

Causal Inference Project:

Impact of Scholarships on Student Success

Anushka Mukherjee, Lucie Marimar, Jort Koks, Jakob Sarrazin

April 04th, 2025

Machine Learning for Econometrics

ENSAE, IP Paris

Bruno Crépon, Matthieu Doutreligne

1. Motivation

Retention and Completion: A Core Challenge for Universities

- **High dropout rates** are a persistent issue in higher education, especially during the first years of study.
- **Timely graduation** is crucial for both students (career entry) and universities (funding, reputation)
- ⇒ **Financial constraints** are a major barrier to academic success — especially for socio-economically disadvantaged students.



Scholarships as a Tool to Improve Student Retention and Graduation

- **Scholarship programs** are widely used as an intervention, but:
 - Their **causal effect** on student outcomes is difficult to measure
 - Many studies show correlations, but few rigorously identify causality.
- This study uses a **causal machine learning framework (DML)** to estimate the **true effect of scholarships**, adjusting for observed confounders.
- Findings can inform **policy decisions** on financial aid allocation and **targeting of support** for at-risk students.

2. PICO & Research Question

Population, Intervention, Comparison, Outcome

- P - Undergraduate students at a Portuguese university (N = 4,424), with data on demographics, socio-economic background, and prior academic performance.
- I - Receiving a scholarship during university studies.
- C - Students without scholarships, adjusted for observed confounders (grades, family background, gender, etc.).
- O - Two binary outcomes observed 3 years after enrollment:
 1. Dropout vs. Enrolled/Graduated
 2. Graduated vs. Dropout/Enrolled

Research Question

RQ1

Does receiving a scholarship **reduce** the likelihood of **dropping out** within 3 years?

RQ2

Does receiving a scholarship **increase** the likelihood of **graduating** within 3 years?

3. Data Overview and Exploratory Analysis

Source

- UCI Machine Learning Repository – Predict Students Dropout and Academic Success

Scope

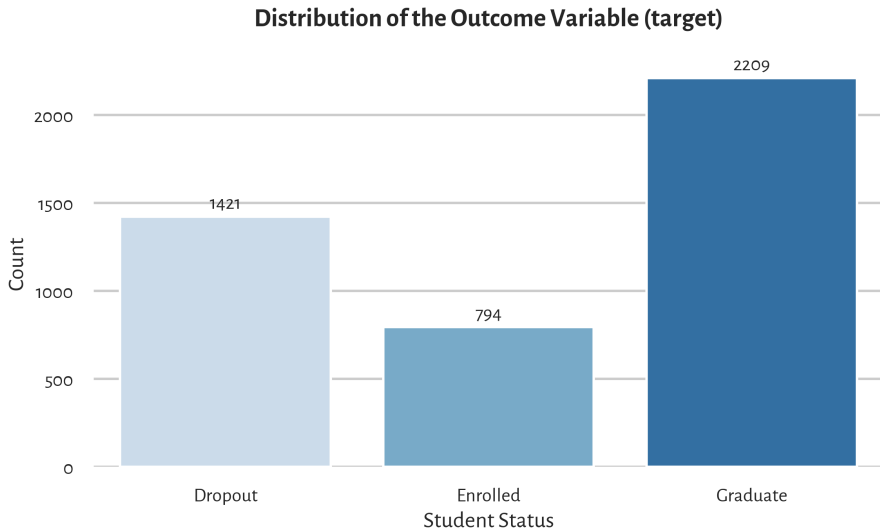
- Administrative records from a Portuguese university → 4,424 undergraduate students across various degree programs

Observation Period

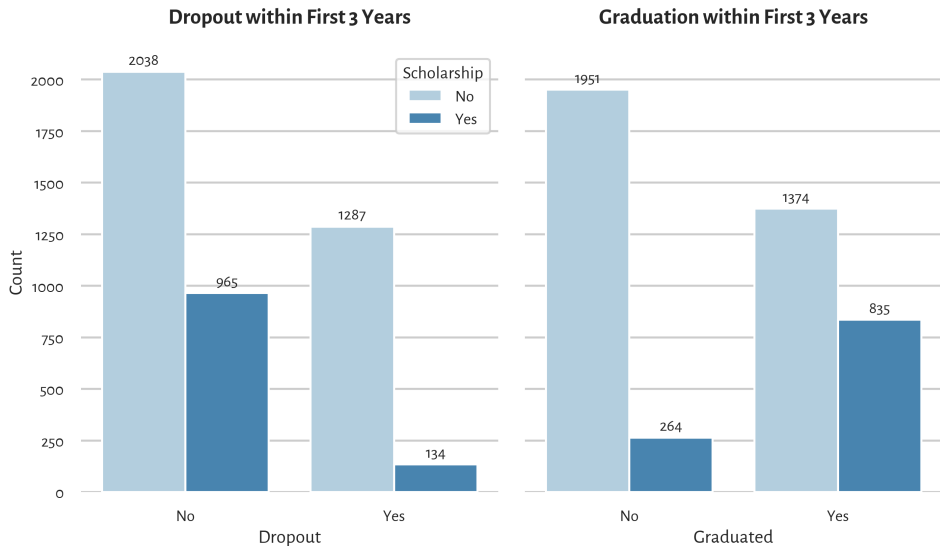
- Students tracked for 3 years after enrollment.

Variables

Outcome Variable	Treatment Variable	Covariates (Pre Treatment)
<p>Student status after 3 years:</p> <ul style="list-style-type: none">– Dropout– Still enrolled– Graduated <p>→ <i>Re-coded into two binary variables for RQ1 & RQ2</i></p>	<p>Received scholarship or not</p> <p><i>(Binary variable)</i></p>	<ul style="list-style-type: none">– Academic performance before university– Family background– Economic context– Demographics



Treatment vs Outcome Variable



4. Causal Graph and Covariate Selection

5. Causal Effect Estimation Using Double Post Lasso

6. Causal Effect Estimation Using Double Machine Learning

X. Conclusion
