

# Model-based matching for causal inference in observational studies

Kellie Ottoboni  
with Philip B. Stark

Department of Statistics, UC Berkeley  
Berkeley Institute for Data Science

March 15, 2016

# Outline

## 1 Introduction

## 2 Matching

- Propensity score matching
- Model-based Matching

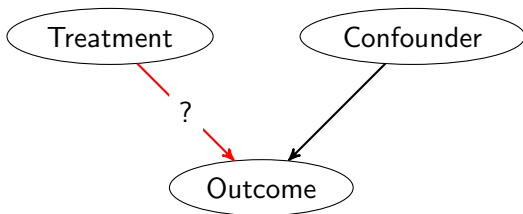
## 3 Examples

- Toads and Packstock in Yosemite
- Salt and Mortality

## 4 Conclusions

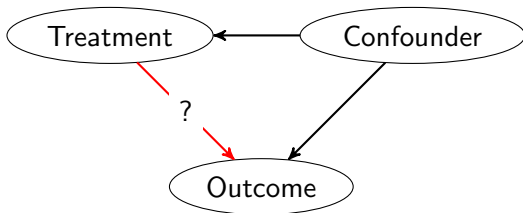
# Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



# Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



# Toads

placeholder for toads intro

# Neyman-Rubin Causal Model

- Population of  $i = 1, \dots, N$  individuals. Each individual has two **potential outcomes**.
- $Y_i(1)$  is individual  $i$ 's outcome if he receives treatment
- $Y_i(0)$  is individual  $i$ 's outcome if he is in the control group
- The treatment effect for individual  $i$  is  $\tau_i = Y_i(1) - Y_i(0)$

$Y_1(1)$	$Y_1(0)$
----------	----------

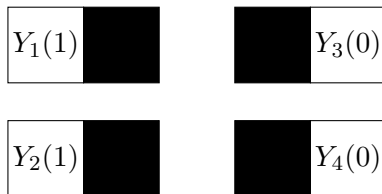
$Y_3(1)$	$Y_3(0)$
----------	----------

$Y_2(1)$	$Y_2(0)$
----------	----------

$Y_4(1)$	$Y_4(0)$
----------	----------

# Fundamental Problem of Causal Inference [Holland, 1986]

- We may never observe both  $Y_i(1)$  and  $Y_i(0)$
- $T_i$  is a treatment indicator: 1 if  $i$  is treated, 0 if  $i$  is control
- The observed outcome for individual  $i$  is
$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$



TO DO: FIX BOX POSITION

# Estimands

- Average treatment effect

$$\mathbb{E}(Y_i(1) - Y_i(0))$$

- Average treatment effect on the treated

$$\mathbb{E}(Y_i(1) - Y_i(0) \mid T_i = 1)$$

- Conditional average treatment effect

$$\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i)$$

- If treatment effect varies by covariates  $X$ , then averages might not be informative



# Goal

**Goal:** test the **strong null hypothesis** of no treatment effect whatsoever.

$$H_0 : Y_i(1) = Y_i(0) \text{ for all } i$$

$$H_1 : Y_i(1) \neq Y_i(0) \text{ for some } i$$

We'd like our test to have power to detect

- non-constant effects
- non-linear effects
- effects with non-constant sign

# Outline

## 1 Introduction

## 2 Matching

- Propensity score matching
- Model-based Matching

## 3 Examples

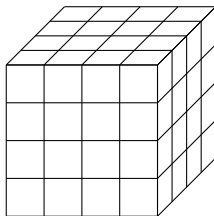
- Toads and Packstock in Yosemite
- Salt and Mortality

## 4 Conclusions

# Matching

How can we estimate the counterfactual for treated individuals?

- **Ideal:** group individuals by  $X_i$  to estimate subgroup treatment effects and then average over subgroups
- **Reality:** many covariates, perhaps continuous, make it difficult to stratify

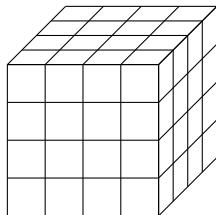


## Aside... the curse of dimensionality

- If  $d$  covariates are split into  $k$  bins, we have  $d^k$  groups.
- To guarantee that we have at least one treated and one control in each group with 95% probability, we need

$$n \geq \frac{2 \log(1 - (0.95)^{1/k^{d+1}})}{\log(\frac{k^d - 1}{k^d})}$$

- If  $d = 5$  and  $k = 2$ ,  $n \geq 225$ .
- If  $d = 10$  and  $k = 2$ ,  $n \geq 10,844$ .



# Matching

- **Solution:** use a one-dimensional score to match or group individuals

# Propensity score matching

- The **propensity score** is an individual's probability of being assigned treatment, conditional on their covariates

$$p(x) = \mathbb{P}(T = 1 \mid X = x)$$

- The propensity score is a balancing score:  $X \perp\!\!\!\perp T \mid p(X)$
- For individuals with the same propensity score, treatment assignment is as if random

# Propensity score matching

## Theorem (Rosenbaum and Rubin [1983])

*If treatment assignment is independent of potential outcomes given  $X$ ,*

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

*and if every unit has a chance of receiving treatment,*

$$0 < p(X) < 1 \text{ for all } X$$

*then  $(Y(1), Y(0)) \perp\!\!\!\perp T \mid p(X)$ .*

In particular, treated units can serve as the counterfactual for controls with the same  $p(X)$

$$\mathbb{E}(Y(t) \mid T = 1, p(X)) = \mathbb{E}(Y(t) \mid T = 0, p(X)) \text{ for } t = 0, 1$$

# Propensity score matching

This result identifies the average treatment effect in terms of quantities we can estimate:

$$\begin{aligned}\mathbb{E}(Y(1) - Y(0)) &= \mathbb{E}_{p(x)} [\mathbb{E}(Y(1) - Y(0) \mid p(x))] \\ &= \mathbb{E}_{p(x)} [\mathbb{E}(Y(1) \mid p(x)) - \mathbb{E}(Y(0) \mid p(x))] \\ &= \mathbb{E}_{p(x)} [\mathbb{E}(Y \mid p(x), T = 1) - \mathbb{E}(Y \mid p(x), T = 0)]\end{aligned}$$



# Propensity score matching

$p(x)$  is usually unknown and estimated by  $\hat{p}(x)$  using logistic or probit regressions

- Assumes a simple functional form for relationship between covariates and treatment
- Assumes that probability of treatment takes same form for all individuals
- May actually worsen balance if estimated incorrectly [Diamond and Sekhon, 2012]

Matching complicates inference

- Standard errors are difficult to compute for matching estimators [Abadie and Imbens, 2006, 2008]
- Rarely used in hypothesis testing procedures
- There's no “optimal” way to match [Austin, 2014]

# Model-based Matching

**Idea:** Instead of modeling the propensity score, model the outcome

Computing  $\hat{Y}$ , the “best” prediction of the outcome based on all covariates except for the treatment, buys us two things:

- $\hat{Y}$  is a score on which to stratify observations
- Using residuals  $Y - \hat{Y}$  improves precision by removing variation due to  $X$  [Rosenbaum, 2002]

# Model-based Matching

Suppose that outcomes have the form

$$Y_i(t) = f(t, X_i) + \varepsilon_i$$

for  $i = 1, \dots, N$  and  $t = 0, 1$ . Let  $X_i$  be fixed and suppose that the  $\varepsilon_i$  are IID with  $\mathbb{E}(\varepsilon_i) = 0$ .

Under the strong null hypothesis,  $f(0, X_i) = f(1, X_i)$  for each  $i$ .

Thus, our best guess of  $Y_i$  needn't involve the treatment:

$$\hat{Y}_i = \hat{f}(X_i)$$

# Permutation tests

**Basic idea:** If, under the null hypothesis, the probability distribution of the data is invariant under permutation of treatment assignments, then once we observe the actual data, we know other possible data sets that are equally likely.

Suppose we have  $S$  strata and the  $s$ th stratum contains  $N_s$  units,  $n_s$  of which are treated. Then there are

$$\prod_{s=1}^S \binom{N_s}{n_s}$$

equally likely assignments to treatment, conditional on the number of treated units in each stratum.

TO DO: FLESH OUT

# Permutation tests

On the other hand, if the null hypothesis is false, then the residuals will capture some of the effect of treatment:

$$Y_i - \hat{Y}_i \not\perp T_i$$

We will use the average difference in means across strata as our test statistic:

$$\tau(Y, T) = \frac{N_s}{N} \sum_{s=1}^S \left| \frac{n_s}{N_s} \sum_{i \in s: T_i=1} (Y_i - \hat{Y}_i) - \frac{N_s - n_s}{N_s} \sum_{i \in s: T_i=0} (Y_i - \hat{Y}_i) \right|$$

TO DO: KEEP NOTATION STRAIGHT. S CAN'T BE INDEX AND SET

# Permutation tests

We approximate the null distribution using this invariance principle.

- Within strata, permute treatment assignments to obtain new treatment vector  $T_1^*$ .
- Compute the test statistic  $\tau(Y, T_1^*)$ .
- Repeat a large number  $B$  times to get a distribution  $\tau(Y, T_1^*), \dots, \tau(Y, T_B^*)$ .
- The p-value of the test is

$$p = \mathbb{P}(\tau(Y, T) \geq \tau(Y, t)) \approx \frac{\sum_{i=1}^B \mathbb{I}(\tau(Y, T_b^*) \geq \tau(Y, T))}{B}$$

# Outline

## 1 Introduction

## 2 Matching

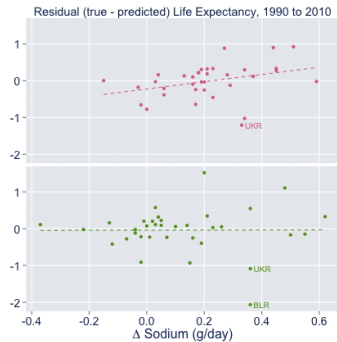
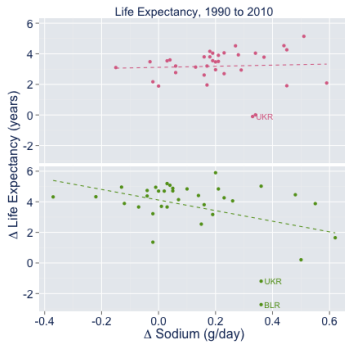
- Propensity score matching
- Model-based Matching

## 3 Examples

- Toads and Packstock in Yosemite
- Salt and Mortality

## 4 Conclusions

# Salt



Female Male



# Outline

## 1 Introduction

## 2 Matching

- Propensity score matching
- Model-based Matching

## 3 Examples

- Toads and Packstock in Yosemite
- Salt and Mortality

## 4 Conclusions

# Future Directions

- Do different test statistics give greater power when the treatment effect is nonlinear?
- What is the optimal way to stratify?
- How to quantify uncertainty – standard errors and confidence intervals?

# References

- Alberto Abadie and Guido W. Imbens. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1):235–267, January 2006. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2006.00655.x. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2006.00655.x/abstract>.
- Alberto Abadie and Guido W. Imbens. On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, 76(6):1537–1557, November 2008. ISSN 1468-0262. doi: 10.3982/ECTA6474. URL <http://onlinelibrary.wiley.com/doi/10.3982/ECTA6474/abstract>.
- Peter C. Austin. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6): 1057–1069, March 2014. ISSN 1097-0258. doi: 10.1002/sim.6004. URL <http://onlinelibrary.wiley.com/doi/10.1002/sim.6004/abstract>.
- Alexis Diamond and Jasjeet S. Sekhon. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics*, 95(3): 932–945, October 2012. ISSN 0034-6535. doi: 10.1162/REST\_a\_00318. URL [http://dx.doi.org/10.1162/REST\\_a\\_00318](http://dx.doi.org/10.1162/REST_a_00318).
- Paul W. Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396): 945–960, 1986. ISSN 0162-1459. doi: 10.2307/2289064. URL <http://www.jstor.org/stable/2289064>.
- Paul R. Rosenbaum. Covariance Adjustment in Randomized Experiments and Observational Studies. *Statistical Science*, 17(3):286–327, August 2002. ISSN 0883-4237, 2168-8745. doi: 10.1214/ss/1042727942. URL <http://projecteuclid.org/euclid.ss/1042727942>.
- Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, April 1983. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/70.1.41. URL <http://biomet.oxfordjournals.org/content/70/1/41>.