

# The Effect of Age of Enrollment on the Probability of Graduating in Courses - Project Proposal

Gur Keinan 213635899

Yarden Adi 212585848

August 25, 2024

In this project, we aim to investigate a crucial causal question: *What is the causal effect of being an adult student (over 21) on the probability of completing academic courses?* This research question is significant as it addresses the impact of age of enrollment on academic success, a topic of interest to academic researchers, educators, and policymakers.

**Datasets** We will use a Kaggle dataset called *Predict Dropout or Academic Success* which contains information about students' pre-academic background, age, academic performance, social and economic status, and other relevant variables. It contains 4424 records and 37 variables, including the age of enrollment. The dataset includes a trinary outcome variable indicating whether a student dropped out, graduated, or is still enrolled in the course.

**Challenges in the data** The dataset contains some post-treatment variables, which may interfere with the causal analysis we plan to conduct. For example, the dataset includes the average grade of the students after the first 2 semesters, which is a post-treatment variable that may be affected by the treatment (age of enrollment). We will address these issues by conducting a careful analysis with and without those variables, as we have learned in class. Additionally, there might be confounders we are unaware of, which could bias our results. We will thoroughly investigate the data and the existing literature on the topic to identify potential confounders and address them in our analysis.

Our project will have the following structure:

1. **Data Preprocessing:** We will clean the data, handle missing values, and transform the data into a suitable format for the analysis.
2. **Exploring the Data:** We will conduct a descriptive analysis of the data to understand the distribution of the variables and identify potential patterns and trends.
3. **Discussing the Needed Assumptions for Causal Inference:** We will discuss the assumptions needed for causal inference and assess whether they hold in our context. Such assumptions include but are not limited to the Stable Unit Treatment Value Assumption (SUTVA), the Ignorability Assumption, the Overlap Assumption, and the other conditions we learned in class.
4. **Causal Analysis** We will utilize multiple methods to estimate the causal effect of being an adult student on the probability of graduating in courses. We will employ methods such as propensity score matching, inverse probability weighting, and regression adjustment, ensuring the robustness of our findings.
5. **Results:** We will present the results of our analysis, including the estimated causal effect and the uncertainty around the estimate.
6. **Discussion:** We will discuss the implications of our findings and the limitations of our analysis. We will also suggest potential avenues for future research.

Overall, our findings will provide valuable insights into the effect of age of enrollment on the probability of graduating in courses and contribute to the existing literature on the topic.