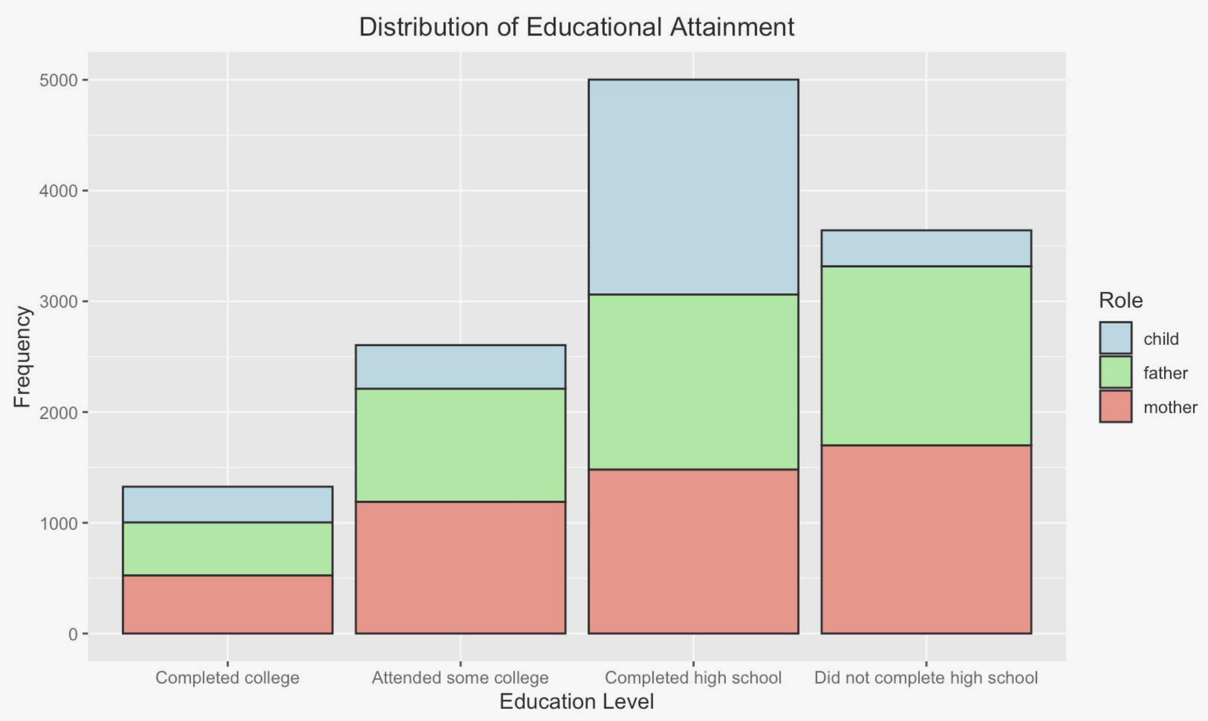
The Influence of Early Education on Children's Eventual Academic Attainment

Isaac Wecht, Li-Heng Pan, Sagar Raichandani, Sarah Prakriti Peters, YouJung Koo
BROWN

GitHub Link : https://github.com/harren7/familycausalinference

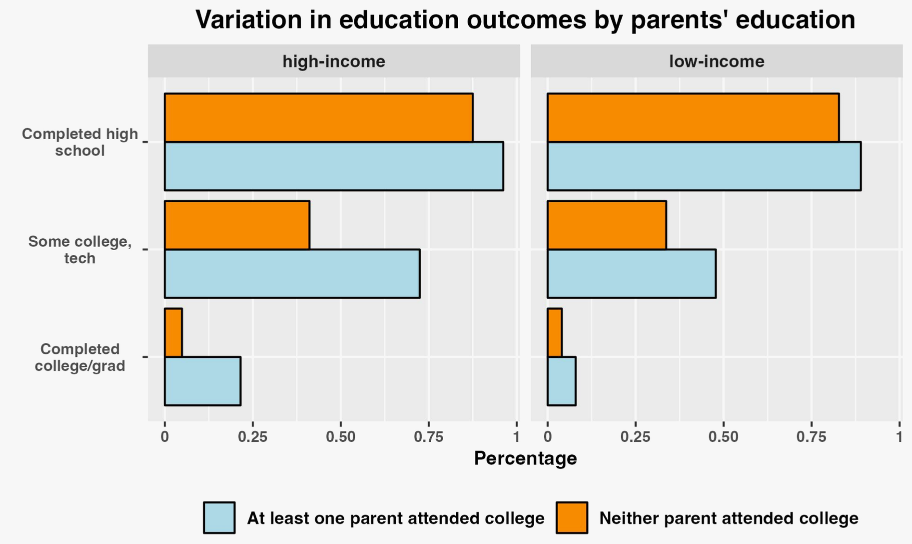
Introduction



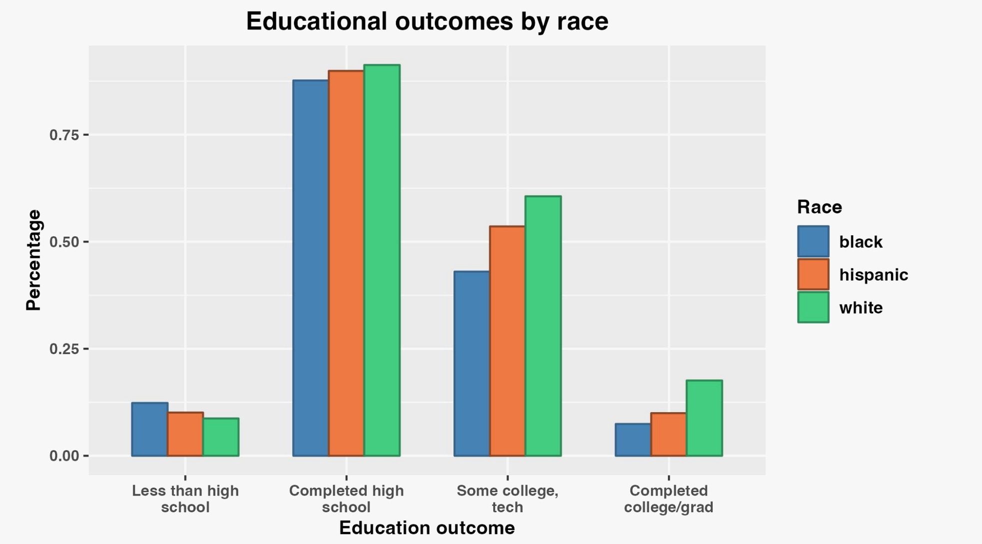
Education is often portrayed as a vehicle for social mobility, leading to higher lifetime earnings. However can education, in and of itself, drive the same generational benefits after controlling for lower economic status? We hypothesize that early childhood education positively correlates with increased long-term academic achievement among children.

We seek to investigate the causal impact of early childhood education on the academic outcomes for children of low-income families.

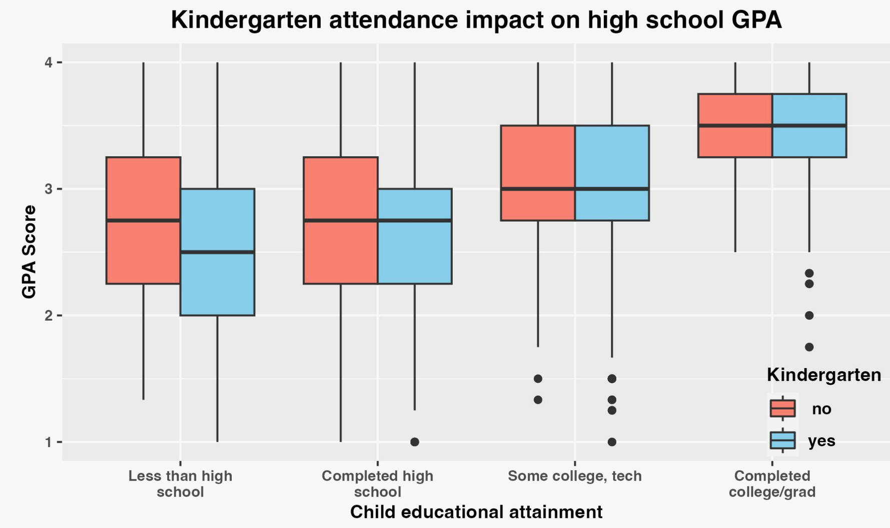
Exploratory Data Analysis



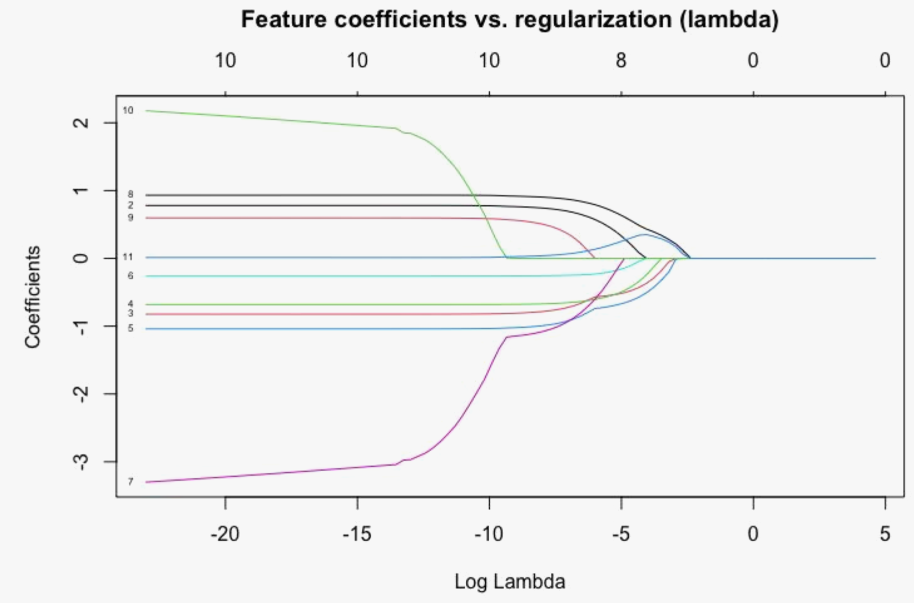
The children of highly educated parents often attain higher levels of education. This effect is compounded by income.



Racial disparities are tied closely with educational outcomes. Black students are more likely to drop out of high school, and least likely to complete college.



Children who achieve higher levels of education earn higher GPAs during high school. A child's kindergarten attendance has no impact.



Large values for the Lasso penalty drive coefficients for all features to zero. The optimal value for the Lasso penalty was 0.001.

Methods

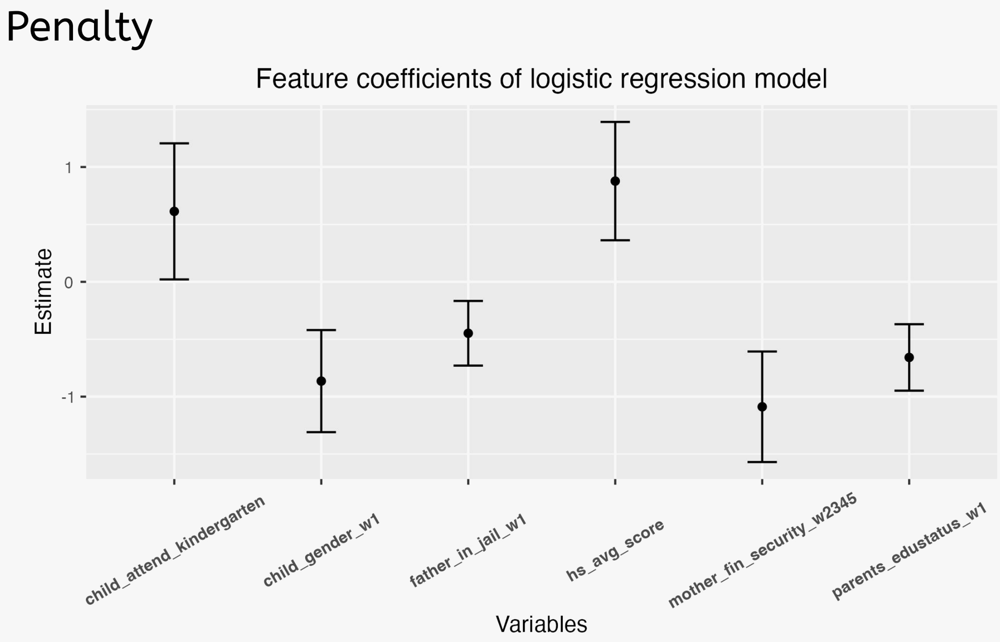
Logistic Regression

Overview: Predict whether a child will graduate from college based on parents' education, social factors, and family finances.

Evaluation: McFadden's R2 (goodness of fit), F1 (accuracy of prediction)

Techniques: cross-validation with L1 Penalty

Data dimensions: (2500, 7)



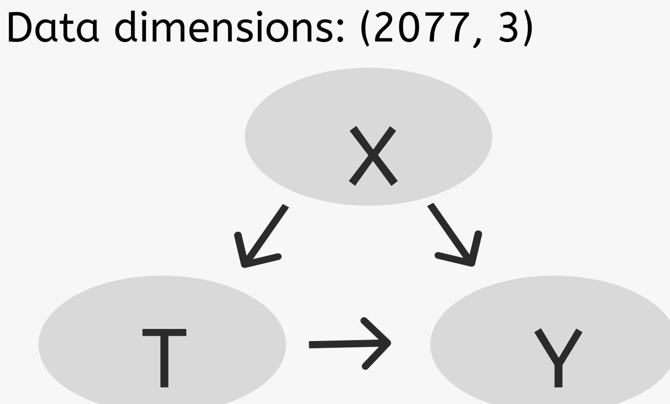
Casual Inference

Overview: To analyze the casual relation between access to early education and long-term academic outcomes, we modeled the propensity scores of the treatment as a logistic regression model. The propensity scores were then applied as weights to an outcome model that measures the relation between the target and treatment.

Evaluation: Average Treatment Effect

Techniques: inverse propensity weighting

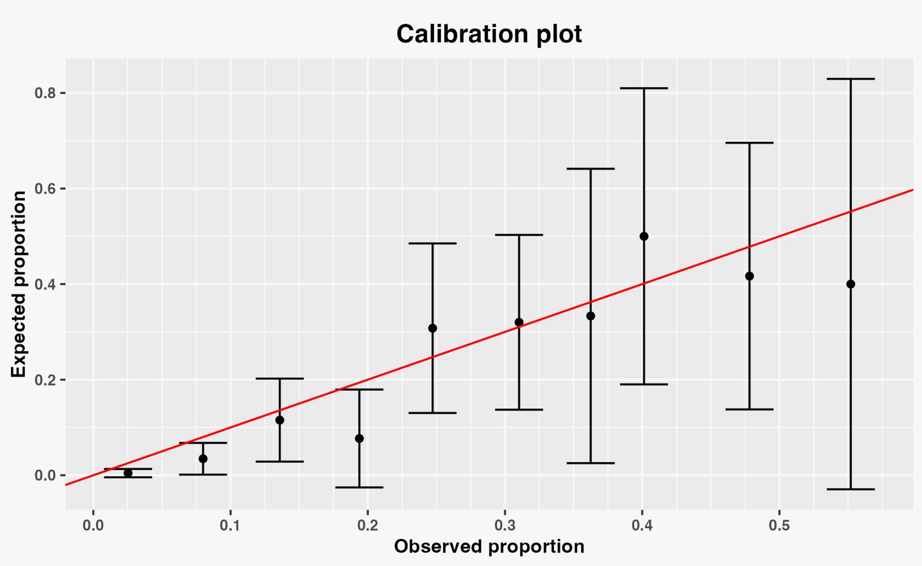
Data dimensions: (2077, 3)

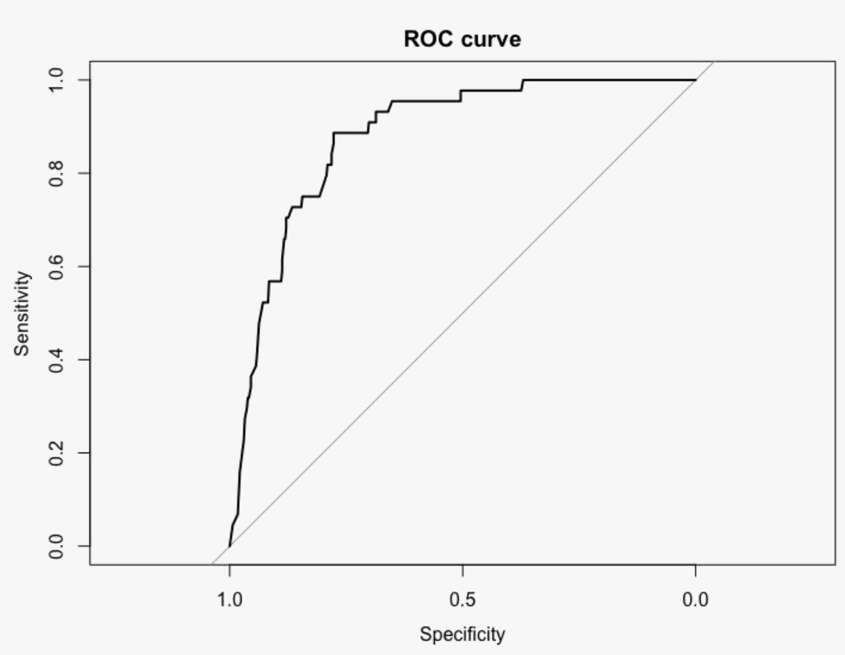


Variable	Type	Features
X	Confounders	Mother's income, Parents' education, Mother's race
Y	Target	College Graduation
T	Treatment	Kindergarten Attendance

Results

Logistic Regression

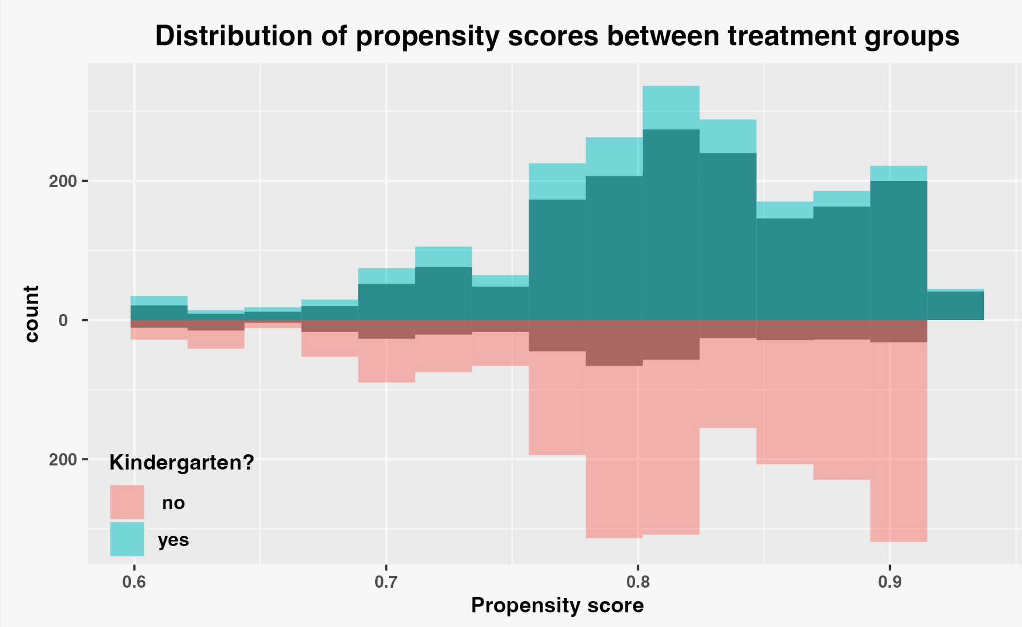




McFadden's R2 score: 0.204
Precision: 0.323

F1: 0.432 (at p = 0.264)
Recall: 0.612

Causal Inference



A mirrored histogram of the distribution of propensity scores between treatment groups. Dark bars represent unweighted distribution and lighter bars represent weighted by the ATE weights.

Feature	Estimate	Std. err	p-value
Intercept	1.68	0.174	4.46×10^{-22}
Mother's income	0.077	0.025	2.73×10^{-3}
Parents' education (Factor = Neither)	-0.55	0.130	1.90×10^{-5}
Mother's race (Factor = Hispanic)	-0.373	0.144	9.67×10^{-3}
Mother's race (Factor = White)	-0.803	0.145	2.71×10^{-8}

ATE_low: 4.45% ATE_high: 6.92%

ΔATE: 2.47% ATE_total: 6.19%

Reference

[1] Alves, M., "Causal Inference for the Brave and True"
[2] Deng, A., "Causal Inference and Its Applications in Online Industry"
[3] De Vito, R, "Probability, Statistics, and Machine Learning"

Contributions

Isaac: EDA, Poster Design
Liam: EDA, Logistic Regression
Sagar: Logistic Regression, Causal Inference
Sarah: Data Collection, Causal Inference
YouJung: Data Collection, Poster Design