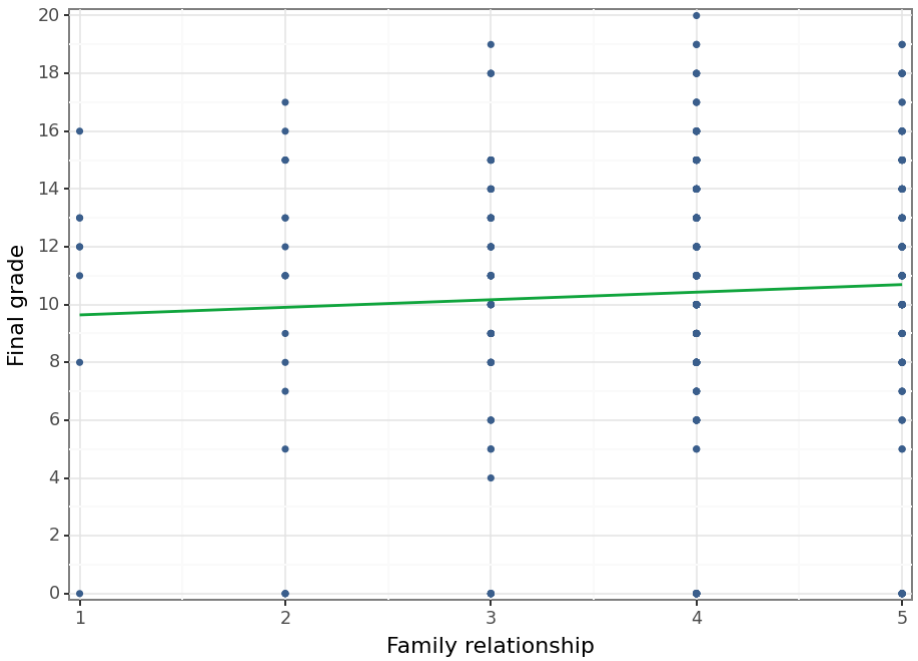

Figure 3: Scatterplot of Regression Model