

# How Family Background And Resources Affect Academic Success

## Motivation

When we began university, we noticed that some peers excelled academically with seemingly less effort compared to others. This observation sparked our curiosity: What factors contribute to their success? Could their family background or access to resources play a role? These questions motivated us to explore the relationship between family environment, socioeconomic factors, [4] and the academic performance of undergraduate engineering students. Our aim is to better understand how these external influences shape academic success and highlight the disparities that may exist among students.

## Data Collection

Our data comes from Mendeley Data [1], specifically the “**Data of Academic Performance Evolution for Engineering Students**” dataset published in 2020. This dataset was originally collected by the Technological University of Bolívar as part of a Master’s Degree in Engineering project, using data from the Colombian Institute for the Evaluation of Education (ICFES). It includes detailed records of 12,411 undergraduate engineering students, covering their academic performance and social class factors such as parents’ education levels, socioeconomic stratum (0-6), household size, and ownership of resources like computers and cars [3].

This dataset provides comprehensive information on both academic outcomes and the social class variables relevant to our study. It was collected from official educational databases maintained by ICFES. Both ICFES and the Technological University of Bolívar are reputable institutions in Colombia, which adds credibility to the data. Since the dataset has been anonymized and is publicly available on the trusted platform Mendeley Data, it is both suitable and ethical for our research purposes.

## Methods & Analysis

We conducted a multiple linear regression analysis to investigate how social class factors affect the academic performance of undergraduate engineering students. Our outcome variable was the **OVERALL\_SCORE**, calculated by summing five national standardized test scores: Quantitative Reasoning (QR\_PRO), Critical Reading (CR\_PRO), Citizen Competencies (CC\_PRO), English (ENG\_PRO), and Written Communication (WC\_PRO). This composite score, with a maximum of 500 [3][5], provided a measure of each student's academic performance:

$$\text{OVERALL\_SCORE} = \text{QR\_PRO} + \text{CR\_PRO} + \text{CC\_PRO} + \text{ENG\_PRO} + \text{WC\_PRO}$$

Our predictor variables included social class factors such as parents' education levels (EDU\_FATHER and EDU\_MOTHER), the parent's occupations (OCC\_FATHER and OCC\_MOTHER), socioeconomic stratum (0-6)[4], household size (PEOPLE\_HOUSE), and ownership of resources, computers, mobile phones, and cars. We also considered the student's gender (GENDER) and the university they attended. To account for the potential influence of university reputation on academic performance, we categorized universities into tiers (1-5) based on their world rankings obtained from reputable international ranking systems [2].

Our multiple linear regression model aimed to quantify the impact of these variables on the overall academic score. The mathematical representation of our model is:

$$\text{OVERALL\_SCORE}_i = \beta_0 + \beta_1 \text{GENDER}_i + \beta_2 \text{EDU\_FATHER}_i + \beta_3 \text{EDU\_MOTHER}_i + \beta_4 \text{STRATUM}_i + \beta_5 \text{PEOPLE\_HOUSE}_i + \beta_6 \text{COMPUTER}_i + \beta_7 \text{MOBILE}_i + \beta_8 \text{CAR}_i + \beta_9 \text{UNIVERSITY\_TIER}_i + \beta_{10} \text{OCC\_FATHER}_i + \beta_{11} \text{OCC\_MOTHER}_i + \epsilon_i$$

where  $\epsilon_i$  represents the error term for student  $i$ .

## Methods & Analysis

We analyzed data from 12,411 students, providing a substantial sample size to underline the reliability of our findings. By incorporating these variables into our model, we aimed to understand how social class factors, access to resources, and university influence academic performance among engineering students.

### Data Preparation and Transformation

- Reformatted variables
  - STRATUM : Converted to a categorical variable with levels 0 to 6
  - PEOPLE\_HOUSE: Converted to a numerical variable
  - COMPUTER, CAR, and MOBILE: Converted to binary indicators (1=Yes, 0=No)
  - Discarded records with missing (N/A) values for key predictors
- Outcome Transformation
  - Applied a powertransformation to meet the normality assumption
  - Using log-likelihood estimation method, the optimal  $\lambda$  was determined to be 1.47, rounded to 1.5. → OVERALL\_SCORE $^{1.5}$

### Assumption Checking and Model Refinement

- Diagnostic Checks
  - Multicollinearity: Assessed using Variance Inflation Factors; no significant
  - Partial F-tests:
    - COMPUTER and CAR: Found to be statistically insignificant at a 95% C.I; null hypothesis was not rejected, and were removed from the model
    - PEOPLE\_HOUSE: Marginally insignificant but retained due to its relevance to the research question
  - Outliers:
    - Identified both high-leverage and high-residual outliers in the dataset
    - Applied a filtering approach to exclude observations that were both high leverage and had disproportionate influence on the regreession coefficients, ensuring data integrity without introducing bias.

### Final Regression Model

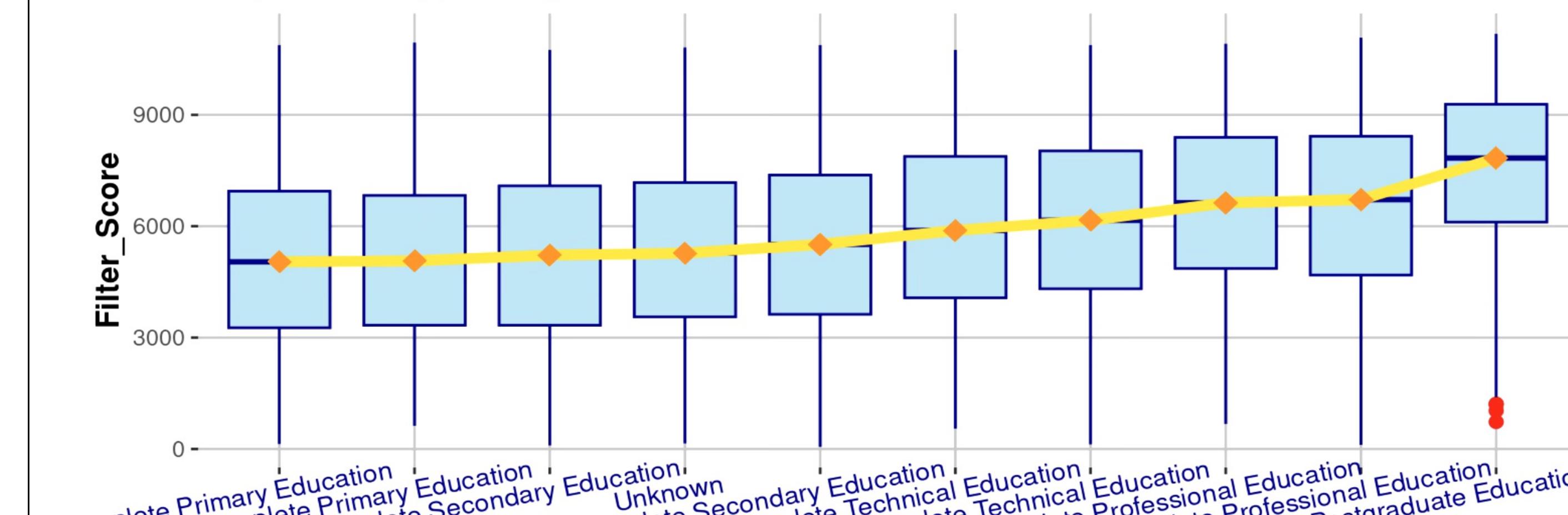
$$\text{Filter\_Score}_i = \beta_0 + \beta_1 \text{GENDER}_i + \beta_2 \text{EDU\_FATHER}_i + \beta_3 \text{EDU\_MOTHER}_i + \beta_4 \text{OCC\_FATHER}_i + \beta_5 \text{OCC\_MOTHER}_i + \beta_6 \text{STRATUM}_i + \beta_7 \text{MOBILE}_i + \beta_8 \text{UNIVERSITY}_i + \beta_9 \text{PEOPLE\_HOUSE}_i + \epsilon_i$$

## Results

Family background and social class factors significantly influence academic success, measured by the power-transformed overall score (Score $^{1.5}$ ). Students with parents who have higher education levels consistently achieve higher median scores (Figure 1). For example, the median score increase by approximately 42% from Score $^{1.5} = 5511.02$  (complete secondary education) to Score $^{1.5} = 7835.57$  (postgraduate education) for fathers. Similar trends are observed for mother's education levels. These differences demonstrate a strong positive correlation between parental education levels and student performance. Socioeconomic status, [4] also shows a significant impact on academic success, as seen in Figure 3. The power-transformed median score increases by 37% from Score $^{1.5} = 6131.51$  (stratum 3) to Score $^{1.5} = 8423.65$  (stratum 6). This finding highlights the important role of socioclass factors in academic achievement. University tier further amplifies the effects of family background. The median Score $^{1.5}$  increase by approximately 61%, from Score $^{1.5} = 5537.54$  (Tier 3) to Score $^{1.5} = 8916.67$  (Tier 5), as shown in Figure 2. This demonstrates that institutional factors, such as the quality and prestige of universities, significantly contribute to academic performance. Our research aimed to determine how family background influence academic success. The analysis shows that higher parental education levels, socioeconomic status,

and attending a prestigious universities are highly associated with better academic performance.

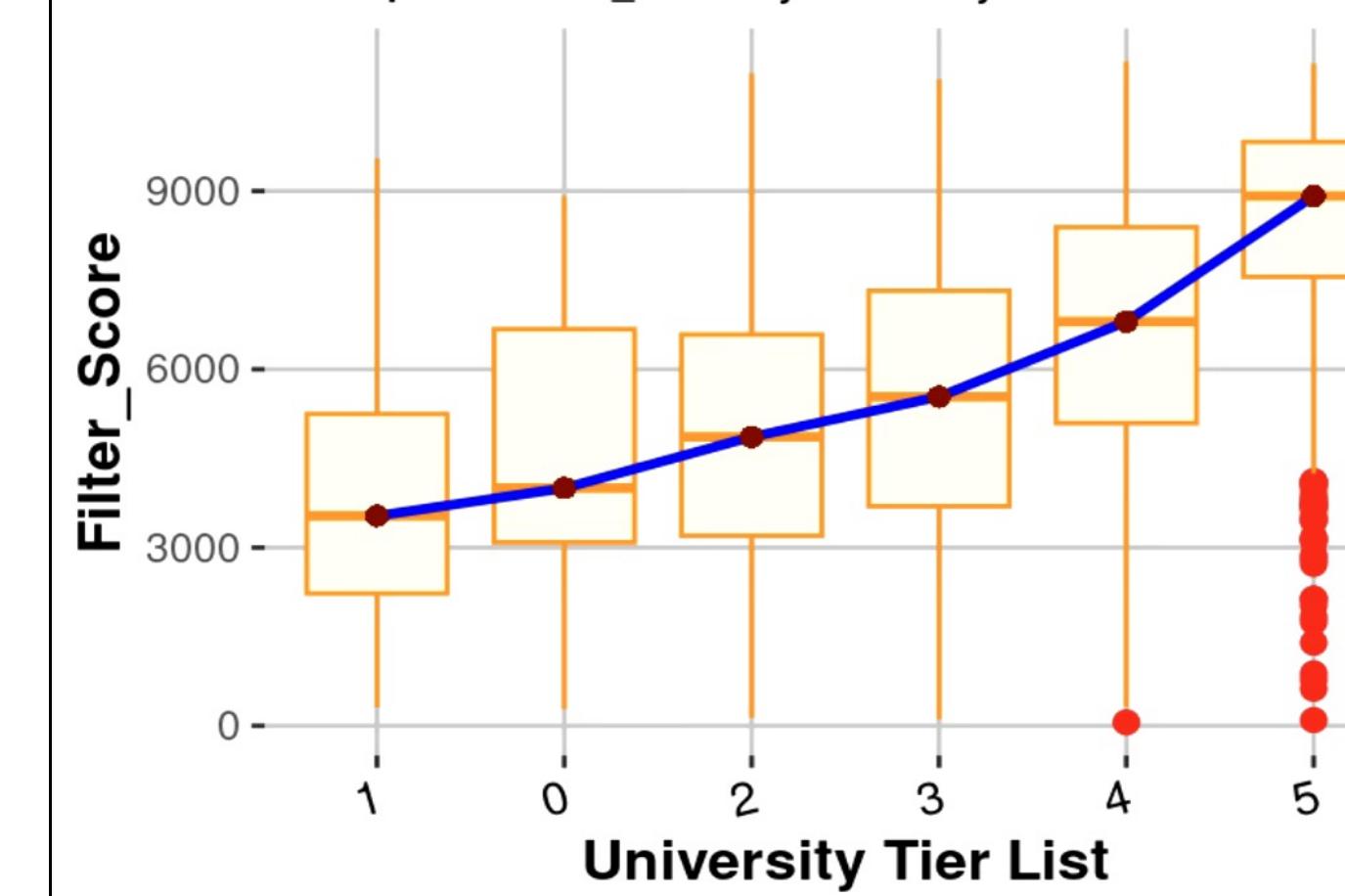
Boxplot of Filter\_Score by Father's Education Level with Trend Line



Father's Education Level (Ordered by Median Filter\_Score)

Figure 1: Boxplot of the power-transformed Filter\_Score ( $\lambda=1.5$ ) across different categories of father's education level, ordered by the median Filter\_Score for each category. Yellow trend line connects the medians, highlighting the aradual increase in Filter Score with higher levels of father's education.

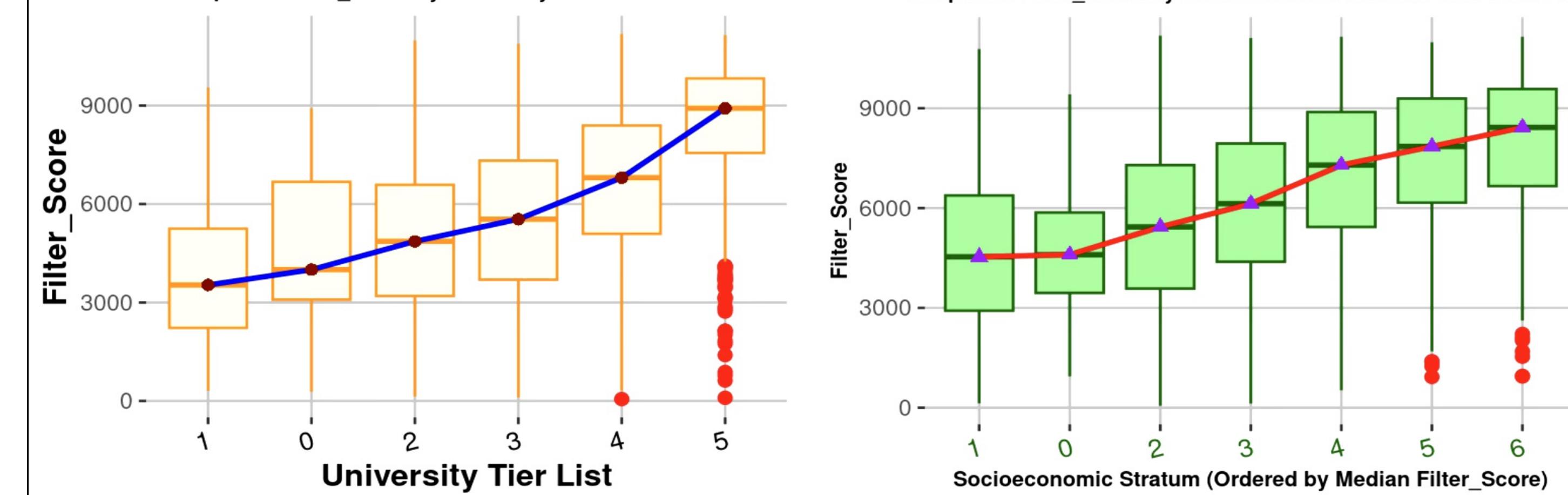
Boxplot of Filter\_Score by University Tier with Trend Line



University Tier List

Figure 2: Boxplot of the power-transformed Filter\_Score ( $\lambda=1.5$ ) across university tiers. Blue trend line connects the medians, highlighting the positive relationship between Filter\_Score and university tier.

Boxplot of Filter\_Score by Socioeconomic Stratum with Trend Line



Socioeconomic Stratum (Ordered by Median Filter\_Score)

Figure 3: Boxplot of the power-transformed Filter\_Score ( $\lambda=1.5$ ) across socioeconomic strata. Orange trend line connects the medians, highlighting the positive relationship between Filter\_Score and socioeconomic stratum.

## Conclusion

Our analysis indicates that family background and social class significantly impact academic performance among Colombian engineering students. Students whose fathers have postgraduate degrees score 42% higher than those whose fathers completed secondary education in Score $^{1.5}$ . A similar trend is observed for mothers educational level. Socioeconomic status also shows a impact. Student in stratum 6 achieve a median (Score $^{1.5}$ ) that is 37% higher than student in stratum 3. These disparities informed our regression model, emphasizing the influence of parental education and socioeconomic status. These findings answer our research question by demonstrating that higher parental education and socioeconomic status are associated with better academic performance. However, focusing solely on Colombian engineering students limits generalizability, and potential issues with outliers and nonlinear relationships suggest the model may not fully capture all influencing factors.

Future studies should include more diverse populations and explore advanced modeling techniques to account for nonlinear effects. Adding variables such as parental involvement could further enhance our understanding of factors influencing academic success.

## Reference

[1] Data of Academic Performance Evolution for Engineering Students (2020). Mendeley Data. <https://data.mendeley.com/datasets/83tcx8psxv/1>

[2] University World Rankings Data (2024). [https://docs.google.com/spreadsheets/d/1b2dU7jN7cUjbtQkM-MPlwbjhbYnmL3\\_zrSEDxafXl/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1b2dU7jN7cUjbtQkM-MPlwbjhbYnmL3_zrSEDxafXl/edit?usp=sharing)

[3] Delahoy-Domínguez, E., Zuluaga, R., & Fontalvo-Herrera, T. (2020). Dataset of academic performance evolution for engineering students. <https://doi.org/10.1016/j.dib.2020.105537>

[4] Chica-Olmo, J., Sánchez, A., & Sepulveda-Murillo, F. H. (2019). Assessing Colombia's policy of socio-economic stratification: An intra-city study of self-reported quality of life. <https://doi.org/10.1016/j.cities.2019.102560>

[5] Colombia Education Info. (n.d.). Undergraduate and postgraduate admission test. <https://www.colombiaeducation.info/tests/undergraduate-and-postgraduate-admission-test.html>