11. Matching Estimators

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Credit: Created using Pixiz (https://en.pixiz.com/template/It-s-a-Match-Tinder-mockery-with-customizable-text-3150)

Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates
  - Units with opposite value of some instrument
  - Exploit two possible sources of variation for identification!
    - Exploit **cross-sectional** variation in treatment.
    - Exploit variation in treatment **within a unit over time** (before/after)
  - At a discontinuity in treatment assignment (will cover in W13)
- Can we make our identifications strategies work better?
  - $\rightsquigarrow$ matching or weighting

The Problem with Regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{\text{ATT}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

The Problem with Regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

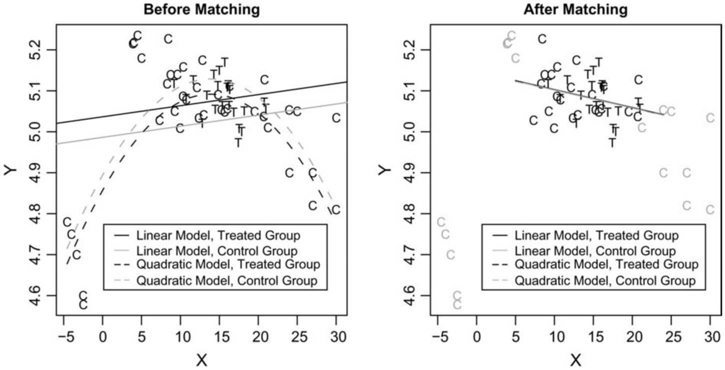
$$\tau_{\text{ATT}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression (ECI W5).

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i(Y_i - \widehat{\mu}_0(\mathbf{X}_i))$$

- Common solution: use a parametric model for $\widehat{\mu}_0(\mathbf{X}_i)$
    - For example, could assume it is linear: $\mu_0(\mathbf{x}) = \mathbf{x}'\beta$
    - Regression, MLE, Bayes, etc.
    - But this model might be wrong $\rightsquigarrow$ wrong causal estimates.

Model Dependence



Source: *Figure 1* in Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." Political analysis 15, no. 3 (2007): 199-236.

What is Matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$
  - Matching? = pruning bad controls (King, 2020)

What is Matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$
  - Matching? = pruning bad controls (King, 2020)

- Matching has strong advantages:
  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.
  3. Makes counterfactual comparisons more transparent.

What is Matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$
  - Matching? = pruning bad controls (King, 2020)

- Matching has strong advantages:
  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.
  3. Makes counterfactual comparisons more transparent.

- What matching isn't: a solution for selection on unobservables.
  - Matching is an **estimation** technique, not an identification strategy.

Types of Matching

- Assumptions:
    - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
    - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$

Types of Matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$

- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$.
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution in treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0,\ j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j(0) \mid D_j = 0,\ j \text{ is matched}]$$

  - Issue: not feasible with high-dimensional or continuous $\mathbf{X}_i$

Types of Matching

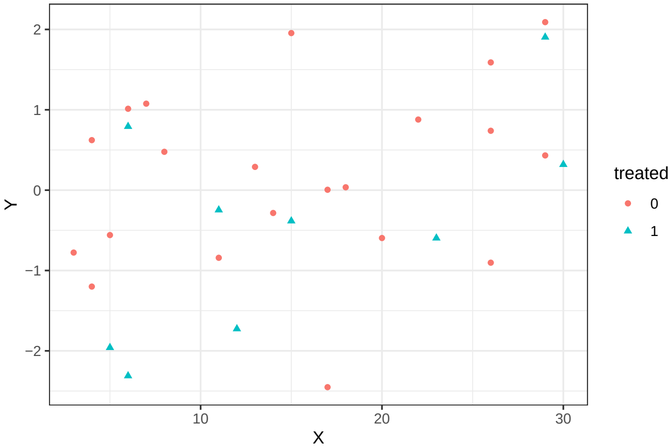
- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$.
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution in treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0,\ j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j(0) \mid D_j = 0,\ j \text{ is matched}]$$

  - Issue: not feasible with high-dimensional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)
  - Discretize and group covariates into substantively meaningful bins
  - Exact match on these bins $\rightsquigarrow$ accounts for interactions
  - Have to drop treated units in bins with no controls $\rightsquigarrow$ changes estimand
  - Allows you to control bias/variance tradeoff through coarsening

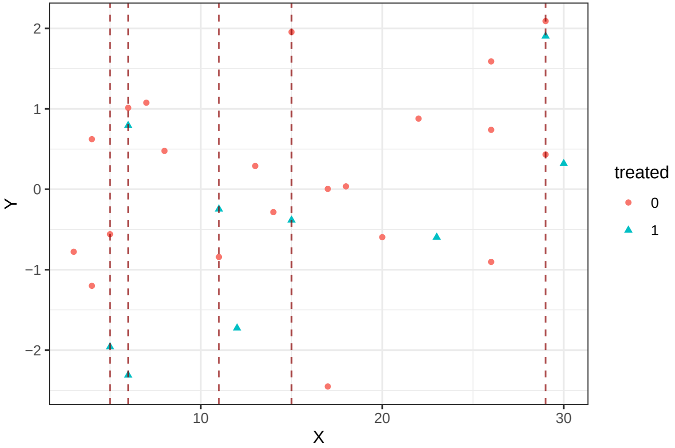
Exact Matching Illustrated

- How would we implement exact matching?



Exact Matching Illustrated

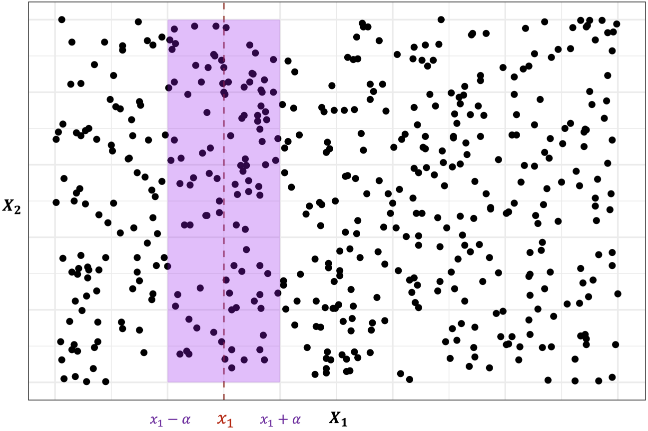
- Only keep data with **identical** X_is!



- What to do when we have continuous $\mathbf{X}_i$?

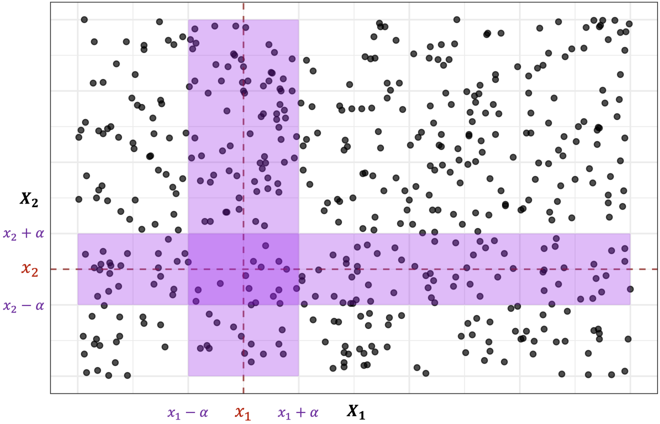
Curse of Dimensionality

- Often, we will have multiple covariates to match on:
  - Total observations in the raw data = 500
  - Remaining observations = 105



Curse of Dimensionality

- After performing 'exact matching' on 2 covariates:
  - Remaining observations = 17



- Q: would we still have observations left when we match on 3, 4, 5 covariates? What to do?

Matching in High Dimensions

- Even CEM can break down with high-dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
  - Reducing dimentionality: mapping two vectors to a single number.

Matching in High Dimensions

- Even CEM can break down with high-dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
    - Reducing dimentionality: mapping two vectors to a single number.
- **Mahalanobis distance**:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)'\widehat{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{x}_j)}$$

    - $\widehat{\Sigma}$ is the estimated variance-covariance matrix of the observations:

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

Matching in High Dimensions

- Even CEM can break down with high-dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
  - Reducing dimentionality: mapping two vectors to a single number.

- **Mahalanobis distance**:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \widehat{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{x}_j)}$$

  - $\widehat{\Sigma}$ is the estimated variance-covariance matrix of the observations:

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$$

- **Estimated propensity score** (Rosenbaum and Rubin, 1983):

$$D(\mathbf{x}_i, \mathbf{x}_j) = |\widehat{\pi}(\mathbf{x}_i) - \widehat{\pi}(\mathbf{x}_j)\ | \ = |\widehat{\mathbb{P}}(D_i = 1 \mid \mathbf{x}_i) - \widehat{\mathbb{P}}(D_j = 1 \mid \mathbf{x}_j)|$$

  - Some use linear predictor: $\text{Dist}_{ij} = \left|\text{logit}(\widehat{\pi}(\mathbf{x}_i)) - \text{logit}(\widehat{\pi}(\mathbf{x}_j))\right|$

Other Matching Choices

- **Matching ratio**: how many control units per treated unit?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance

Other Matching Choices

- **Matching ratio**: how many control units per treated unit?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance
- **With or without replacement**: same control matched to multiple treated?
  - With replacement gives better matches & matching order doesn't matter.
  - Without replacement simplifies variance estimation.

Other Matching Choices

- **Matching ratio**: how many control units per treated unit?
    - Lower reduces bias (only use the closest matches)
    - Lower increases variance
- **With or without replacement**: same control matched to multiple treated?
    - With replacement gives better matches & matching order doesn't matter.
    - Without replacement simplifies variance estimation.
- **Caliper**: drop poor matches?
    - Only keep matches below a distance threshold, $D(\mathbf{X}_i, \mathbf{X}_j) \leq c$
        - Rosenbaum and Rubin (1985): use c size equiv. to $0.25 \times$ sd of the PS.
    - Reduces imbalance, but if you drop treated units, estimand changes.
        - $\rightsquigarrow$ If we drop treated units, what are we estimating other than the ATT?

More on Propensity Scores

- Covariates are balanced conditional on true propensity scores:

$$D_i \perp\!\!\!\perp \mathbf{X}_i \mid \pi(\mathbf{X}_i)$$

More on Propensity Scores

- Covariates are balanced conditional on true propensity scores:

$$D_i \perp\!\!\!\perp \mathbf{X}_i \mid \pi(\mathbf{X}_i)$$

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\underbrace{(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i}_{\texttt{conditional unconfoundedness}} \iff (Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

More on Propensity Scores

- Covariates are balanced conditional on true propensity scores:

$$D_i \perp\!\!\!\perp \mathbf{X}_i \mid \pi(\mathbf{X}_i)$$

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\underbrace{(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i}_{\texttt{conditional unconfoundedness}} \iff (Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data, we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate $\widehat{\pi}(\mathbf{x})$

More on Propensity Scores

- Covariates are balanced conditional on true propensity scores:

$$D_i \perp\!\!\!\perp \mathbf{X}_i \mid \pi(\mathbf{X}_i)$$

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\underbrace{(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i}_{\texttt{conditional unconfoundedness}} \iff (Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data, we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate $\widehat{\pi}(\mathbf{x})$
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient? **No idea!**
  - Have to check if $\mathbf{X}_i$ is actually balanced.
  - Some what deflates the benefits of PS balancing/matching.
- $\rightsquigarrow$ "propensity score tautology"

Assessing Balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
  - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
  - Difficult to assess balance across many dimensions $\rightsquigarrow$ summaries.

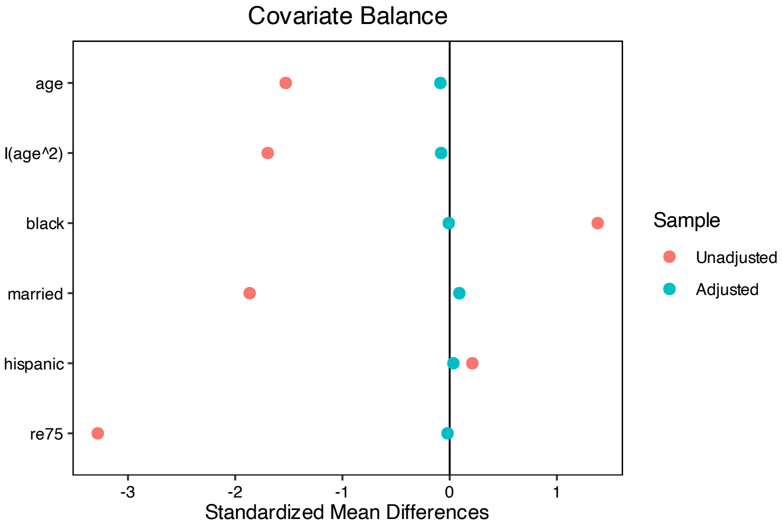
Assessing Balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to assess balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
    - Differences-in-means/medians, standardized.
    - Visualize balance using Love plot.
    - QQ plots/KS statistics for comparing the entire distribution of X_i.

Assessing Balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
  - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
  - Difficult to assess balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
  - Differences-in-means/medians, standardized.
  - Visualize balance using Love plot.
  - QQ plots/KS statistics for comparing the entire distribution of X_i.
- Hypothesis tests for balance are problematic:
  - Dropping units can lower power ($\uparrow$ p-values) without a change in balance.

Assessing Balance

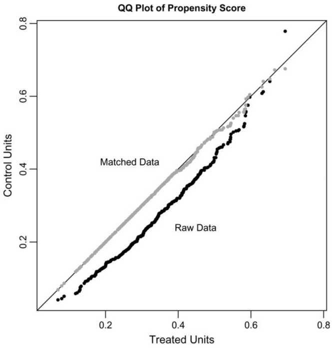


Assessing Balance

	Nonadopters		Adopters		Differences in Means
	Mean	Std dev	Mean	Std dev	
Customer spending:					
Purchase amount (monetary)	21.133	14.251	21.207	15.407	-0.074
Number of transactions (frequency)	1.318	0.578	1.321	0.536	-0.003
Number of transactions per trip	1.161	0.450	1.149	0.422	0.012
Days since last purchase (recency)	118.011	83.374	117.085	85.634	0.926
Number of books purchased (quantity)	1.819	1.173	1.848	1.256	-0.029
Maximum price of the books purchased	12.915	5.782	12.957	6.311	-0.042
Consumption variety:					
Number of unique books	1.733	1.101	1.766	1.149	-0.033
Number of unique genres	1.272	0.511	1.286	0.512	-0.014
Number of unique authors	1.975	1.448	2.019	1.473	-0.044
Number of unique publishers	1.685	0.988	1.696	0.970	-0.011
Concentration on personal favorites:					
Genre concentration (norm. HHI)	0.907	0.161	0.903	0.156	0.004
Author concentration (norm. HHI)	0.753	0.266	0.750	0.256	0.003
Publisher concentration (norm. HHI)	0.788	0.239	0.791	0.228	-0.003
Best-seller purchases:					
Shares of top 10 best sellers	0.076	0.219	0.074	0.201	0.002
Shares of top 50 best sellers	0.147	0.298	0.142	0.272	0.005
Usage of Sales Channels:					
Share of in-store mobile sales	0.147	0.328	0.190	0.362	-0.043
Share of m-commerce sales	0.071	0.237	0.096	0.268	-0.025
Share of e-commerce sales	0.086	0.261	0.094	0.267	-0.008
Share of cash register sales	0.103	0.279	0.107	0.278	-0.004
Customer demographics:					
Age	31.193	7.310	31.151	7.394	0.042
Gender (male = 1)	0.333	0.471	0.341	0.474	-0.008

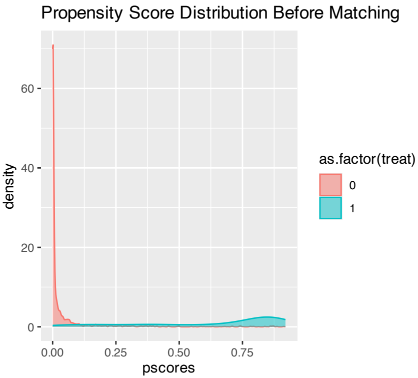
Note. Means and standard deviations of the observed variables, calculated in the 11-month pre-introduction period before February 1, 2015. The unit of analysis is an individual customer.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

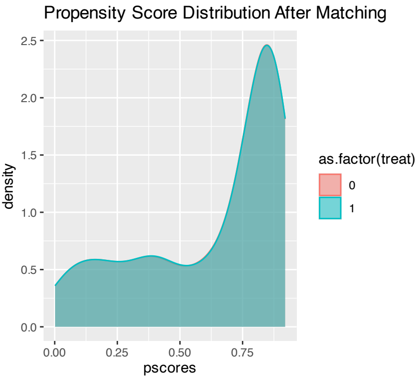
Source: *Table 3* of Yoo et al. "Mobile Payment and In-Store Mobile Purchase Behavior" KAIST Working Paper Series



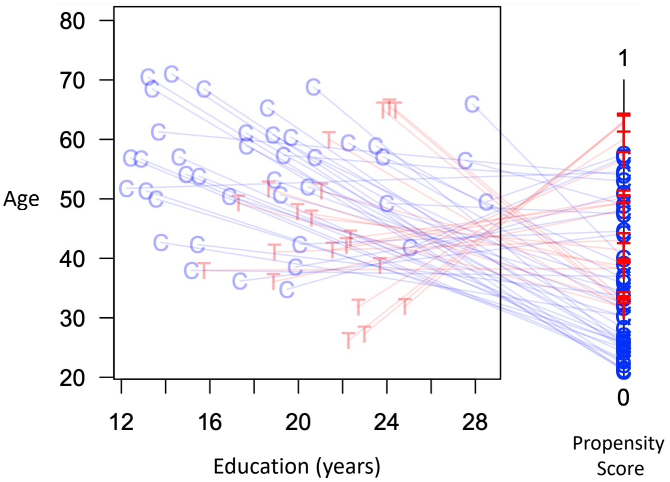
Source: *Figure 3* in Ho, Daniel E., et al. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." Political analysis 15.3 (2007): 199-236.

Assessing Balance



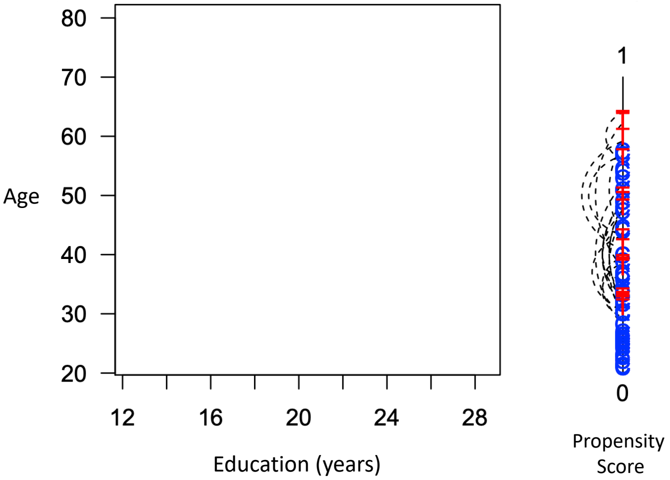


PSM Illustrated



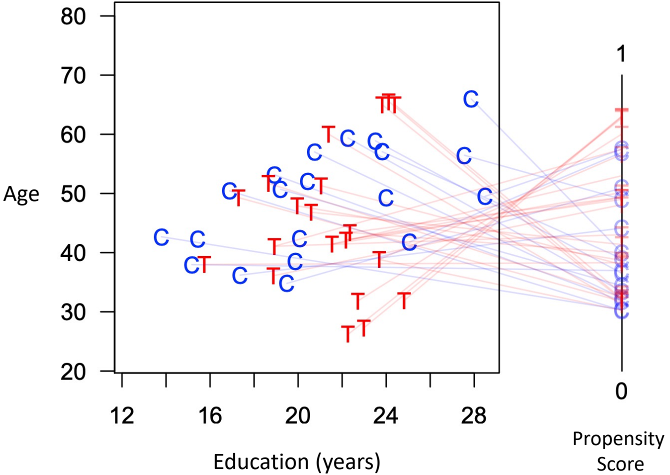
Credit: Figure from Gary King’s talk on “The Balance-Sample Size Frontier in Matching Methods for Causal Inference,” at University of Michigan, January 24, 2014.

PSM Illustrated



Credit: Figure from Gary King's talk on "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," at University of Michigan, January 24, 2014.

PSM Illustrated



Credit: Figure from Gary King's talk on "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," at University of Michigan, January 24, 2014.

Example: LaLonde Dataset

- The effectiveness of a job training program (National Supported Work Demonstration; NSW) on wage increases.
  - The federal government instituted a randomized evaluation of this program.
  - How well the result may be recovered when the experimental controls are replaced with a set of observational controls (Population Survey of Income Dynamics; PSID)?
  - Data publically available at NBER data archive.

- **Problem**: Imbalances between the experimental and observational data $\rightsquigarrow$ use matching.

Example: LaLonde dataset

- **Data:**
  - Treated: 297 units from NSW
  - Control: 2490 units from PSID
  - Treatment: Participation in the job training program (`treat`)
  - Outcome: 1978 earnings (in dollars; `re78`)
  - Pre-treatment covariates: age, race, marriage, past earnings, past employment

Example R Codes

- Import ant process data:

```
> pacman::p_load(tidyverse, broom, cobalt, Matching, MatchIt)

> lalonde_nsw <- haven::read_dta(url("http://www.nber.org/~rdehejia/data/nsw.dta"))
> PSID_obs <- haven::read_dta(url("http://www.nber.org/~rdehejia/data/psid_controls.dta"))

> lalonde_ECI <- full_join(lalonde_nsw |>
    filter(treat == 1),
    PSID_obs); lalonde_ECI

# A tibble: 2,787 × 11
   data_id    treat   age education black hispanic married nodegree  re75   re78
   <chr>      <dbl> <dbl>     <dbl> <dbl>    <dbl>   <dbl>    <dbl> <dbl>  <dbl>
 1 Lalonde S. . .     1    37        11     1        0       1        1     0  9930.
 2 Lalonde S. . .     1    22         9     0        1       0        1     0  3596.
 3 Lalonde S. . .     1    30        12     1        0       0        0     0 24909.
 4 Lalonde S. . .     1    27        11     1        0       0        1     0  7506.
 5 Lalonde S. . .     1    33         8     1        0       0        1     0   290.
 6 Lalonde S. . .     1    22         9     1        0       0        1     0  4056.
 7 Lalonde S. . .     1    23        12     1        0       0        0     0     0
 8 Lalonde S. . .     1    32        11     1        0       0        1     0  8472.
 9 Lalonde S. . .     1    22        16     1        0       0        0     0  2164.
10 Lalonde S. . .     1    33        12     0        0       1        0     0 12418.
#  2,777 more rows
#  1 more variable: re74 <dbl>
#  Use `print(n = ...)` to see more rows
```

Example R Codes

- Assessing balance before matching:

```
> bal.tab(x = lalonde_ECI |> dplyr::select(age:re78),
          treat = lalonde_ECI$treat, continuous = "std", binary = "std")

Note: `s.d.denom` not specified; assuming "pooled".
Balance Measures
             Type Diff.Un
age       Contin. -1.1662
education Contin. -0.6862
black      Binary  1.3222
hispanic   Binary  0.2554
married    Binary -1.9513
nodegree   Binary  0.9409
re75      Contin. -1.5662
re78      Contin. -1.2939

Sample sizes
    Control Treated
All    2490     297
```

Example R Codes

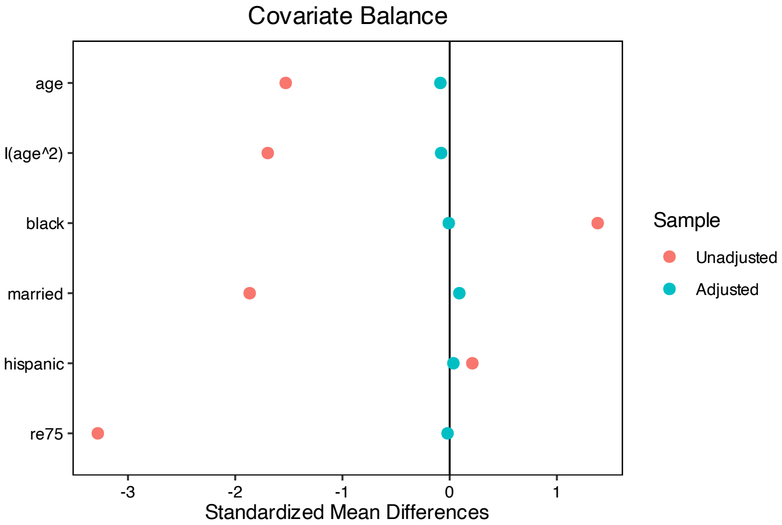
- Estimate propensity score using logistic regression:

```
# Propensity score
> pscores <- glm(treat ~ age + I(age^2) + black + married + hispanic + re75,
                 family = binomial(), data = lalonde_ECI)$fitted.values

# Conduct one-to-one nearest neighbor propensity score matching
> require(Matching)
> match_ps <- Match(Y=lalonde_ECI$re78, Tr=lalonde_ECI$treat,
                    X=pscores, M=1, replace = TRUE, ties = FALSE)
```

- Assess covariate balance after matching using love plot:

```
> cobalt::love.plot(treat ~ age + I(age^2) + black + married + hispanic + re75,
          data = lalonde_ECI,
          stats = "mean.diffs",
          weights = data.frame(Matched = get.w(match_ps)),
          method = c("matching"), binary = "std")
```

Balance test using Love Plot

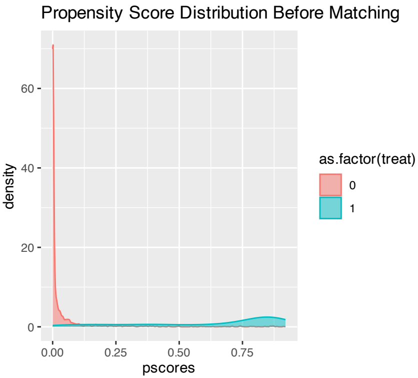


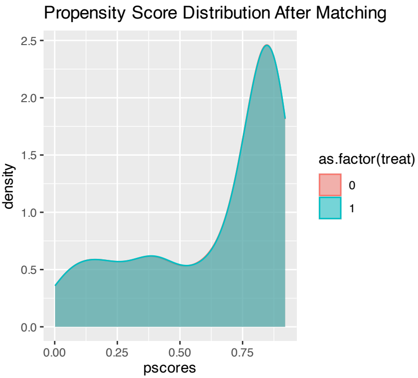
Example R Codes

- Estimate propensity score using logistic regression:

```r
# Attach estimated propensity scores
> lalonde_ECI$pscores <- pscores

# Extract matched indices
> matched_indices <- unlist(match_ps[c("index.treated", "index.control")])
> matched_data <- lalonde_ECI[matched_indices, ]

# Create a density plot of the propensity scores
> ps_dist_before <- lalonde_ECI |>
    ggplot(aes(x = pscores, group = treat,
               color = as.factor(treat),
               fill = as.factor(treat))) +
    geom_density(alpha = .5) +
    labs(title = "Propensity Score Distribution Before Matching")

> ps_dist_after <- matched_data |>
    ggplot(aes(x = pscores, group = treat,
               color = as.factor(treat),
               fill = as.factor(treat))) +
    geom_density(alpha = .5) +
    labs(title = "Propensity Score Distribution After Matching")

> ggpubr::ggarrange(ps_dist_before, ps_dist_after)
```

Propensity Score Distributions





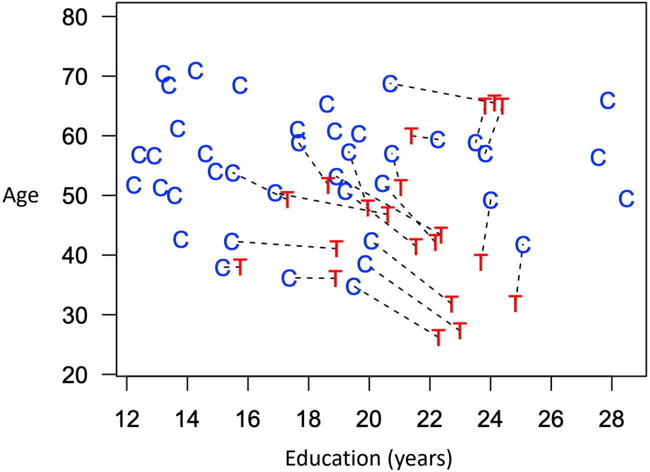
On to the Presentations & Discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

Matching in High Dimensions

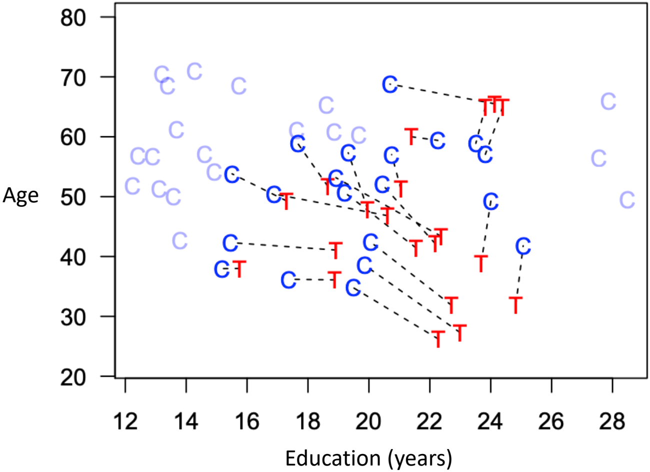
- **Mahalanobis Distance Matching**:
  - Prune observations where $\text{Dist}_{ij} > \text{caliper}$ ($\rightsquigarrow$ caliper? = cutoff point)



Credit: Figure from Gary King's talk on "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," at University of Michigan, January 24, 2014.

Matching in High Dimensions

- **Mahalanobis Distance Matching**:
  - Prune observations where $\text{Dist}_{ij} > \text{caliper}$ ($\rightsquigarrow$ caliper? = cutoff point)



Credit: Figure from Gary King's talk on "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," at University of Michigan, January 24, 2014.

Inverse Probability Weighting (IPW) Adjustments

- Matching (via pruning) has a downside:
  - It throws away data!
  - More problematic when we discard treated units $\rightsquigarrow$ changes estimand!

- Propensity scores can also be used as inverse weights directly when estimating the causal estimand.

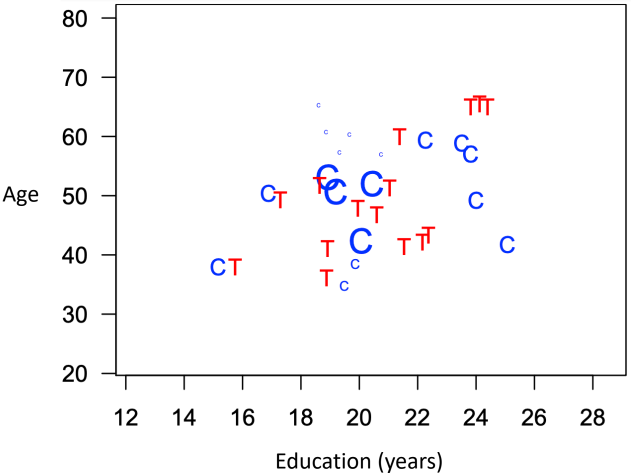
- Horvitz-Thompson IPW estimator for treatment effects:

$$\widehat{\text{ATE}} = \widehat{\tau}_{\text{ipw}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{D_i Y_i}{\widehat{\pi}(\mathbf{X}_i)} - \frac{(1 - D_i) Y_i}{1 - \widehat{\pi}(\mathbf{X}_i)} \right)$$

  - Under no unmeasured confounders, $\mathbb{E}[\widehat{\tau}_{\text{ipw}}] \xrightarrow{p} \tau_{\text{ATE}}$
  - Intuition: up-weight units that have smaller ‘treatment’ probability (underrepresented) $\rightsquigarrow$ sample more representative of population.

IPW Adjustments

- Applying weights to units instead of pruning.



Credit: Figure from Gary King's talk on "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," at University of Michigan, January 24, 2014.

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$
  - Bias grows with **matching discrepancy**/imbalance.

- **Bias correction**: estimate $\widehat{\mu}_0(\mathbf{x})$ with regression and estimate bias.

$$\widehat{Y}_i(0) = Y_{j(i)} - (\widehat{\mu}_0(\mathbf{X}_i) - \widehat{\mu}_0(\mathbf{X}_{j(i)}))$$

  - Imputation of missing potential outcome now matching + regression.
  - Generalizes easily to any number of matches.

Sampling Variance

- Matching with replacement: cluster on the match.
  - Can either use clustered SEs or cluster bootstrap.
  - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.
  - Same control unit matched to multiple treated: no easy clustering.
  - $K_M(i)$ is the number of times a unit is used as a match
- Assuming units are well-matched so bias can be ignored,

$$\mathbb{V}(\widehat{\tau}_m) = \frac{1}{n_1}\left(\underbrace{\mathbb{E}\left[(\tau(\mathbf{X}_i) - \tau_{\text{ATT}})^2 \mid D_i = 1\right]}_{\text{variance of CATE on treated}} + \underbrace{\mathbb{V}[\widehat{\tau}_m \mid \mathbb{X}, \mathbf{D}]}_{\text{conditional variance}}\right)$$

  - Abadie and Imbens (2006) provides matching-based variance estimators.