



Who's Your Neighbor?

Using Matching Methods to Evaluate Intervention Effectiveness

*Michelle Borckardt, MPA
Gateway Technical College*

*AIRUM Conference
November 2024*





About Gateway

- Public 2-year technical college in southeastern Wisconsin
- First publicly funded technical college in America
- In 2023-24:
 - 7,715 degree-seeking students (3,301 FTE)
 - 20,442 students overall
- Hispanic Serving Institution

Session Objectives



Understand propensity score matching (PSM) theory & methodology



Learn how we used PSM to assess tutoring impact



Practical steps to run PSM in SPSS

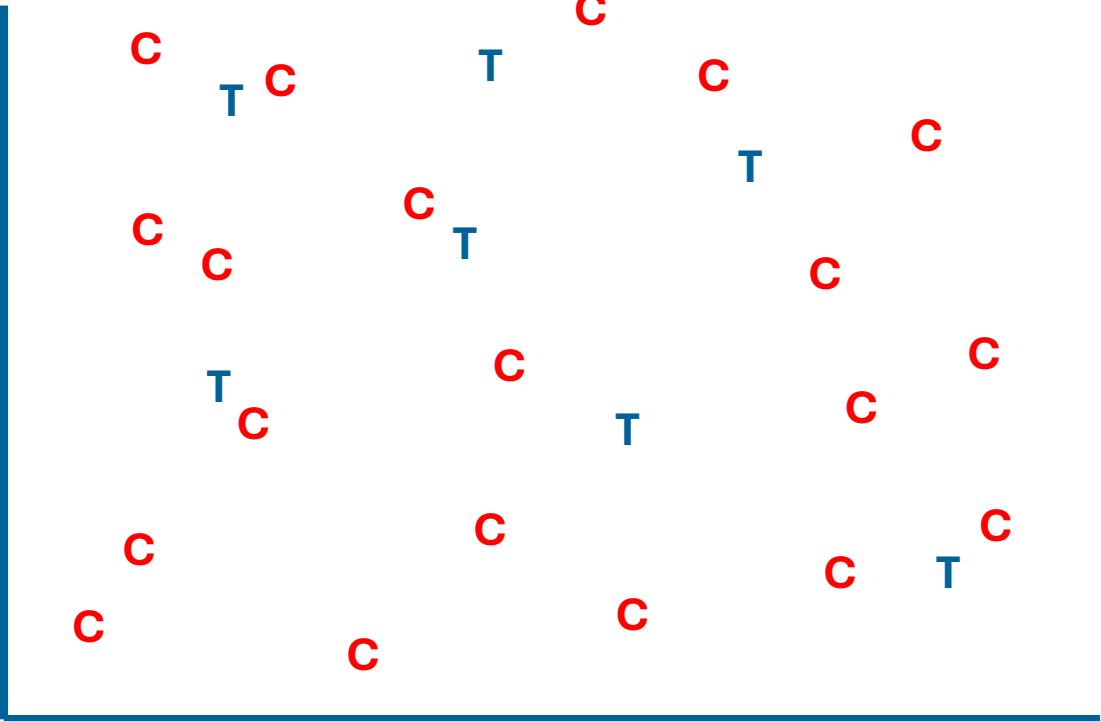


Highlight impact on assessment culture

Simple Comparison

The Problem:
Self-Selection
Bias

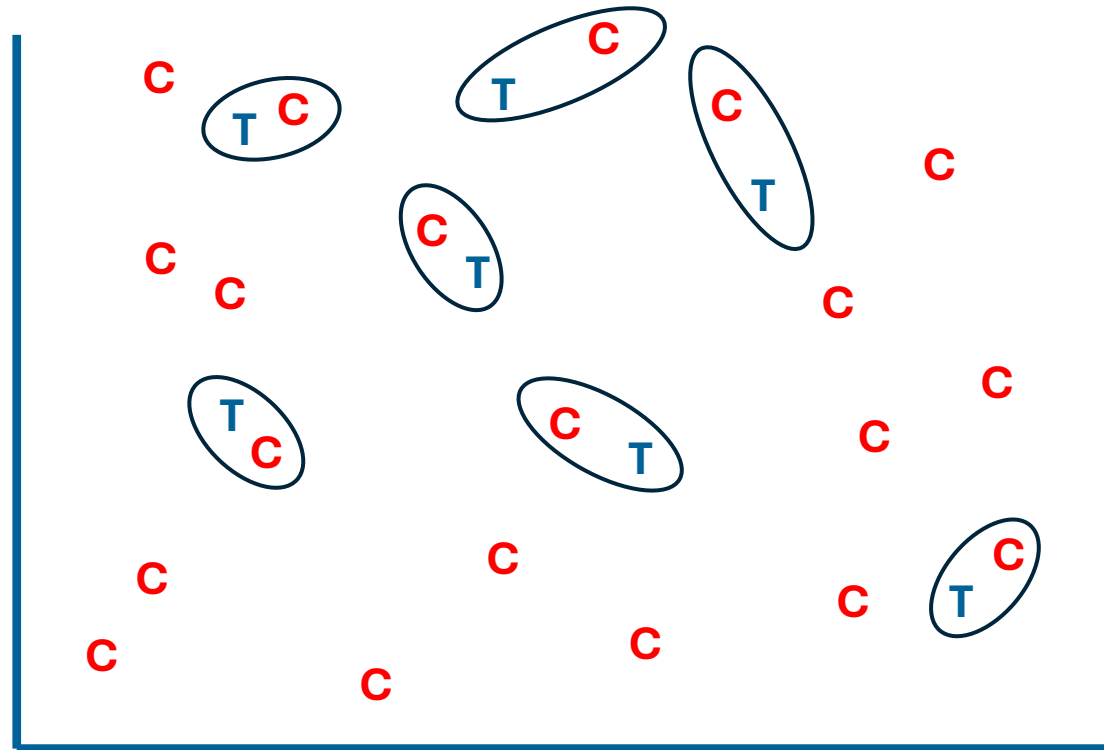
T = Treatment
C = Control



Nearest Neighbors

Match each student in treatment group with “nearest neighbor” among the control group

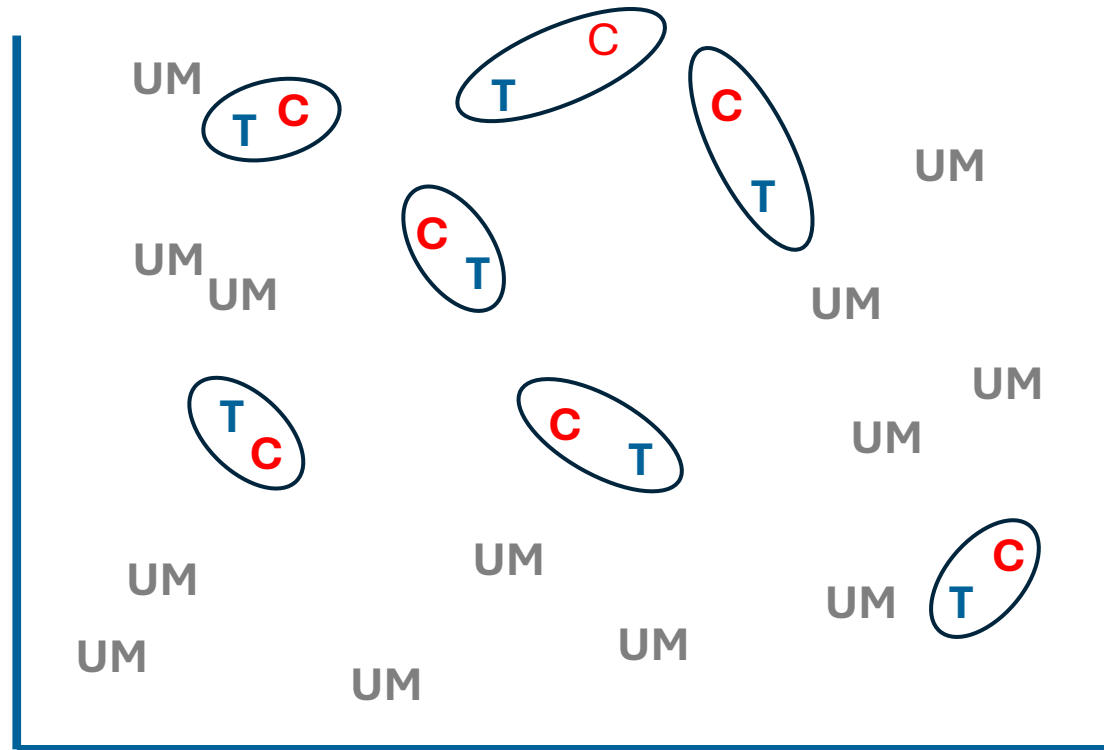
T = Treatment
C = Control



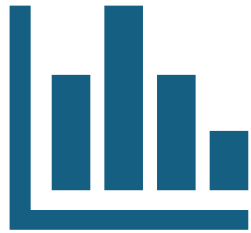
Matched Comparison

Unmatched students aren't included in final analysis

T = Treatment
C = Control
UM = Unmatched



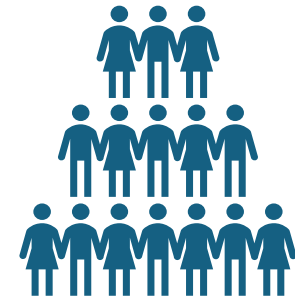
Propensity Score Matching (PSM): The Basics



Assign a **propensity score** (the student's probability of using the intervention)



Match students with similar propensity scores



Creates treatment and control groups with **reduced self-selection bias**

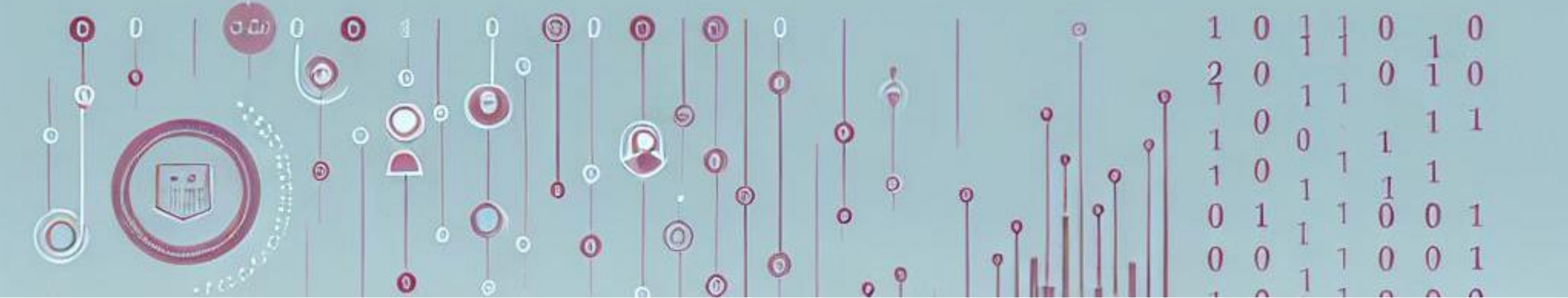
How PSM Works: The Process

1. Literature review to identify potential covariates
2. Logistic regression
 - Confirm relevant covariates
 - Estimate propensity scores
3. Match students based on propensity scores
4. Check covariate balance
5. Analyze outcomes on matched data



Case Study: Tutoring Program Evaluation

Research Question: Does tutoring participation improve student course success?

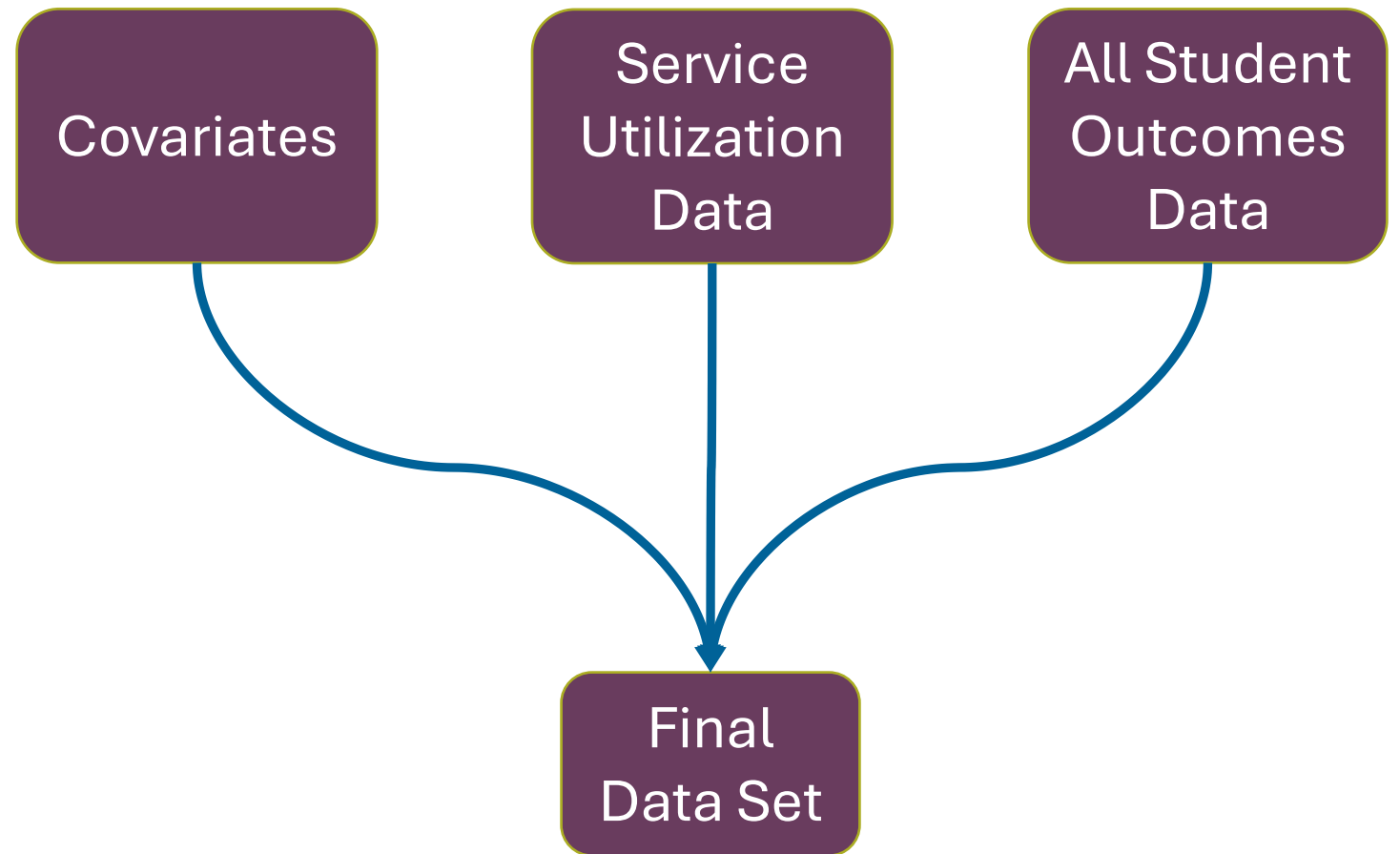


Sample Covariates

- Placement test scores
- High school GPA
- First-time student
- FYE participation
- Enrollment intensity
- Cumulative GPA
- Age
- Race/ethnicity
- Gender
- Pell recipient

Data Preparation

- Merge three streams of data
- Wide format
- Time considerations
 - Anything occurring **before** treatment is a **covariate** (e.g., high school GPA)
 - Anything occurring **after** treatment is an **outcome** (e.g., course pass rates)



Confirming Covariates: Results of Regression Analyses

Covariate	Related to Treatment Condition?	Related to Outcome Variable?	Use Covariate?
Placement test results	YES	YES	YES
High school GPA	YES	YES	YES
First-time student	NO	NO	NO
FYE Participation	YES	NO	YES
Age	YES	YES	YES
Race/Ethnicity	YES	YES	YES
Gender	YES	NO	YES
Pell-recipient	YES	NO	YES



Propensity Score Matching in SPSS

In the menu bar, select:

Data / Propensity Score Matching

Group Indicator: whether they received tutoring services or not (0 = no, 1 = yes)

Predictors: All covariates of interest

Case ID: Student ID

Match Tolerance: How similar must the propensity scores be in order to match two cases?

Naming Conventions: You must give names to the output dataset, propensity variable name, and match ID variable. These can be whatever you want, but I typically use descriptive names such as:

- **Output Dataset:**
PropensityScoresAdded
- **Propensity Variable:**
PropensityScore
- **Match ID Variable:** MatchID

Propensity Score Matching

Variables:
Sort: None

- Fiscal Year
- Full-time
- First-Generation
- Scholarship Recipient?
- Used Advising?
- EFC
- Unmet Need
- Total Expenses
- First-time Student?
- Total Tutoring Hours Yr 1
- Pass Rate

This procedure runs a logistic regression on the group indicator and then uses the resulting propensity variable to select controls for cases

Output Dataset Name (must not already exist):
PropensityScoresAdded

Dataset of Matched Controls
 Retain dataset of controls used
Dataset Name (must not already exist):

Group Indicator:
Used Tutoring Year 1?
1 indicates a case, and 0 indicates a control

Predictors:
English Placement
Math Placement
HS GPA Transformed
FYE Participation
Age
Black?
Hispanic?
Other Minority?
Male?
Pell Recipient?

Categorical predictors are converted to factors.
Propensity Variable Name (must not already exist):
PropensityScore
Match Tolerance:
0.05
Case ID:
Student ID
Match ID Variable Name(must not already exist):
MatchID

OK Paste Reset Cancel Help

Propensity Score Matching in SPSS

The steps on the previous slide create a separate data set file with two new variables: propensity scores (added for all cases) and Match IDs (added for matched cases). The file still retains all student records, so now we must remove the unmatched cases.

In the menu, navigate to:
Data / Select Cases

In the “If condition is satisfied” screen, enter the following formula:

MatchID \sim 0

Back in the Select Cases screen, under Output, check the option “Copy selected cases to a new data set” and give the new dataset a descriptive name.

That new dataset provides only the matched cases, which you should save as a new file for your records.

The image shows two overlapping SPSS dialog boxes. The background box is the main 'Select Cases: If' dialog, and the foreground box is a zoomed-in view of the 'Select Cases' dialog's 'If condition is satisfied' section.

Select Cases: If

- Function group: All Arithmetic
- Condition: `matchid ~= 0`

Select Cases

Select

- All cases
- If condition is satisfied
 - If...: `matchid ~= 0`
- Random sample of cases
 - Sample...
- Based on time or case range
 - Range...
- Use filter variable:
 -

Output

- Filter out unselected cases
- Copy selected cases to a new dataset
 - Dataset name: `Matched_Cases`
- Delete unselected cases

Current Status: Do not filter cases

Buttons: OK, Paste, Reset, Cancel, Help

Evaluate Covariate Balance

Test each covariate (whether or not it was used in the propensity score calculation) to ensure there is no statistically significant difference between treatment and control groups

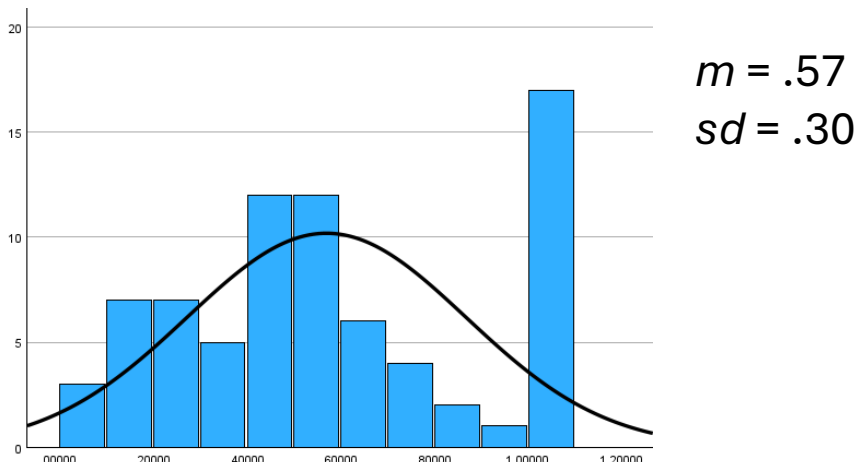
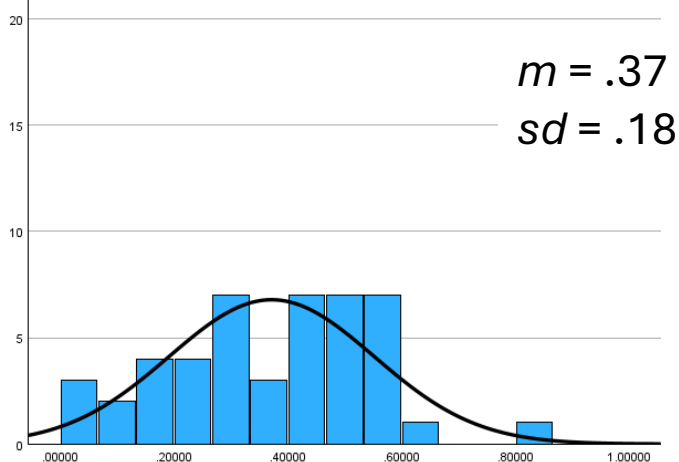
- Continuous covariates → T-tests
- Categorical covariates → Chi-square

Evaluate Improvement to Self-Selection Bias

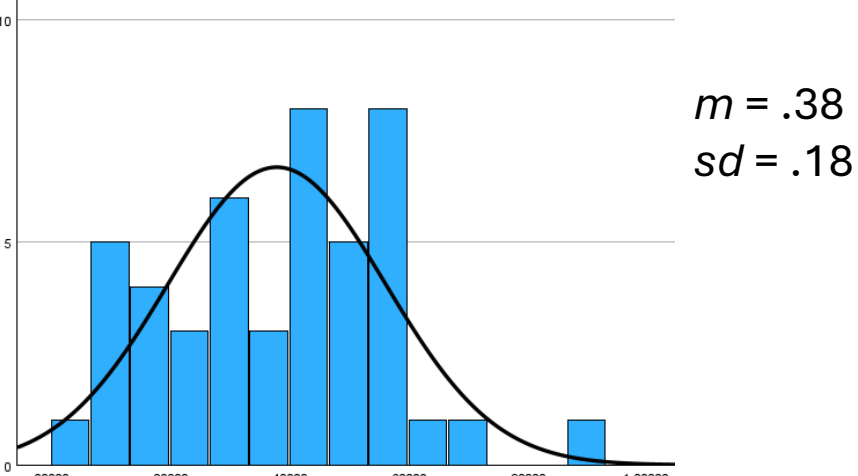
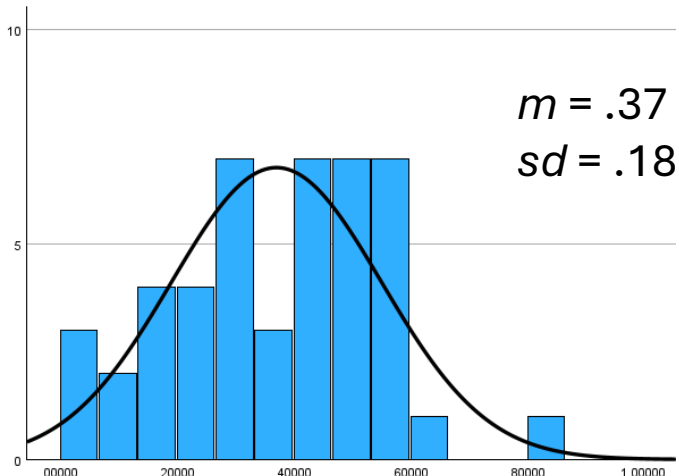
Control Group

Treatment Group

Before
Matching

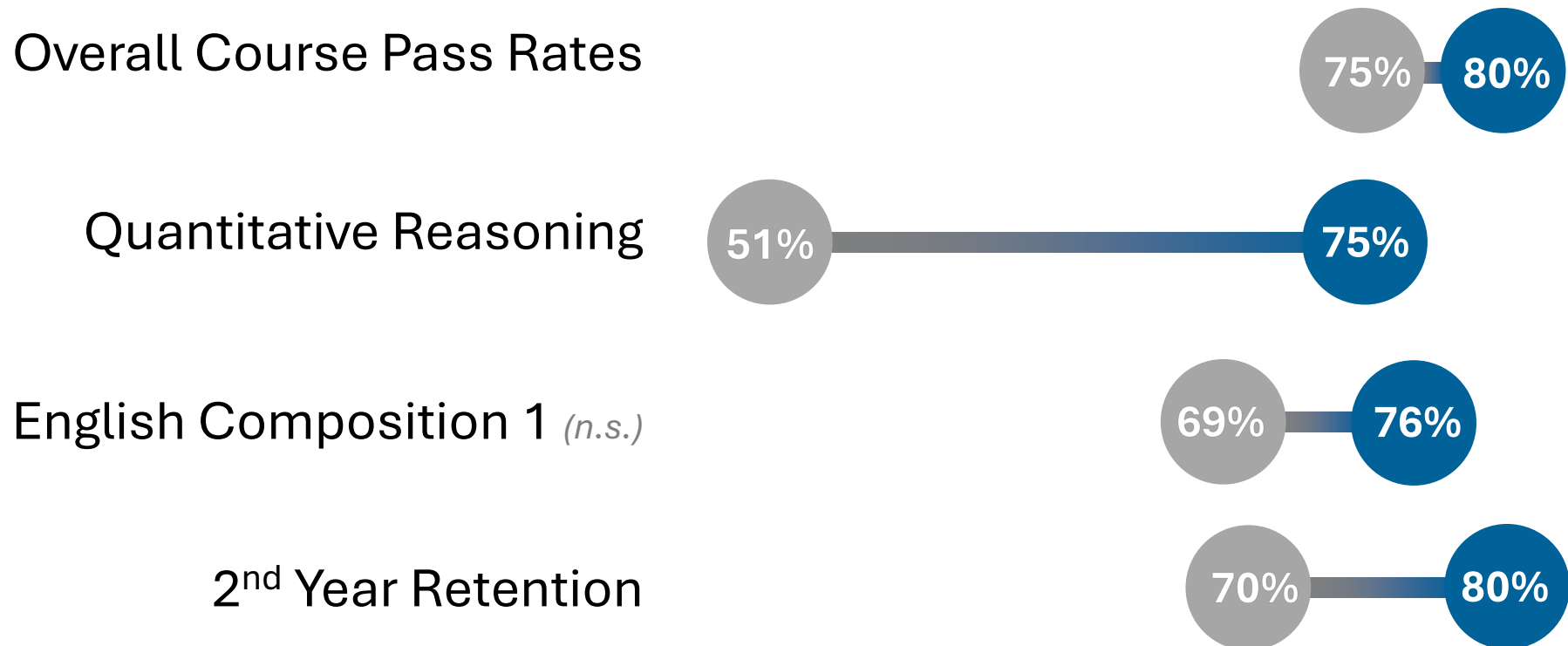


After
Matching



Goals: Histograms of propensity scores are normally distributed, visually similar with similar standard deviation and $m = 0.5$

Students who used tutoring had improved course outcomes and retention rates compared to similar peers who did not use tutoring.





Impact on Assessment Culture

Prior State

- No prior way to assess voluntary programs reliably
- Departments relied on hunches for decisions



The Change

- AIR Forum
- Post-finals statistics musings
- Tutoring asked for an outcomes evaluation
- Took the leap!



Growing Interest

- Tutoring assessment results shared college wide
 - Included a simplified methodology explanation (think cheesy clip art of apples and oranges)
- Many new requests for similar evaluations
- Organizational learning theory
 - Academic vs faculty advising
 - Inclusive access pilot

The Imperative

- **Elevate 2028 strategic plan** now prioritizes systematic assessment and data-informed decision-making
- **HLC Criterion 4:** Data-informed decision-making
- **Grant applications** and “evidence-based practice”



Challenges and Solutions

Messy service utilization data

If possible, automate ETL; try and get in the room early

Limited covariate selection

Literature review; new data warehouse; develop strong partnerships with IT

Too many data requests

Code snippet libraries; automate as much as possible

Rapid Insight/Veera Construct

The screenshot displays the Veera Construct software interface. The main workspace shows a data pipeline workflow with the following steps:

- Input 1:** Retention Rates with Covariates Added.xlsx \Export
- Process 1:** Filter
- Process 2:** Convert_2
- Input 2:** 03 - Tutoring Use Data.xlsx\Raw Data
- Process 3:** Convert
- Process 4:** Cleanse
- Process 5:** Aggregate
- Process 6:** Transpose
- Process 7:** Merge (receives input from both paths)
- Process 8:** Transform
- Process 9:** Cleanse_2
- Process 10:** Rename
- Output:** G:\My Drive\Research Projects & Presentations\Academic Assessment\TVAIRUM - M\Retention Rates with Covariates Service Utilization\Export

The left sidebar contains tool categories: Input, Compile, Process, and Output, each with various icons for data manipulation tasks.

- Jobs
 - 01_Common Code Snippets
 - 01_Covariate Code Snippet Library
 - Cumulative Hours Taken So Far
 - Demographics
 - EFC and Unmet Need
 - Emergency Grant Recipients
 - Enrollment Intensity
 - First-Semester PS Credit Hours
 - First-Time PS Students
 - FYE Course Enrollment
 - GPA - Cumulative
 - GPA - High School
 - GPA - Specific Year
 - Pell Grant Recipients
 - Placement Scores - ACT Composite
 - Placement Scores - English
 - Placement Scores - Math
 - Scholarship Recipients
 - Took Remedial English
 - Took Remedial Math
 - Used Academic Advising
 - 01_Service Utilization Data
 - 01_Student Success Indicators

Key Takeaways



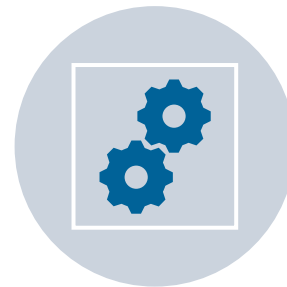
PSM reduces self-selection bias



Develop key relationships and find early adopters



Keep learning as an IR professional



Build IR capacity and automate what you can

Resources

- Bai, H., & Clark, M. H. (2019). *Propensity score methods and applications*. SAGE Publications, Inc.
- Pan, W., & Bai, H (Eds.) (2015). *Propensity score analysis: Fundamentals and developments*. The Guildford Press.
- Porter, S. (2023, August 25). *Introduction to matching and propensity score analysis* [Webinar]. Percontor Educational Research and Training.
<https://www.percontor.org/research-methods/introduction-to-matching-and-propensity-score-analysis/>
- Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. The Guildford Press.

Questions & Discussion



Thank You & Contact Info

Michelle Borckardt, borckardtm@gtc.edu

Gateway Technical College