Jaehoon Lee, PhD & Kwanghee Jung
Educational Psychology, Leadership, and Counseling
Texas Tech University

# FINDING and BUILDING DEVELOPMENTAL MATH and ACADEMIC SUPPORT SERVICES: IDENTIFYING SUCCESS and CHALLEGES for HISPANIC and LOW-INCOME STUDENTS

# BACKGROUND

Alvin Community College (ACC) proposed ***student service programs***, in collaboration with Texas Tech University (TTU), and received a 5-year DOE grant.

ACC
Your College Right Now

- ***AP #1***: Improve *academic achievement* of "Hispanic and low-income" students in ***developmental and gateway math*** courses
- ***AP #2***: Increase the *number* of "Hispanic and low-income students" who complete a ***STEM college degree***

# BACKGROUND

Alvin Community College (ACC) proposed ***student service programs***, in collaboration with Texas Tech University (TTU), and received a 5-year DOE grant.

SUCCESS THROUGH ENGAGED ADVISING

CHOOSE CONTINUE COMPLETE

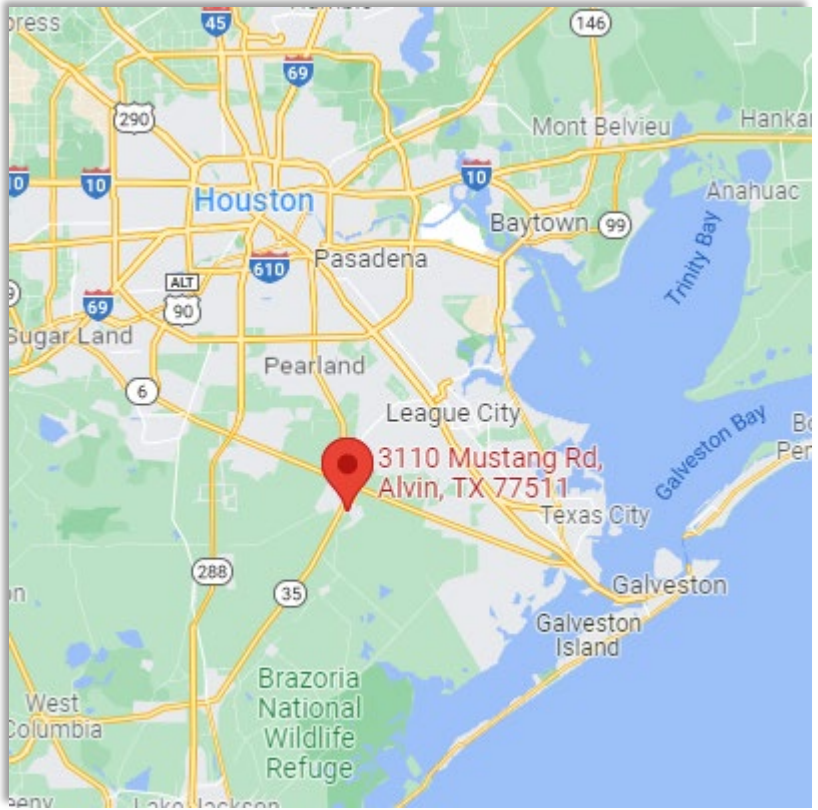
ALVIN COMMUNITY COLLEGE QUALITY ENHANCEMENT PLAN

- ***ModMath***: *Accelerated curriculum* helps students succeed in *Developmental Math* (with *MyMath Lab* and *Math Success Center*)
- ***STEM Coach***: *Retention program* provides *advising* and *tutoring* services
- ***STEM Bridge***: *Consultation* offers *career exploration & planning* and meeting with family members to review financial options and college benefits and requirements

# BACKGROUND

ACC main campus in Alvin, TX

- Only community college in a 421-sq-mile service region in the Brazoria County
- Student population: 5,293
- >70% comes from within the service region (*28% Hispanics*)
- Major cities: Alvin, Pearland, Rosharon, Angleton, Manvel



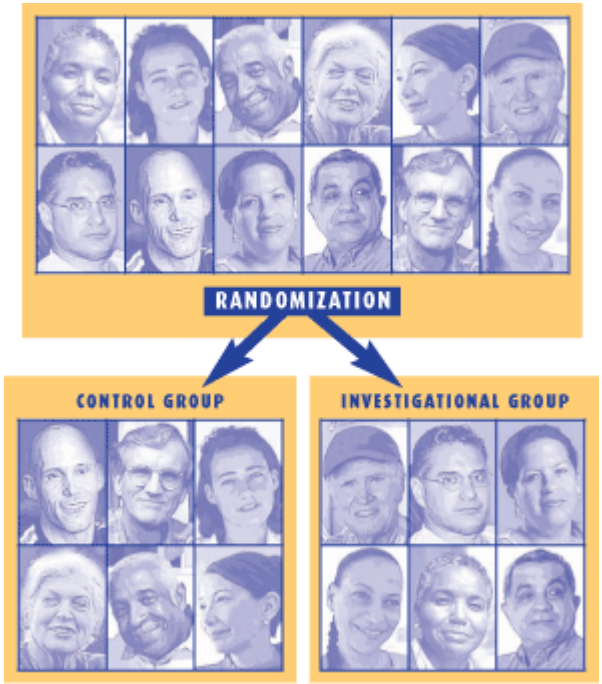
# THE TASK

TTU is in charge of ***"summative" program evaluation*** examining the effects of the *ModMath* program on students' academic achievement.


- *MATH 0310 & 0314*: *Traditional* vs. *ModMath* versions
- *3 cohorts*: AY 18-19, AY 19-20, AY 20-21;
  289 *Traditional* and 248 *ModMath* students
- *Key outcomes*: GPA, time spent for advancement (0310 → 0314 → College Algebra), degree completion, etc.

# BUT, THERE ARE ALWAYS CHALLENGES

***In an ideal setting,***




- Students are *randomly* assigned to the *Treatment* condition and the *Control* condition
  - By the virtue of ***random assignment***, the two condition groups are "comparable" at baseline
  - Thus, we can make a ***causal inference*** that any observed differences between the two groups are solely due to the *Treatment*

# BUT, THERE ARE ALWAYS CHALLENGES

***In reality,***

- Students were *not randomized*. Rather, they (unwittingly) "elected" the *ModMath* version of MATH 0310 & 0314 or the *Traditional* version of the courses
  - *ModMath* students and *Traditional* students may be considerably dissimilar in some personal characteristics — ***selection bias***
  - Thus, when the two groups show different academic achievement, this could be due to the *ModMath* program, personal factors, *or both*

# BUT, THERE ARE ALWAYS CHALLENGES

***In reality,***

- Students were *not randomized*. Rather, they (unwittingly) "elected" the *ModMath* version of MATH 0310 & 0314 or the *Traditional* version of the courses

- Required to address ***WWC Standards*** for *baseline equivalence*

| 0.00 ≤ ES Difference ≤ 0.05 | 0.05 < ES Difference ≤ 0.25 | ES Difference > 0.25 |
|---|---|---|
| Satisfies baseline equivalence | Statistical adjustment required to satisfy baseline equivalence | Does not satisfy baseline equivalence |

# CHALLENGE ACCEPTED.

# PSM

***A solution: Propensity score methods***

- In real settings, it is often *infeasible* or *unethical* to randomly assign people into different (*Treatment* and *Control*) conditions
    - New drug testing for acute cancer

- In such case, ***propensity score methods*** are useful to account for possible *selection bias* and thereby allow us for addressing questions of *causal inference*

# PSM

***Propensity score*** (Rosenbaum & Rubin, 1983)

- "How *likely* does a person receive or select the treatment (*T*) *given* his/her personal characteristics (*X*) at baseline?"

$$P_i\left(T_i \middle| X_i\right)$$

# PSM

***Propensity score*** (Rosenbaum & Rubin, 1983)

- PS exists *both* in *randomized* trials and in *observational* studies
- In *randomized* trials, the "true" PS is *known and equal* for all individuals (e.g., 0.5 in coin toss)

# PSM

***Propensity score*** (Rosenbaum & Rubin, 1983)

- In *observational* studies, the "true" PS is *unknown*

$$\hat{P}_i\left(T_i \middle| X_i\right)$$

  - People already in the *Treatment* and *Control* conditions
  - PS is *estimated* for each person using his/her "actual" *treatment status* (*T*) and values on the *covariates* (*X*) measured at baseline
  - PSM utilize this *conditional probability* to "recreate" a situation that *would have been expected* in a *randomized* trial

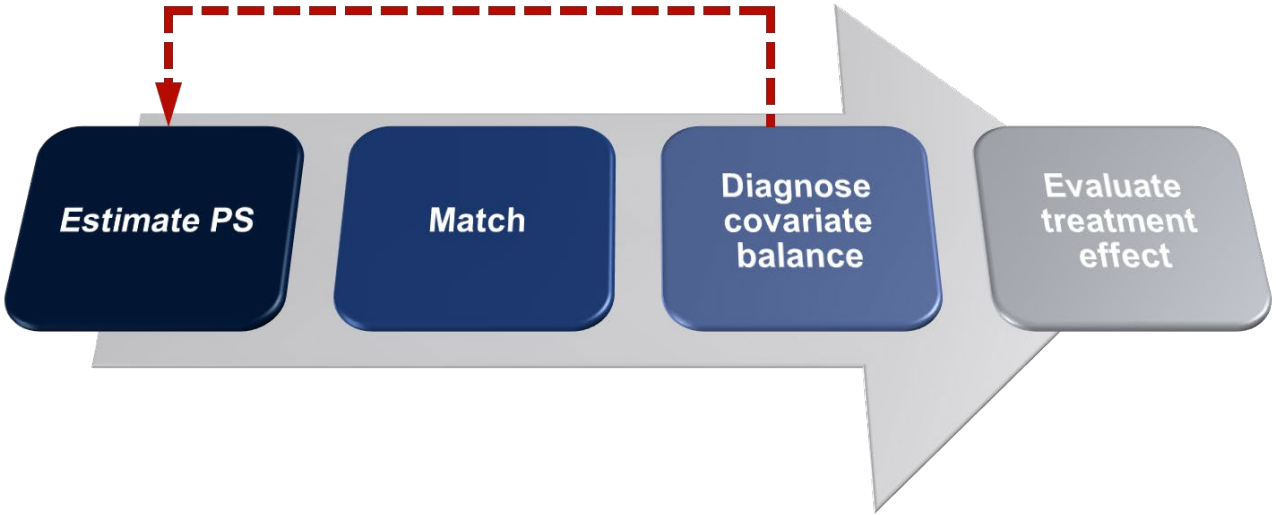
# PSM

***4 popular methods*** (Austin, 2009; Rosenbaum, 2002; Rosenbaum & Rubin, 1983)

- ***Matching*** treated *persons* with untreated *persons*
- ***Weighting*** *data*
- ***Stratifying*** *sample*
- ***Adjusting*** *parameter estimates*

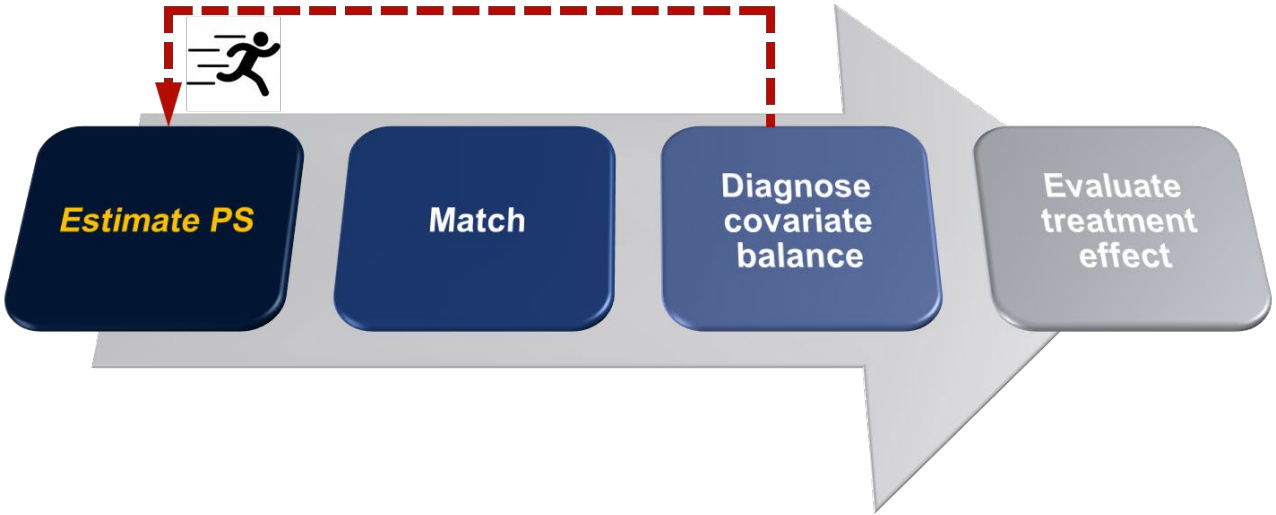
# PROCESS

***Propensity score matching***




- Iteratively check *balance* on the *covariates* (*X*) between treated persons and untreated persons in the *"matched" sample*

# ESTIMATING PS



# ESTIMATING PS

***2 popular methods of estimating PS***

- *Parametric*: ***Logistic regression***
- *Non-parametric*: ***Generalized boosted modeling***

# ESTIMATING PS: LR

***Logistic regression***

$$\hat{P}_i\left(T_i \middle| X_i\right)$$

$$\Leftrightarrow \ln\left(\frac{P_i(T_i=1)}{1-P_i(T_i=1)}\right) = \beta_0 + \beta_1\left(X_{1i}\right) + \beta_2\left(X_{2i}\right) + \ldots + \beta_n\left(X_{ni}\right) + e_i$$

$T_i$ = "actual" *treatment status* (1 = treatment, 0 = no treatment)

$X_i$, ... , $X_{ni}$ = values on the *covariates* measured at baseline

# ESTIMATING PS: LR

***Alvin ModMath Study***

| ***Variable*** | ***Name*** | ***Value*** |
|---|---|---|
| Treatment status | *treated* | *ModMath* (1), *Traditional* (0) |
| DOB | *dob* | yyyy-mm-dd |
| Gender | *gender* | Female / Male |
| Race / Ethnicity | *race* | *Asian, Black, Hispanic, White* |
| Financial aid | *aid* | Pell / No |
| Grade | *grade1* | A=4, B=3, C=2, D=1, F=0 |
| | *grade2* | Pass (1) / Fail (0) |

# ESTIMATING PS: LR


## *Logistic regression*

```
> param <- matchit(treated ~ dob + gender + race + aid,
                   family=binomial, data=dat)
> param$distance
```

| | SID | DOB | GENDER | Ethnicity | RACE | AID | TREATED | ps |
|---|---|---|---|---|---|---|---|---|
| 1 | 99144 | 1978-05-24 | Female | Non Hispanic/Latino | White | PELL | 0 | 6.249817e-01 |
| 2 | 812172 | 1983-12-28 | Female | Hispanic/Latino | Hispanic | No | 0 | 3.941388e-01 |
| 3 | 929284 | 1999-06-03 | Male | Non Hispanic/Latino | White | No | 1 | 5.015871e-01 |
| 4 | 1022256 | 1994-04-23 | Female | Non Hispanic/Latino | White | No | 1 | 4.214361e-01 |
| 5 | 1055844 | 2001-03-13 | Male | Non Hispanic/Latino | White | No | 1 | 4.945763e-01 |

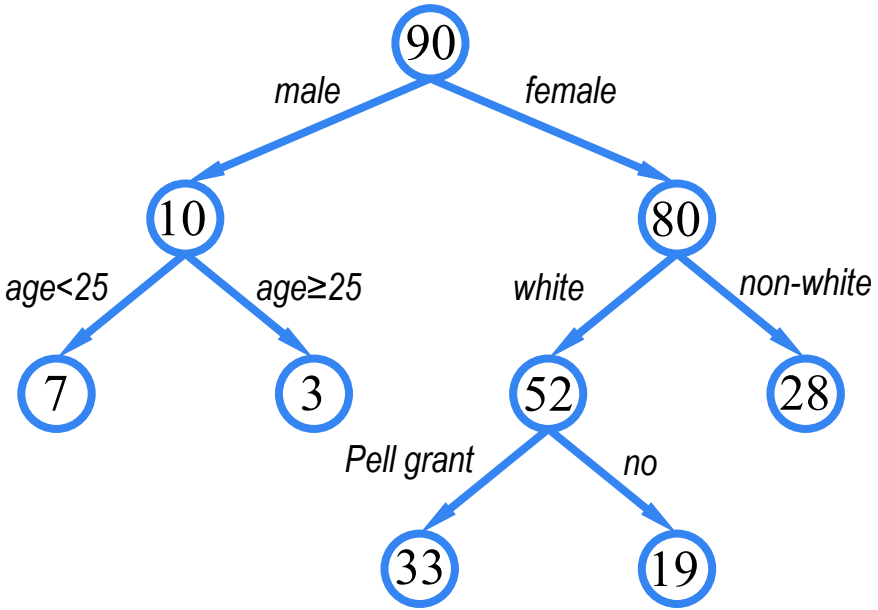
# ESTIMATING PS: GBM

***Generalized boosted modeling*** allows for multiway *product terms* modeled "naturally" as a result of sample splitting. (Friedman, 2001; McCaffrey, Ridgeway, & Morral, 2004)

- ***Step 1***: Randomly select 50% of the sample — "training data".
- ***Step 2***: Predict *treatment status* using ***Classification and Regression Trees (CART)***.

# ESTIMATING PS: GBM

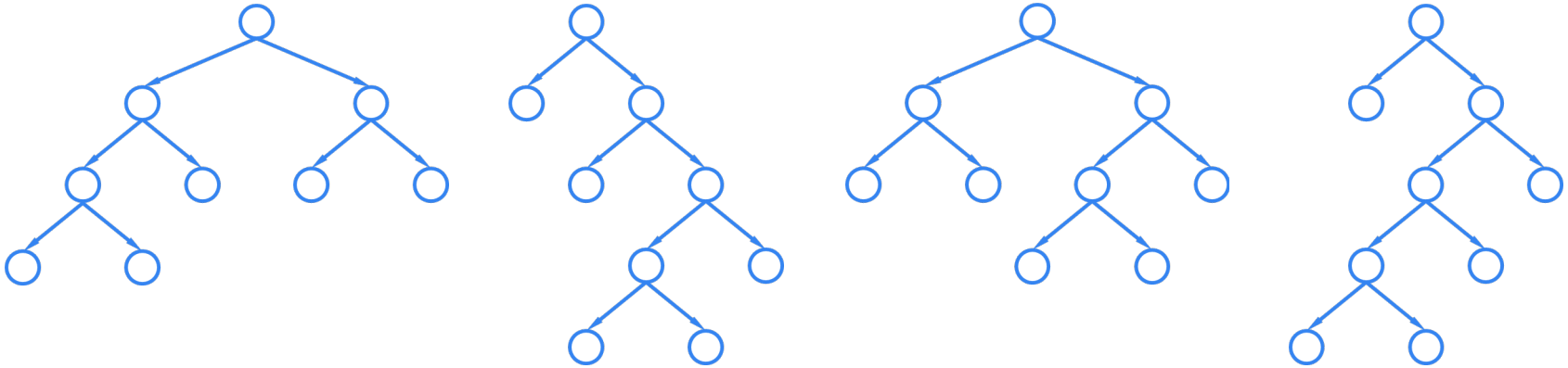
- ***Step 2***: Predict *treatment status* using ***CART***.




- The sample is split by the covariate that, among all covariates, best predicts treatment status
- The difference between "estimated" PS and "actual" treatment status — *residual* — is computed within each split subset
- Additional splits are made by predicting the *residual* with the remaining covariates

# ESTIMATING PS: GBM

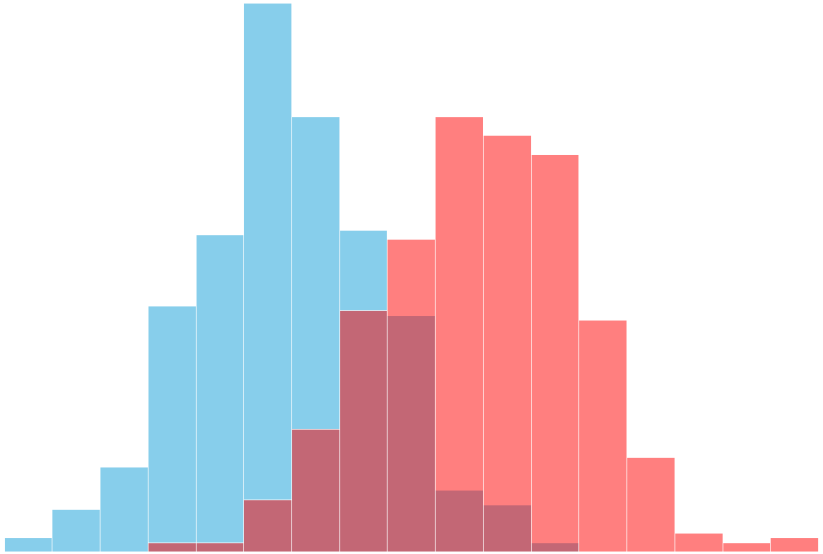
- ***Step 3***: Many trees are formed by repeating ***Steps 1 & 2***.



- ***Step 4***: The trees are combined together to calculate a final PS estimate for each person in the sample.

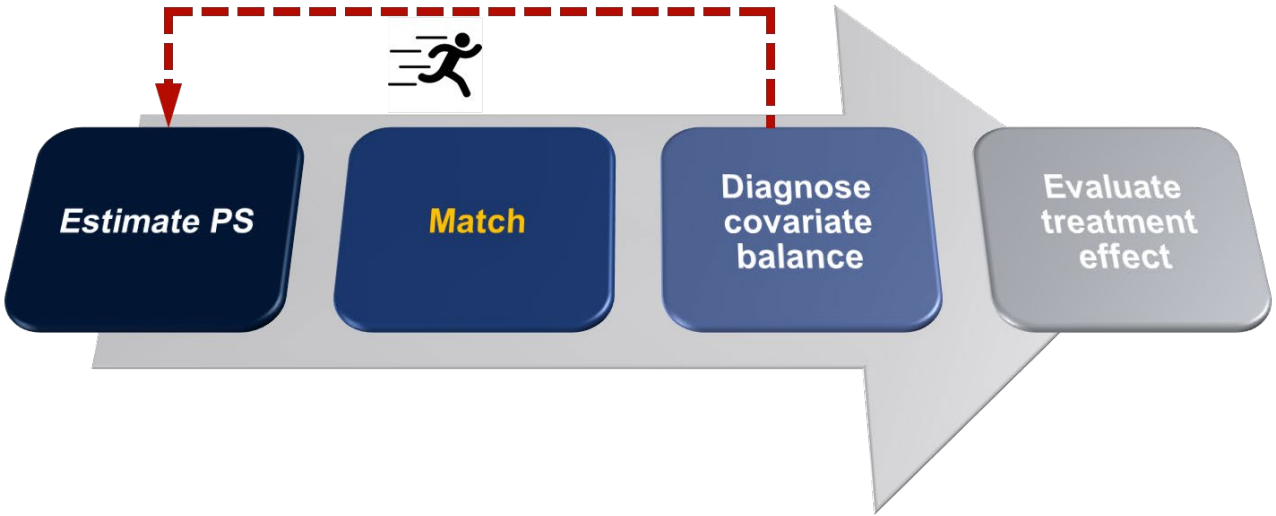
# ESTIMATING PS: COMMON SUPPORT

After estimating PS, ensure that there is a "substantial" *overlap* in PS between *treatment* persons and *untreated* persons — "*common support*".



- A large area of common support increases the confidence that the observed treatment effect can be *generalized* to the entire *population* being represented by the sample.

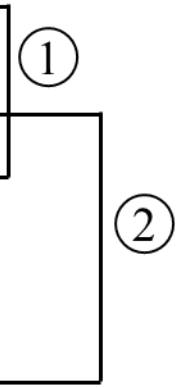
# MATCHING


Estimate PS
Match
Diagnose covariate balance
Evaluate treatment effect

# NEAREST NEIGHBOR MATCHING

In ***nearest neighbor matching***, a treated person is matched to an untreated person if their PS are *most similar* — in the "smallest" distance.

| ID | Treated | PS |
|---|---|---|
| 1 | 1 | .57 |
| 2 | 1 | .36 |
| 3 | 0 | .54 |
| 4 | 0 | .60 |
| 5 | 0 | .17 |
| ⋮ | ⋮ | ⋮ |

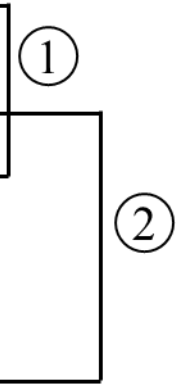



- Find a match for treated persons, one by one

# NEAREST NEIGHBOR MATCHING

In ***nearest neighbor matching***, a treated person is matched to an untreated person if their PS are *most similar* — in the "smallest" distance.

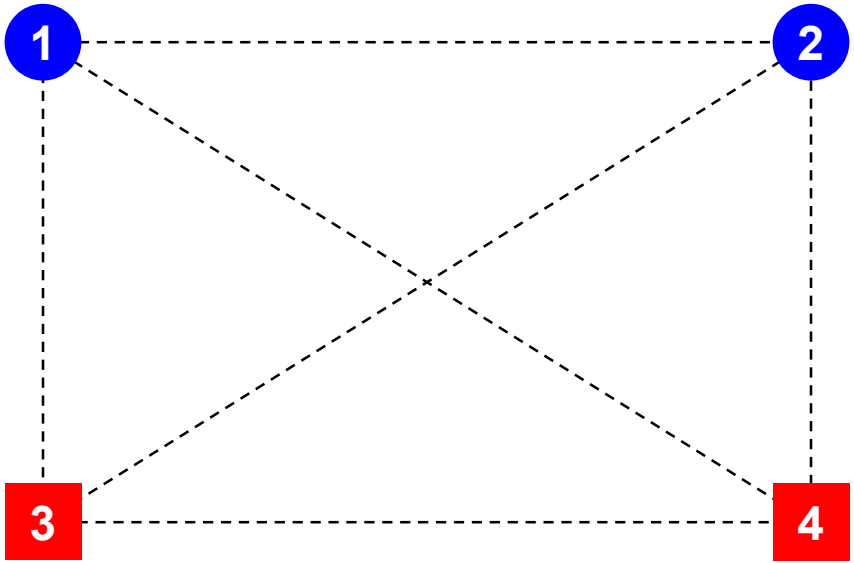
| ID | Treated | PS |
|---|---|---|
| 1 | 1 | .57 |
| 2 | 1 | .36 |
| 3 | 0 | .54 |
| 4 | 0 | .60 |
| 5 | 0 | .17 |
| ⋮ | ⋮ | ⋮ |




- Often the PS is *not* close for possible pairs
- To avoid bad matches, define a "*caliper*" — the *maximum* distance in PS by which matches are allowed
- *0.25 x standard deviation of the logit of PS*
  (Rosenbaum & Rubin, 1985)

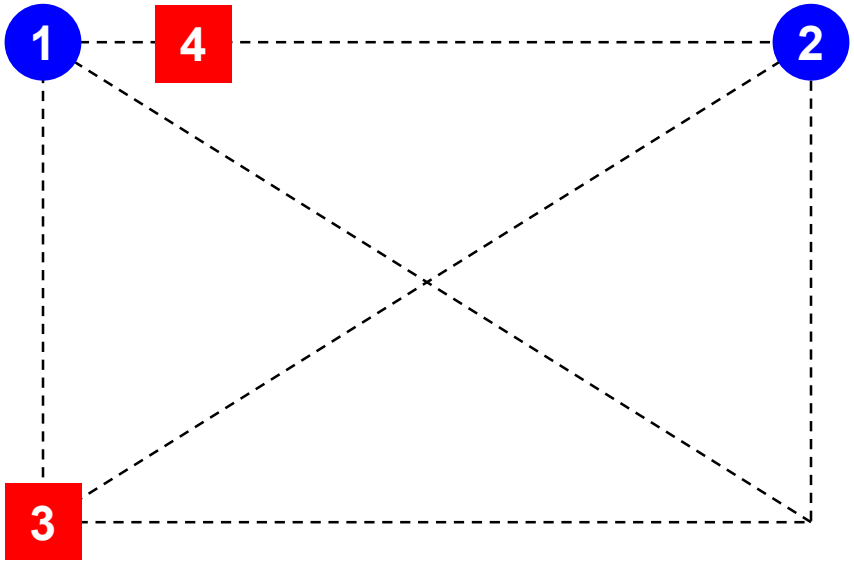
# OPTIMAL MATCHING

In ***optimal matching***, matches are formed by *minimizing* the *global distance* in PS, defined as the sum of PS distances in the whole matched sample.



# OPTIMAL MATCHING

In ***optimal matching***, matches are formed by *minimizing* the *global distance* in PS, defined as the sum of PS distances in the whole matched sample.



# MATCHING

## *Nearest neighbor matching*

```
> matchit(treated ~ dob + gender + race + aid,
          data=dat,
          method="nearest", caliper=0.25)
```

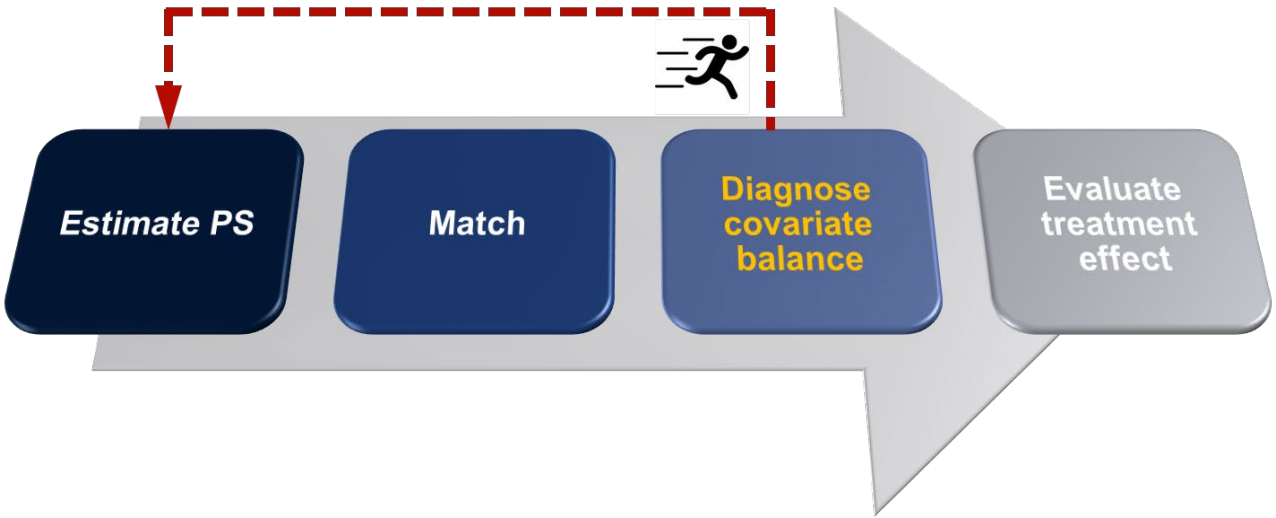
```
Sample Sizes:
          Control Treated
All           105      91
Matched        82      82
Unmatched      23       9
Discarded       0       0
```

## *Optimal matching*

```
> matchit(treated ~ dob + gender + race + aid,
          data=dat,
          method="optimal"
```

```
Sample Sizes:
          Control Treated
All           105      91
Matched        91      91
Unmatched      14       0
Discarded       0       0
```

# BALANCE DIAGNOSTICS



# BALANCE DIAGNOSTICS

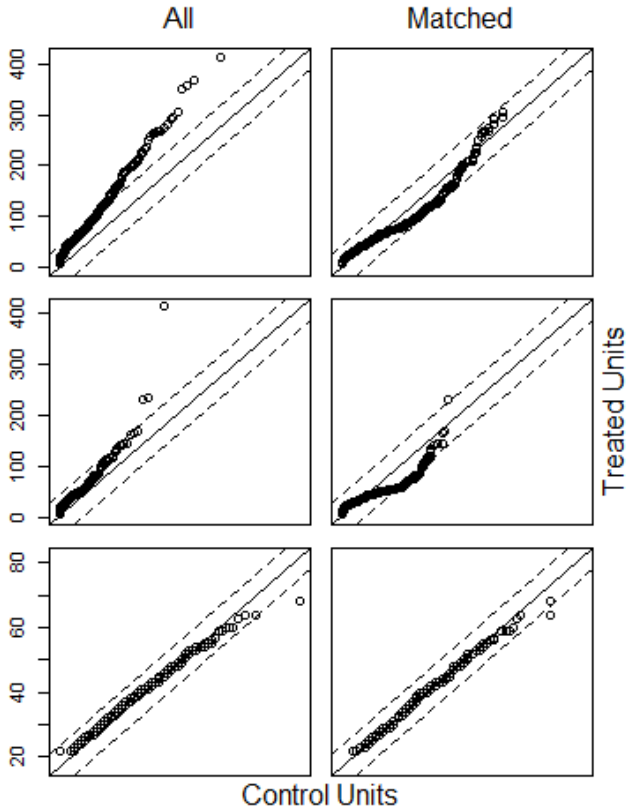
Once a matching is successfully implemented, the next step is to examine if *balance* is made on the covariates. (Austin, 2009; Flury & Riedwyl, 1986)

- Inspection of *distributions* — ***Q-Q plot***
- *Standardized difference* in *means* — ***Cohen's d***

# BALANCE DIAGNOSTICS: Q-Q PLOT

***Q-Q plot***

- The *distribution* of a covariate in the *Treatment* group is plotted against the distribution in the *Control* group
- *Deviations from a 45-degree line* indicate that the distributions are dissimilar.



# BALANCE DIAGNOSTICS: D

***Cohen's d***

- Continuous covariates:

$$\frac{\left(\bar{X}_T - \bar{X}_U\right)}{\sqrt{\frac{s_T^2 + s_U^2}{2}}}$$

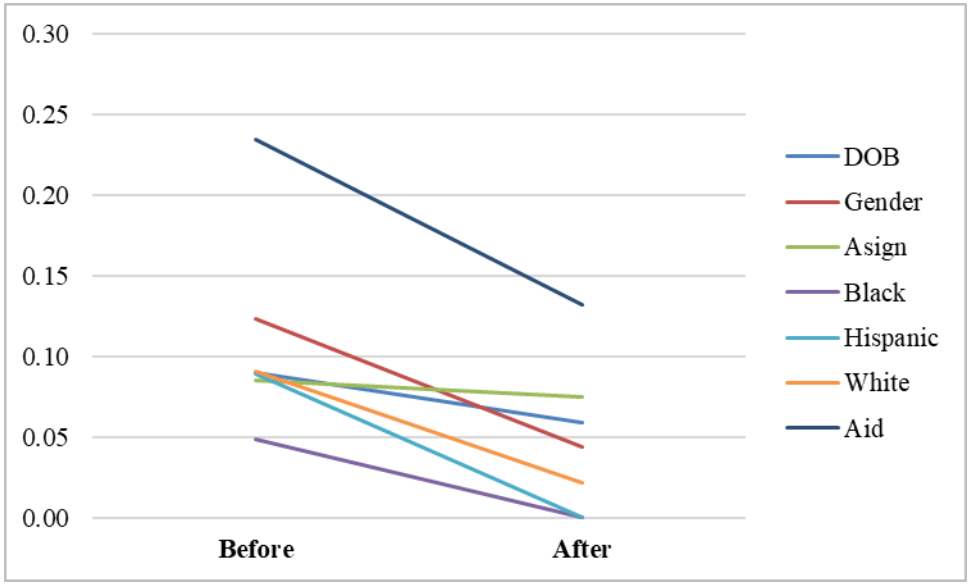
Binary covariates:

$$\frac{\left(\hat{P}_T - \hat{P}_U\right)}{\sqrt{\frac{\hat{P}_T\left(1-\hat{P}_T\right) + \hat{P}_U\left(1-\hat{P}_U\right)}{2}}}$$

  - A covariate with *n-categories* is *dichotomized* into *n* variables (by dummy-coding) and then examined for balance

# BALANCE DIAGNOSTICS: D

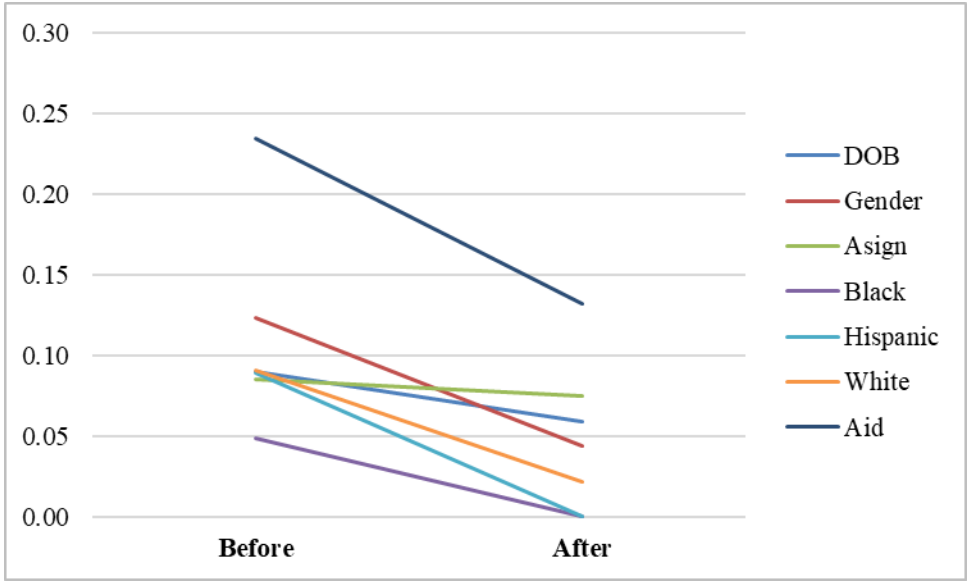
## *Cohen's d*




| Percent Balance Improvement: | Std. Mean Diff. |
|---|---|
| distance | 63.5 |
| DOB | 34.2 |
| GENDERFemale | 64.3 |
| GENDERMale | 64.3 |
| RACEAmerican/Alaska Native | 100.0 |
| RACEAsian | 11.8 |
| RACEBlack or African American | 100.0 |
| RACEHispanic | 100.0 |
| RACERace/Ethnicity Unknown | 100.0 |
| RACEWhite | 75.4 |
| AIDNo | 43.7 |
| AIDPELL | 43.8 |

# BALANCE DIAGNOSTICS: D

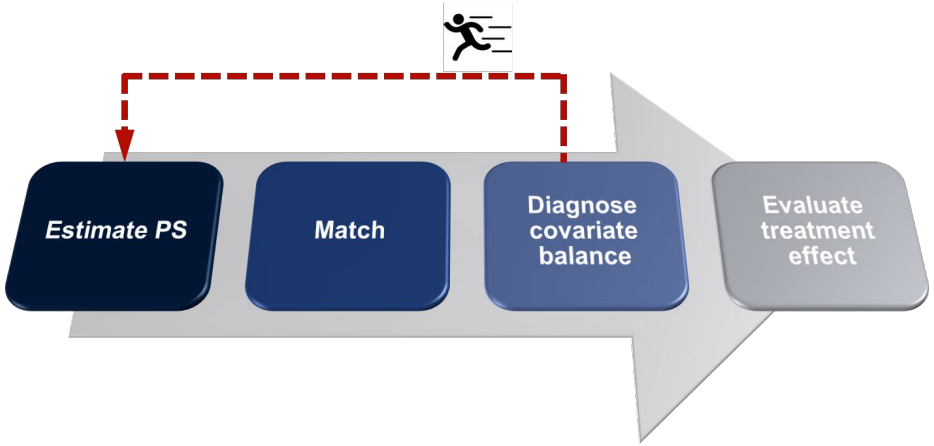
## *Cohen's d*




| Percent Balance Improvement: | Std. Mean Diff. |
|---|---|
| distance | 63.5 |
| DOB | 34.2 |
| GENDERFemale | 64.3 |
| GENDERMale | 64.3 |
| RACEAmerican/Alaska Native | 100.0 |
| RACEAsian | 11.8 |
| RACEBlack or African American | 100.0 |
| RACEHispanic | 100.0 |
| RACERace/Ethnicity Unknown | 100.0 |
| RACEWhite | 75.4 |
| AIDNo | 43.7 |
| AIDPELL | 43.8 |

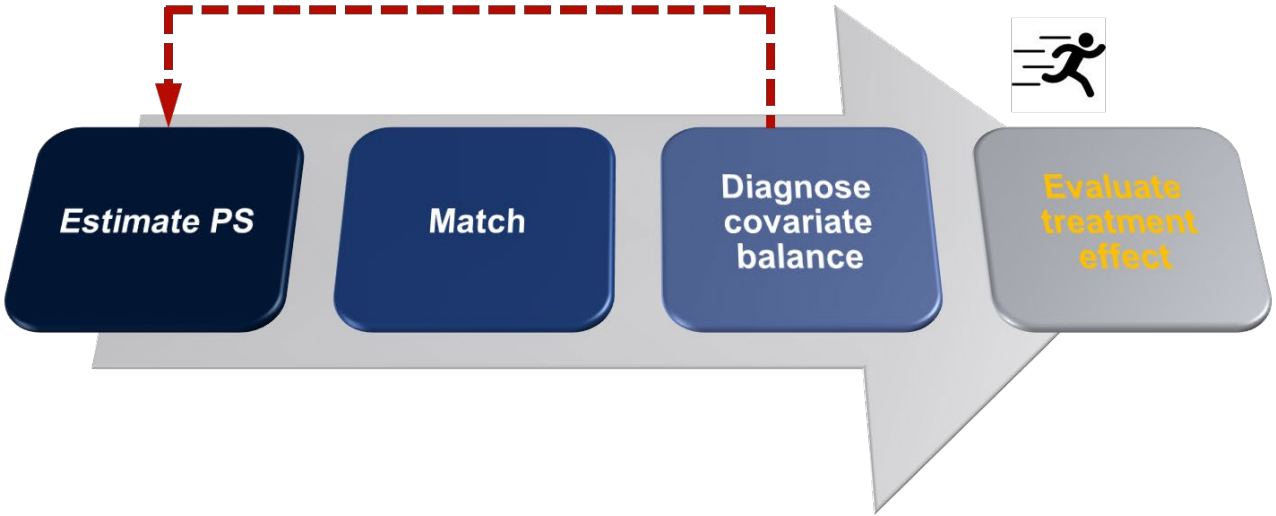
| 0.00 ≤ ES Difference ≤ 0.05 | 0.05 < ES Difference ≤ 0.25 | ES Difference > 0.25 |
|---|---|---|
| *Satisfies baseline equivalence* | *Statistical adjustment required to satisfy baseline equivalence* | Does not satisfy baseline equivalence |

# BALANCE DIAGNOSTICS




- If an *imbalance* is indicated by dissimilar distribution and/or nontrivial *d*,
  - *Transform* or *re-categorize* the unbalanced covariates
  - Add *polynomial terms* of the unbalanced covariates
  - Add *product terms* of the unbalanced covariates and other covariates
  - Use a *smaller caliper* (in case of nearest neighbor matching)

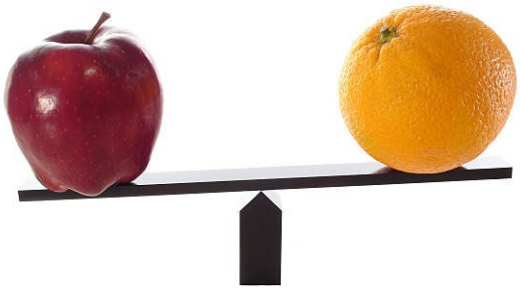
# ESTIMATING TREATMENT EFFECT



# ESTIMATING TREATMENT EFFECT

***Bivariate tests***

- Estimate the *treatment effect* by comparing outcomes between treated persons and untreated persons in the *matched sample*



  - *Continuous* → difference in *means*
  - *Categorical* → difference in *proportions*
  - *Binary* → difference in *probabilities* (relative risk, odds ratio)

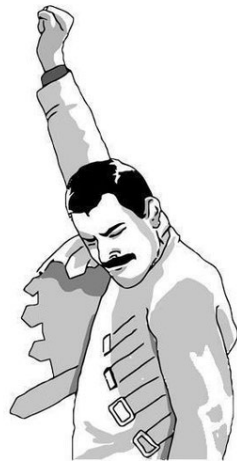
# ESTIMATING TREATMENT EFFECT

***Bivariate tests***

- Estimate the *treatment effect* by comparing outcomes between treated persons and untreated persons in the *matched sample*

  - *Continuous* → difference in *means* → *t*-test
  - *Categorical* → difference in *proportions* → chi-square/Fisher test
  - *Binary* → difference in *probabilities* → chi-square test

$p<0.05$



# ESTIMATING TREATMENT EFFECT

***t-test***

**AY 2020-21 MATH 314**

| *Outcome* | *ModMath* | *Traditional* | *p* | *d* |
|---|---|---|---|---|
| Coaching sessions | 3.8 ± 2.4 | 3.0 ± 2.6 | 0.014 | 0.309 |
| Coaching hours | 1.8 ± 1.2 | 1.5 ± 1.3 | 0.019 | 0.295 |
| Tutoring sessions | 8.0 ± 10.7 | 4.3 ± 8.9 | 0.003 | 0.380 |
| Tutoring hours | 8.5 ± 14.5 | 4.8 ± 11.4 | 0.024 | 0.292 |

- Students in the *ModMath* course (voluntarily) received "significantly" *more coaching and tutoring* than those in the *Traditional* course.

# ESTIMATING TREATMENT EFFECT


***Multivariate tests***

- *Multivariate* analysis is also applicable, of course
  - Linear, logistic, or Poisson regression
  - Survival analysis (e.g., time to transfer to 4-year university)
  - Structural equation modeling
  - Hierarchical linear modeling
- The outcome models can include the covariates used for estimating PS, so as to further *eliminate* *residual imbalance* in "prognostically" important covariates (Harder et al., 2010; Ho et al., 2007)

# ESTIMATING TREATMENT EFFECT

***Concluding remarks***

- *Selection bias* is a major threat to the *validity* of any observational study

- ***PS methodology*** offers researchers an integrative framework where…

  - Not only *"overt" bias* from the *measured covariates* can be corrected in the estimates of the treatment effects,

  - But also *"hidden" bias* from *unmeasured covariates* can be evaluated in terms of robustness of the effect estimates — ***sensitivity analysis***

Behaviour Research and Therapy 98 (2017) 76–90

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

# A practical guide to propensity score analysis for applied clinical research

Jaehoon Lee*, Todd D. Little

*Department of Educational Psychology and Leadership, College of Education, Texas Tech University, United States*

ARTICLE INFO

*Article history:*
Received 22 September 2016
Received in revised form
10 January 2017
Accepted 12 January 2017
Available online 19 January 2017

*Keywords:*
Propensity score
Matching
Subclassification
Weighting
R

ABSTRACT

Observational studies are often the only viable options in many clinical settings, especially when it is unethical or infeasible to randomly assign participants to different treatment régimes. In such case propensity score (PS) analysis can be applied to accounting for possible selection bias and thereby addressing questions of causal inference. Many PS methods exist, yet few guidelines are available to aid applied researchers in their conduct and evaluation of a PS analysis. In this article we give an overview of available techniques for PS estimation and application, balance diagnostic, treatment effect estimation, and sensitivity assessment, as well as recent advances. We also offer a tutorial that can be used to emulate the steps of PS analysis. Our goal is to provide information that will bring PS analysis within the reach of applied clinical researchers and practitioners.

# QUESTIONS?

jaehoon.lee@ttu.edu
kwanghee.jung@ttu.edu