Section 10: Matching

Konstantin Kashin[1]

April 10, 2013

[1]Thanks to Molly Roberts, Brandon Stewart, and former TFs for contributing to this material.

Outline

Outline

Setup

- Let's denote treatment as $T \in 0, 1$. $T = 1$ is treated group, $T = 0$ is control group.
- We have an outcome Y
- SUTVA: stable unit treatment value assumption
  - No interference between units
  - No hidden levels of treatment
- Under SUTVA, we have 2 potential outcomes per unit:

$$Y_i(1) \text{ and } Y_i(0)$$

- Before we think about how treatment was assigned, we can express this information in a potential outcomes table

Potential Outcomes Table

i	Name	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	$Y_1(1)$	$Y_1(0)$
2	Thomas Jefferson	$Y_2(1)$	$Y_2(0)$
3	George Washington	$Y_3(1)$	$Y_3(0)$
4	Alexander Hamilton	$Y_4(1)$	$Y_4(0)$
5	James Madison	$Y_5(1)$	$Y_5(0)$
6	John Jay	$Y_6(1)$	$Y_6(0)$

Individual causal effect for Benjamin Franklin: $Y_1(1)$-$Y_1(0)$

All potential outcomes are fixed pre-treatment characteristics of individuals

Defining the Estimand

We should define an estimand before we think about estimation!
What is a causal quantity we are interested in?

For example, an average treatment effect (ATE):

$$\mathrm{E}[Y_i(1) - Y_i(0)]$$

Fundamental Problem of Causal Inference

i	Name	$Y_i(1)$	$Y_i(0)$	T_i	Y_i^{obs}
1	Benjamin Franklin	$Y_1(1)$	$Y_1(0)$	1	$Y_1(1)$
2	Thomas Jefferson	$Y_2(1)$	$Y_2(0)$	0	$Y_2(0)$
3	George Washington	$Y_3(1)$	$Y_3(0)$	0	$Y_3(0)$
4	Alexander Hamilton	$Y_4(1)$	$Y_4(0)$	1	$Y_4(1)$
5	James Madison	$Y_5(1)$	$Y_5(0)$	1	$Y_5(1)$
6	John Jay	$Y_6(1)$	$Y_6(0)$	0	$Y_6(0)$

The **fundamental problem of causal inference** is that we only observe one potential outcome per unit!

Fundamental Problem of Causal Inference

How do we estimate the ATE from observed data?

$$\mathrm{E}[Y_i(1) - Y_i(0)]$$

Fundamental Problem of Causal Inference

Causal inference is a missing data problem!
We want to impute missing potential outcomes.

Unconfoundedness

Let's assume **unconfoundedness** of treatment:

$$P(T|Y(0), Y(1), X) = P(T)$$

Unconfoundedness translates to the following:

$$\mathrm{E}[Y(1)] = \mathrm{E}[Y(1)|T = 1] = \mathrm{E}[Y(1)|T = 0]$$
$$\mathrm{E}[Y(0)] = \mathrm{E}[Y(0)|T = 1] = \mathrm{E}[Y(0)|T = 0]$$

Now:

$$\begin{aligned} \mathrm{E}[Y(1) - Y(0)] &= \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \\ &= E[Y(1)|T = 1] - E[Y(0)|T = 0] \\ &= E[Y^{obs}|T = 1] - E[Y^{obs}|T = 0] \end{aligned}$$

Causal Effects with Unconfoundedness

	T=1 (Treatment)	T=0 (Control)
$E[Y\|T=t]$	6.6	2.4
% Data	.5	.5

If this is our data, how would we estimate the average causal effect of T on Y? Assuming unconfoundedness:

$$\begin{aligned} \mathrm{E}[Y(1)-Y(0)] &= E[Y(1)|T=1]-E[Y(0)|T=0] \\ &= E[Y^{obs}|T=1]-E[Y^{obs}|T=0] \\ &= 6.6-2.4 \\ &= 4.2 \end{aligned}$$

Unconfoundedness

Recall that unconfoundedness implies:

1. $\mathrm{E}[Y(1)] = \mathrm{E}[Y(1)|T = 1] = \mathrm{E}[Y(1)|T = 0]$
2. $\mathrm{E}[Y(0)] = \mathrm{E}[Y(0)|T = 1] = \mathrm{E}[Y(0)|T = 0]$

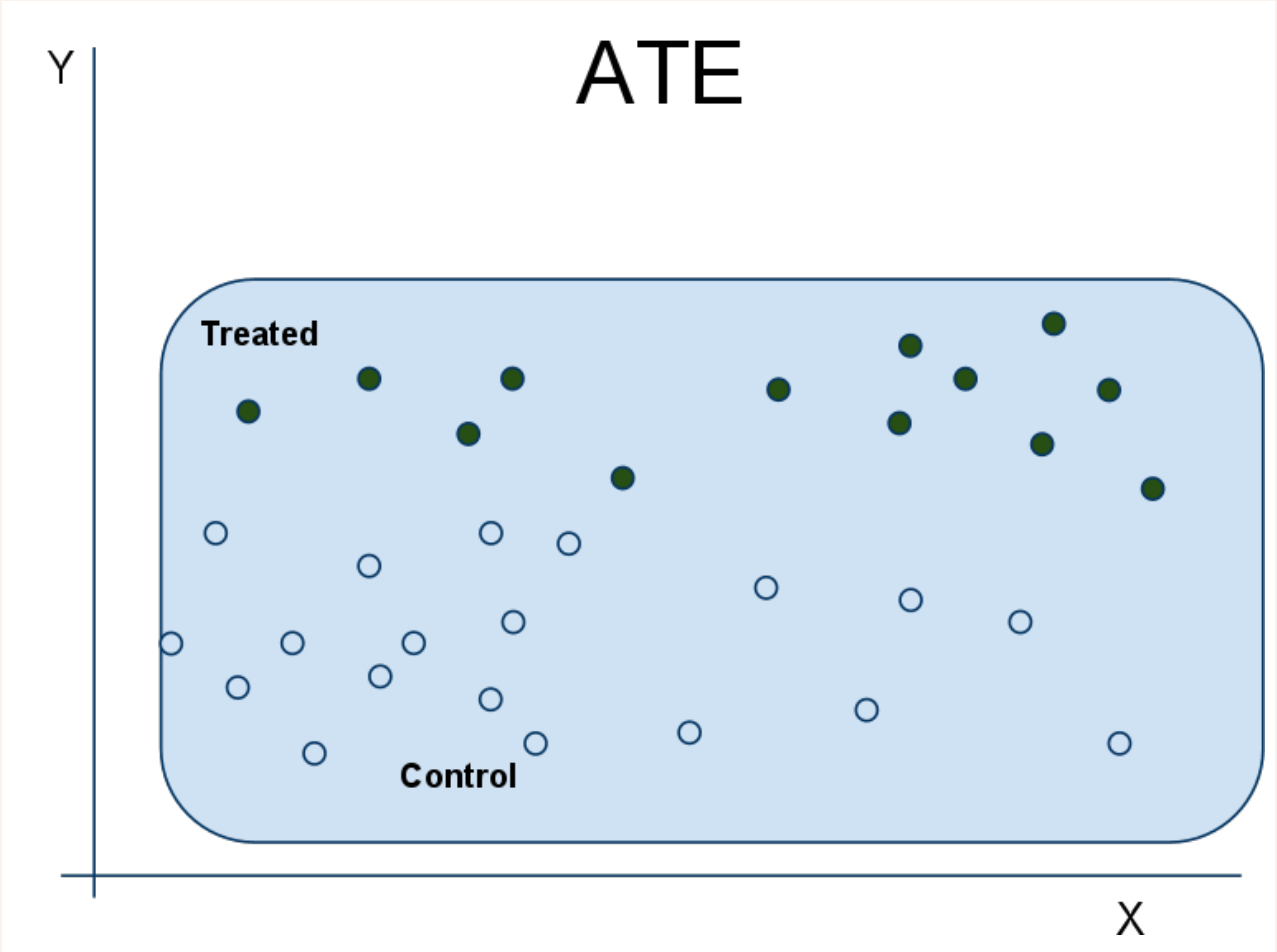
What if we could only assume 2?

We could still calculate the average treatment effect on the treated (ATT):

$$\begin{aligned} E[Y(1) - Y(0)|T = 1] &= E[Y(1)|T = 1] - E[Y(0)|T = 1] \\ &= E[Y(1)|T = 1] - E[Y(0)|T = 0] \\ &= E[Y^{obs}|T = 1] - E[Y^{obs}|T = 0] \end{aligned}$$

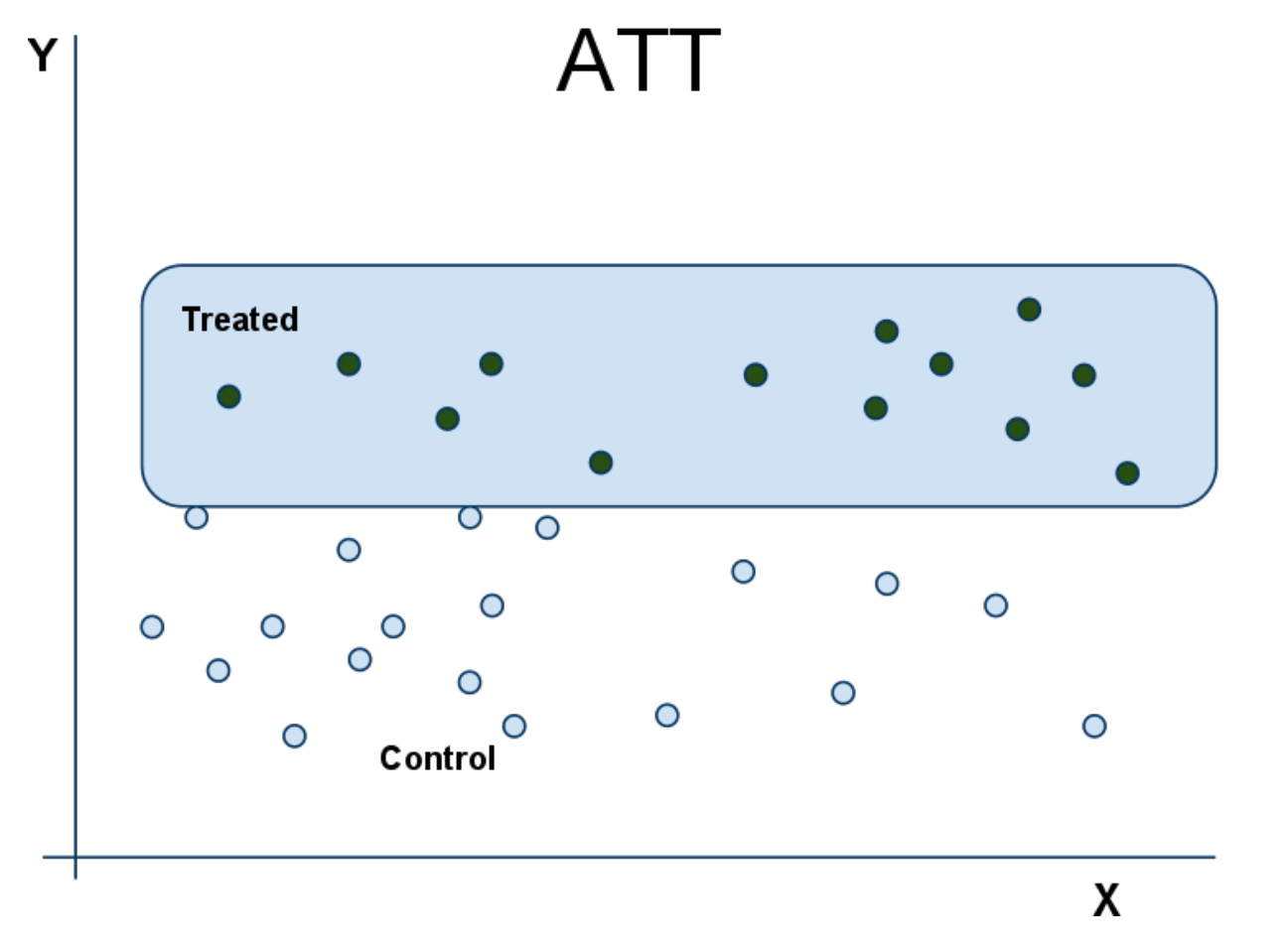
Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



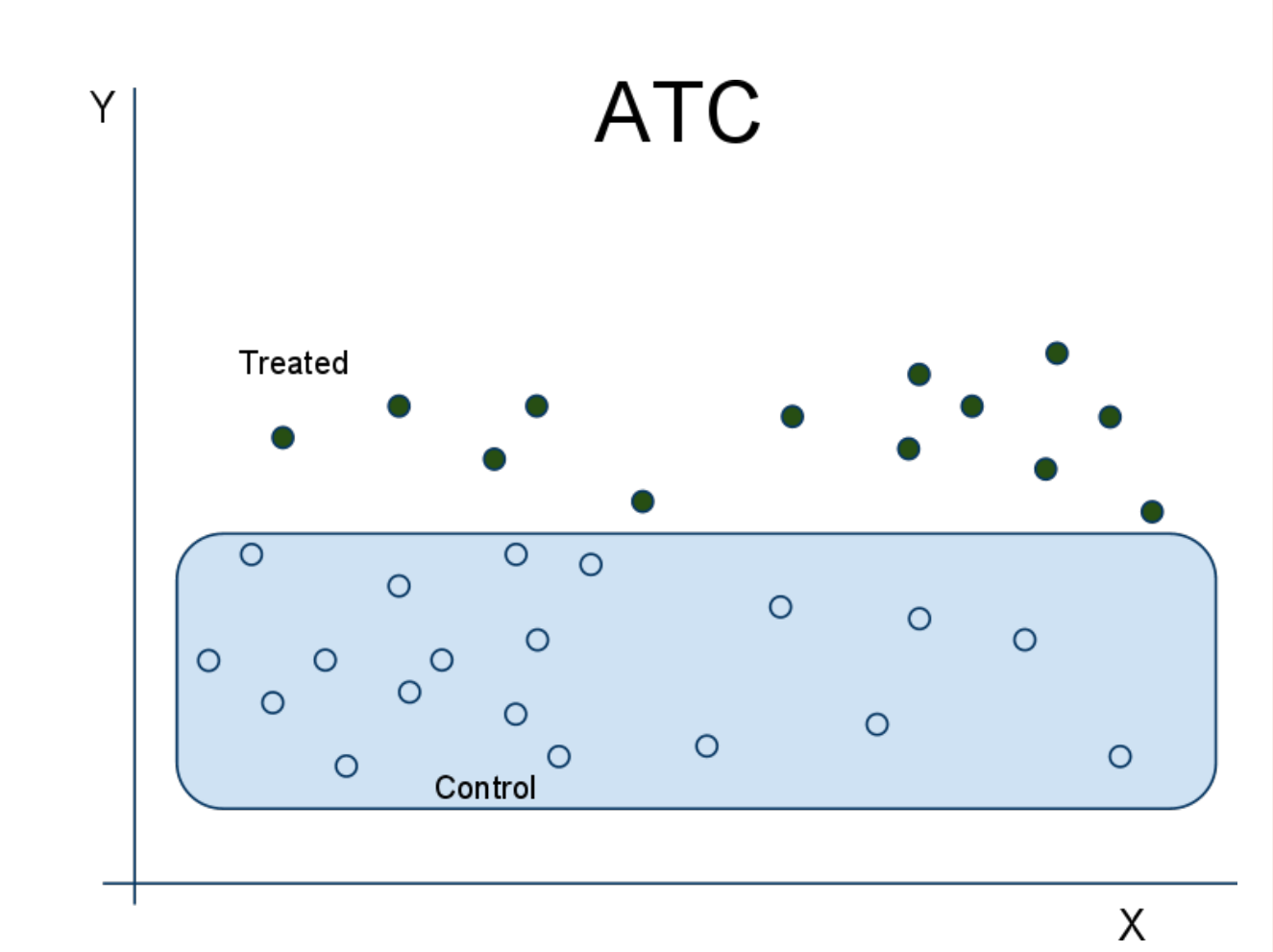
Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



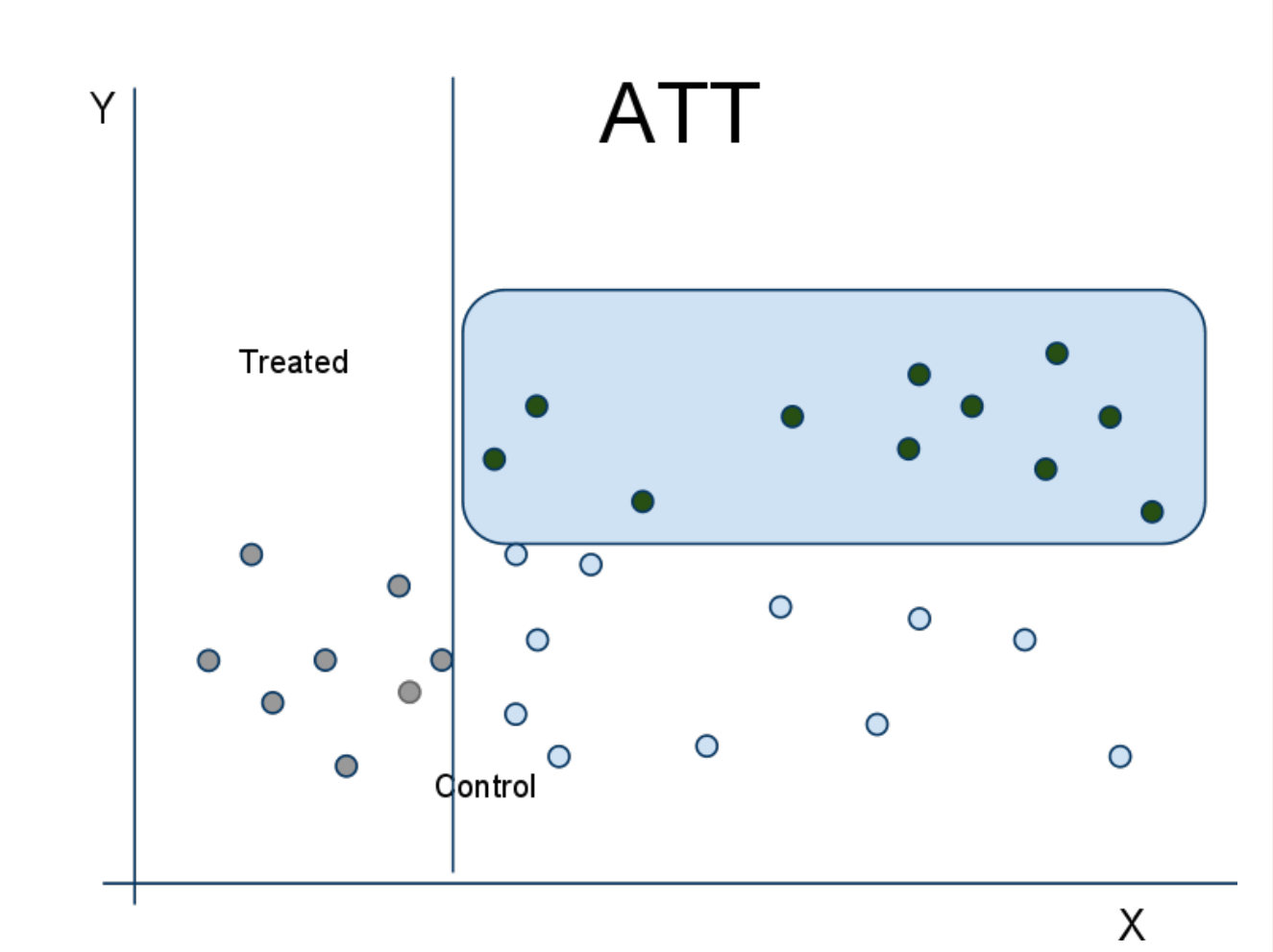
Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



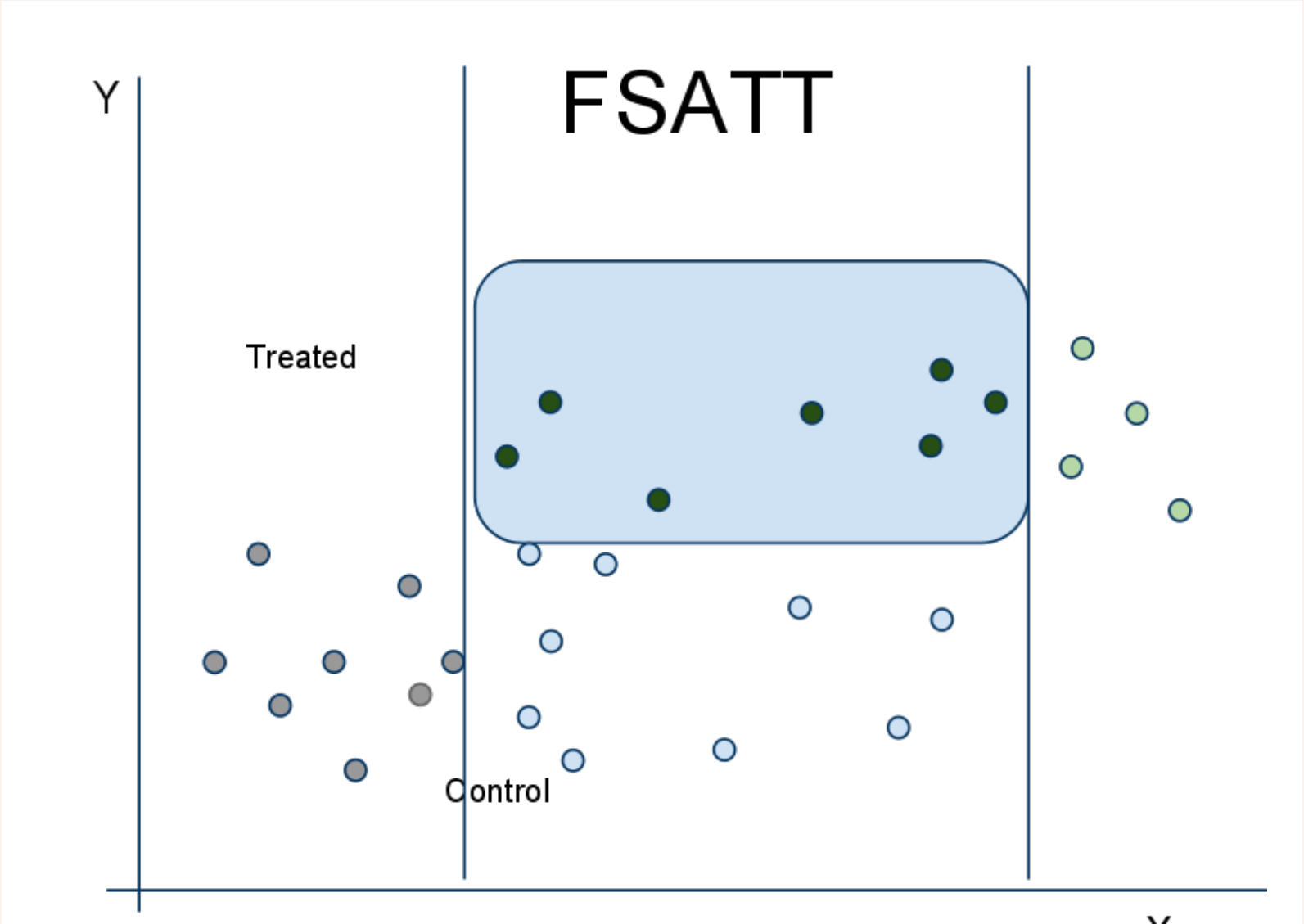
Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



Confounding (with measured covariates)

What if we do not have unconfoundedness?

$$P(T|Y(0), Y(1), X) = P(T|X)$$

- **Consequence**: imbalance between treated and control units in X
- **Possible solution**: matching

Motivations for Matching

1. Matching as stratification: find strata within the data for which treatment assignment is uncorrelated with potential outcomes
2. Matching as weighting:all matching algorithms boil down to a creative reweighting of observed information in an attempt to reconstruct appropriate counterfactuals
3. Matching as preprocessing to reduce model dependence:prune data as with a non-parametric matching analysis, and then fit parametric model

Same data as before, divided according to X

Now suppose that those who take the treatment are systematically different than those who don't take the treatment.

	T=1 (Treatment)	T=0 (Control)
$E[Y\|T = t, X = 0]$	-6	2
% Data	.15	.40
$E[Y\|T = t, X = 1]$	12	4
% Data	.35	.10

In this case, $X = 0$ strata responds negatively to treatment and $X = 1$ strata responds positively to treatment. Moreover, it appears that those who will be negatively affected by treatment are opting for control.

Pre-treatment or Post-Treatment?

There are two possibilities here:

1. If we felt that X was post-treatment and there are no other confounders, we would stick with the ATE.

$$\frac{.15 \cdot -6 + .35 \cdot 12}{.15 + .35} - \frac{.40 \cdot 2 + .10 \cdot 4}{.40 + .10} = 4.2.$$

2. If we felt X was pre-treatment we might assert that *within* strata of X, we have unconfoundedness: $P(T|Y(1), Y(0), X) = P(T|X)$. This implies:
   - Assumption 1a: $E[Y^1|T = 1, X] = E[Y^1|T = 0, X]$
   - Assumption 2a: $E[Y^0|T = 1, X] = E[Y^0|T = 0, X]$

Unconfoundedness within Strata

Unconfoundedness *within* strata of X is expressed as:

$$P(T|Y(1), Y(0), X) = P(T|X)$$

It is equivalent to:

$$Y(0), Y(1) \perp T|X$$

It is also called:

- Conditional ignorability
- Conditional exchangeability
- Selection on observables

Matching as Stratfication

If we assume unconfoundedness within strata, we can think of matching as **stratification**:

$$
\begin{aligned}
E[Y^1 - Y^0] &= E[Y^1] - E[Y^0] \\
&= E_X[E[Y^1|X] - E[Y^0|X]] \\
&= E_X[E[Y|T=1,X] - E[Y|T=0,X]] \\
&= Pr(X=0)(E[Y|T=1,X=0] - E[Y|T=0,X=0]) \\
&+ Pr(X=1)(E[Y|T=1,X=1] - E[Y|T=0,X=1])
\end{aligned}
$$

Matching as Stratification

	T=1 (Treatment)	T=0 (Control)
$E[Y\|T = t, X = 0]$	-6	2
% Data	.15	.40
$E[Y\|T = t, X = 1]$	12	4
% Data	.35	.10

We've stratified:

- match up the estimated potential outcomes within the strata defined by X=0 and X=1
- derive the causal effect in each stratum
- recombine weighting appropriately based on the presence of X=0 and X=1 in the sample.

$$\begin{aligned} E[Y(1) - Y(0)] &= .55(-6 - 2) + .45(12 - 4) \\ &= -0.8 \end{aligned}$$

Matching as Weighting

Estimator where X was not a confounder:

$$ATE_{simple} = \frac{.15 \cdot -6 + .35 \cdot 12}{.15 + .35} - \frac{.40 \cdot 2 + .10 \cdot 4}{.40 + .10} = 4.2$$

Estimator where X is a confounder:

$$ATE_{match} = .55(-6 - 2) + .45(12 - 4) = -0.8$$

The same average outcomes are used to construct these (four different $E[Y|T = t, X = x]$'s) but they are *weighted* differently.

ATT

What if we wanted the ATT: $E[Y(1) - Y(0)|T = 1]$

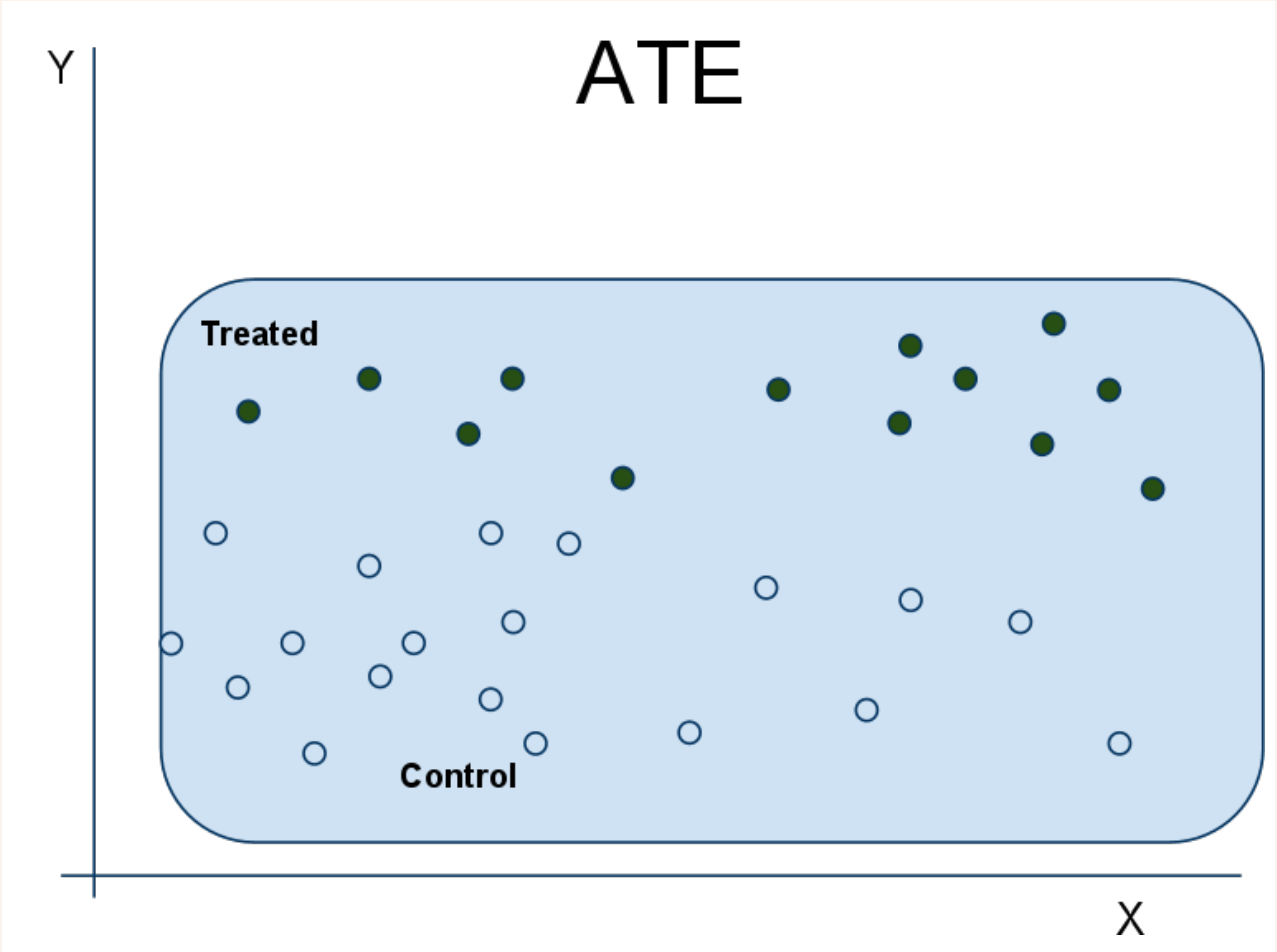
- Do the same things as before (match potential outcomes within strata, derive strata-specific causal effects)
- Weight the strata-specific effects based on the presence of X=1 and X=0 among observations where T=1.

	T=1 (Treatment)	T=0 (Control)
$E[Y\|T = t, X = 0]$	-6	2
% Data	.15	.40
$E[Y\|T = t, X = 1]$	12	4
% Data	.35	.10

$$\begin{aligned} E[Y^1 - Y^0] &= \frac{.15}{.5}(-6 - 2) + \frac{.35}{.5}(12 - 4) \\ &= 3.2 \end{aligned}$$

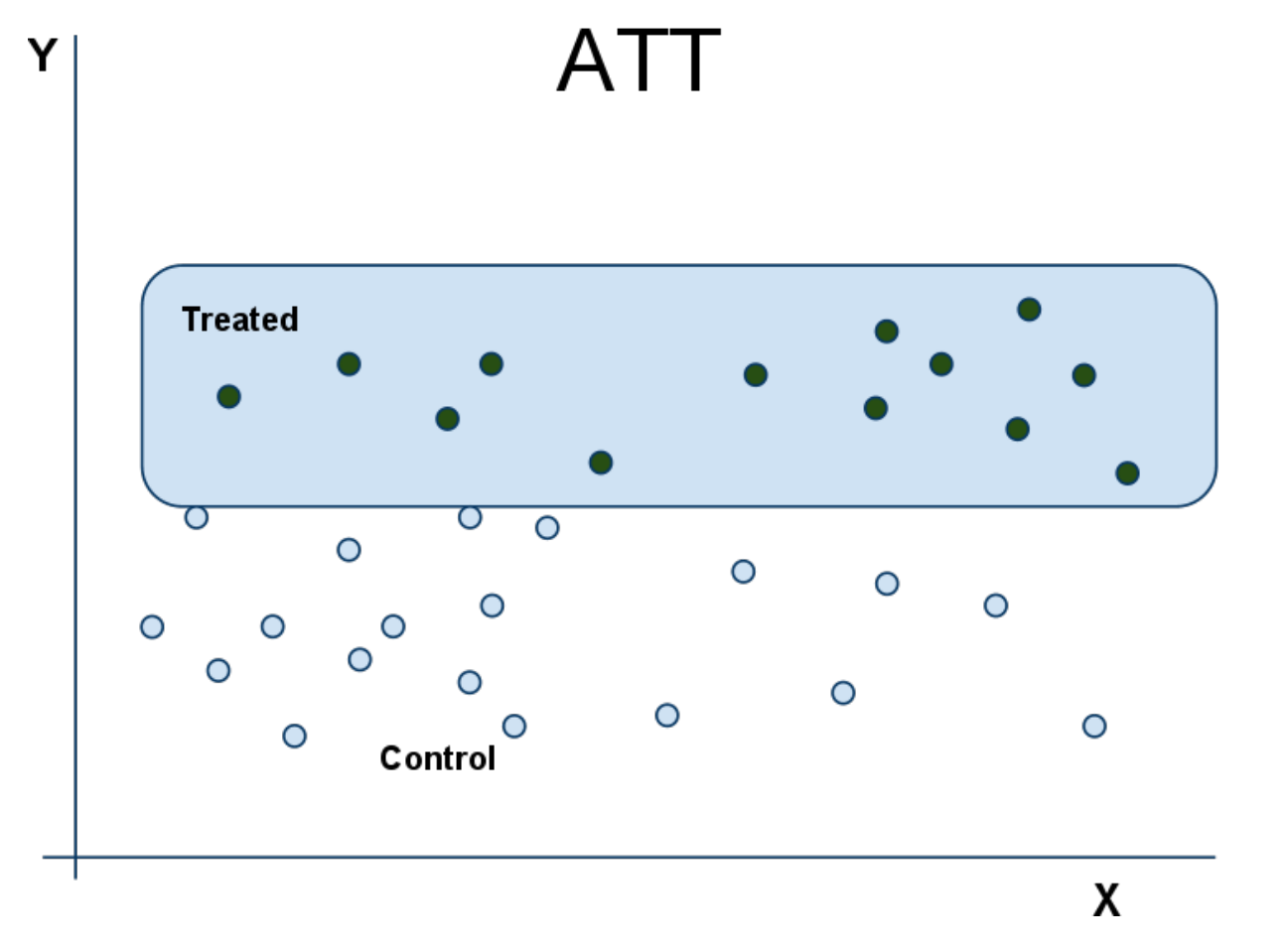
Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



Bracket: ATT vs. ATE

Think of this as the group that we need to find counterfactuals for.



General Strategy

We now have a general strategy for handling non-random treatment assignment:

- 1. Condition on variables such that treatment assignment is uncorrelated with potential outcomes conditional on those covariates.
- 2. Match data according to the strata defined by the values these variables take on.
- 3. Assess our matching procedure (balance checking).

General Strategy

- 4a. Recombine these strata-specific causal effects into an overall treatment effect by appropriately weighting (non-parametric).
- 4bi. Proceed with parametric analysis (double robustness).
- 5. Sensitivity testing for either the confoundedness assumption or the parametric model.

Outline

Things to think about while matching

1. Distance
2. Distance $\Rightarrow$ Matches
3. K to K?
4. With/Without replacement

Distance

Many different conceptions of distance, here are a few:

- 1. **Exact**:
  - Distance = 0 if $X_i = X_j$
  - Distance = ∞ if $X_i \neq X_j$
  - Ideal, but hard for a lot of variables
- 2. **Mahalanobis**:
  - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
  - Doesn't work very well when X is high dimensional because it tries to take into account all interactions between the covariates.
  - Can emphasize importance of a variable by doing Mahalnobis distance matching within calipers

Distance

- 3. **Propensity score**
  - Estimate $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
  - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
  - Overcome the high-dimensionality problem by summarizing covariates into one scalar
  - Has nice asymptotic properties
- 4. **Linear propensity score**:
  - Distance$(X_i, X_j) = |\text{logit}(\pi_i) - \text{logit}(\pi_j)|$

See Stuart, E. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward", *Statistical Science* 25(1): 1-21.

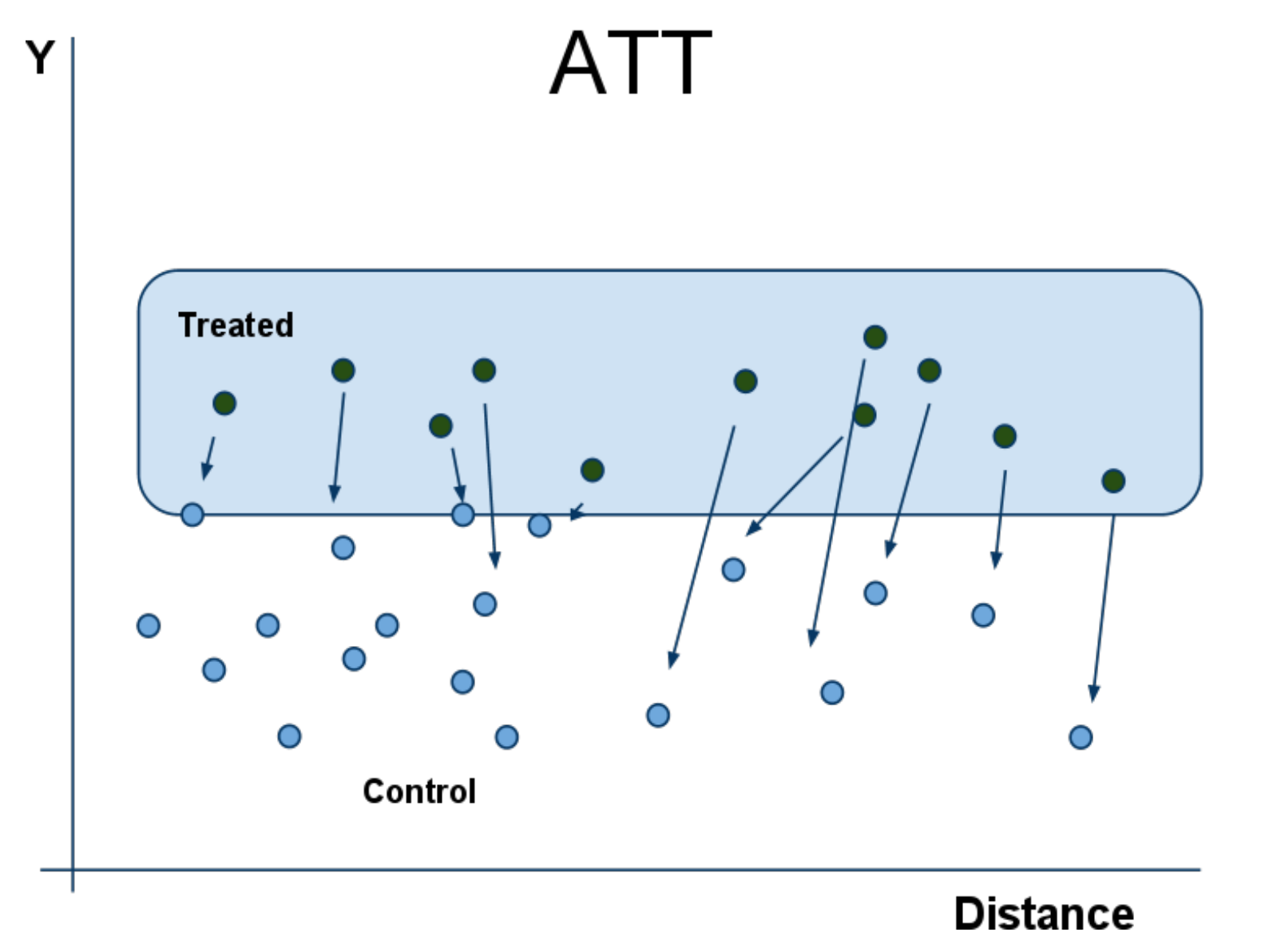
Distance ⇒ Matches

Once we have a distance metric, how do we determine matches?

- 1. K:1 Nearest neighbor matching
  - Almost always is an estimate of the ATT – matches the treated group, then discards the remaining controls.
  - "Greedy"
- 2. Optimal matching
  - Instead of being greedy, minimizes a global distance measure
  - Reduces the difference between pairs, but not necessarily the difference between groups

General points: These methods give weights of zero or 1 to observations, depending on whether they are matched.

Visualizing Nearest Neighbor and Optimal Matches

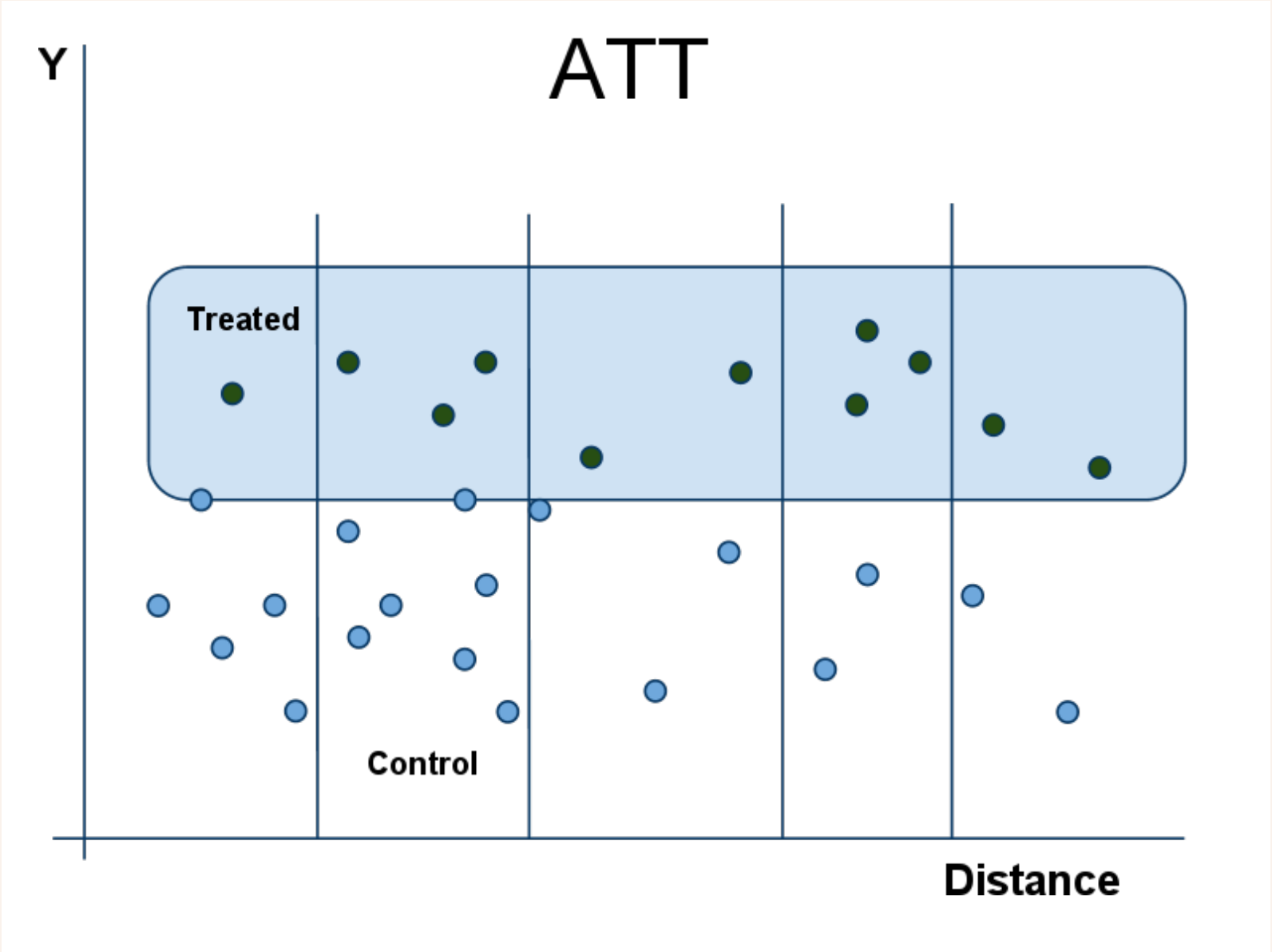


Distance ⇒ Matches

Once we have a distance metric, how do we determine matches?

- 3. Full matching
  - Creates subclasses of variables based on the distance metric
  - Estimates the ATT or ATE within each subgroup, weighting each group by the number of observations.
- 4. Weighting based on propensity scores
  - Weighting based on propensity scores is essentially a very fine subclassification scheme
  - To get the ATE, weight by $w_i = \frac{T_i}{\hat{e}_i} + \frac{(1-T_i)}{(1-\hat{e}_i)}$
  - To get the ATT, weight by $w_i = T_i + (1 - T_i)\frac{\hat{e}_i}{1-\hat{e}_i}$

Visualizing Subclassification



Where does CEM fit into this?

- CEM is exact matching with coarsening.
- Similar to sub-classification, but the classification is not based on a distance metric, it's based on substantive knowledge of the covariates.
- This allows us to get the benefits of exact matching without the problems of high dimensionality.
- CEM also weights to get the ATT and the ATE depending on how many observations are in the coarsened categories.

Pros and Cons of K to K matching

- **Cons:**
  - You are discarding observations
  - Might lead to reduced power
- **Pros**:
  - You get better matches
  - It might not lead to reduced power because power is often driven by the size of the smaller group (treated or control).
  - Power can be increased if you have better precision (reduced extrapolation)

General point: Pros and Cons of K to K matching are very similar to the bias-varaince tradeoff we talked about for matching in class.

Pros and Cons of Matching with Replacement

- **Pros**:
  - You will get better matches
  - Particularly helpful when you have very few control individuals
- **Cons:**
  - More complicated because matched controls are not independent.
  - Should be aware of how many times you are using one control.

Outline

The Goal of Balance

To what extent does the pre-match distribution of $X|T = 1$ look like the distribution of $X|T = 0$?

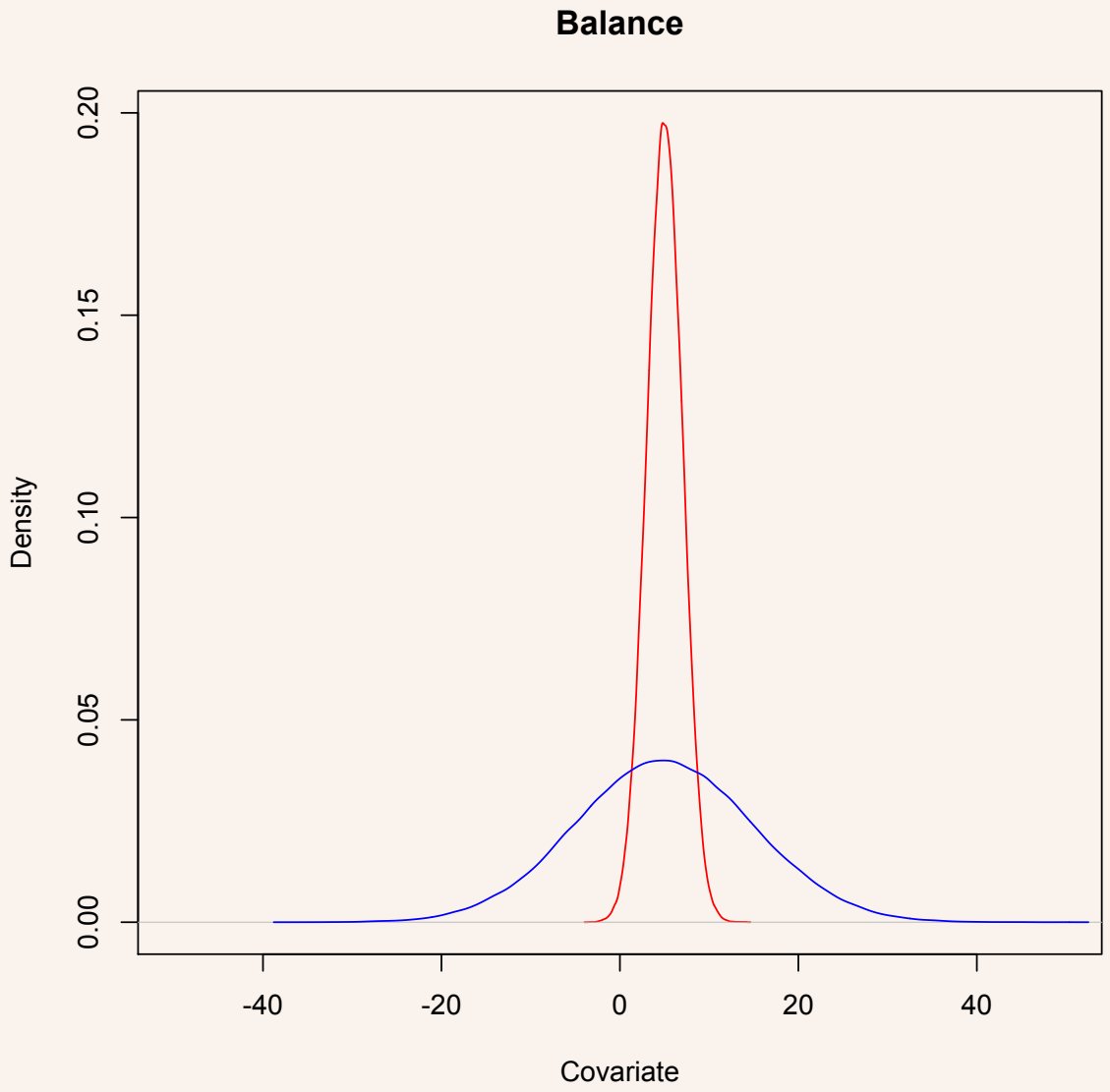
If they are very close, then we have matched well. Illustrative example: exact matching leads to identical multivariate distributions:

$$f(X|T = 1) = f(X|T = 0)$$

Common Measures of Balance

1. Mean or median of each stratifying variable
2. Variance, covariances
3. Empirical densities of each stratifying variable, QQ-plots
4. L_1

Mean or Median: What is wrong with this picture?



L_1

The idea is to divide the distributions of $X|T = 1$ and $X|T = 0$ each into k bins, sort of like a big multivariate (or univariate) histogram. Bin sizes are usually determined automatically.

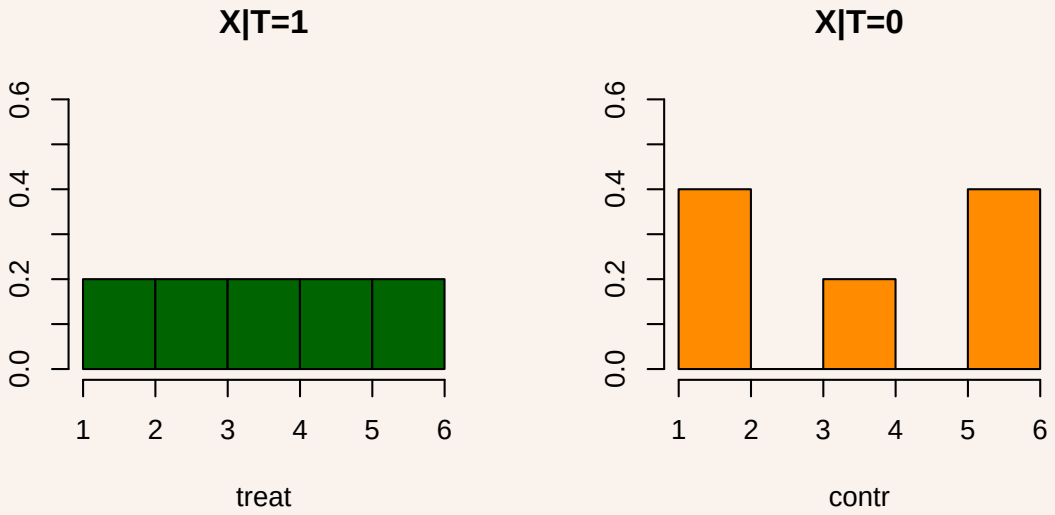
We then have a set of frequencies $f_1, \ldots, f_k$ where f_i is the proportion of treated observations which fall in bin i; likewise $g_1, \ldots, g_k$ are the proportions of control observations falling in bin i. Then

$$L_1(f, g) = \frac{1}{2} \sum_{i=1,\ldots,k} |f_i - g_i|.$$

If balance is perfect this equals 0; if completely imperfect, 1.

L_1

Here is a univariate example:



$$L_1(f,g) = \frac{1}{2}(.2 + .2 + 0 + .2 + .2) = .4.$$

Outline

Introducing the Data

Famous LaLonde dataset: an evaluation of a job training program administered in 1976 (which was not randomly assigned). Main outcome of interest is `re78`, retained earnings in 1978; sole treatment is the job training program (`treated`).

A variety of covariates on which to match:

- `age`, `education` (in years), `nodegree`
- `black`, `hispanic`, `married`
- `re74`, `re75`
- `u74`, `u75` both indicators of unemployment

Get the Data Setup

```
install.packages("MatchIt")
install.packages("cem")
library(MatchIt)
library(cem)
library(Zelig)

## LOAD DATA: Lalonde (1986) dataset; from cem package
data(LL)
```

Look at the Data Before Matching

```
mean(LL$re78[LL$treated == 1]) - mean(LL$re78[LL$treated == 0])
[1] 886.3038
```

```
summary(lm(re78 ~ treated + age + education + black + married +
    + re74 + re75 + hispanic + u74 + u75, data = LL))
```

```
Call:
lm(formula = re78 ~ treated + age + education + black + married +
    nodegree + re74 + re75 + hispanic + u74 + u75, data = LL)

Residuals:
   Min     1Q Median     3Q    Max
-10395  -4431  -1455   3189  53911

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.186e+03  2.638e+03   1.208   0.2276
treated      8.237e+02  4.685e+02   1.758   0.0791 .
age          1.024e+01  3.707e+01   0.276   0.7824
education    2.006e+02  1.805e+02   1.111   0.2669
black       -1.419e+03  8.019e+02  -1.770   0.0772 .
married      4.854e+01  6.522e+02   0.074   0.9407
nodegree    -2.996e+02  7.480e+02  -0.401   0.6889
re74         1.274e-01  7.526e-02   1.693   0.0909 .
re75         6.472e-02  9.145e-02   0.708   0.4794
hispanic     2.993e+02  1.050e+03   0.285   0.7758
u74          1.530e+03  9.378e+02   1.631   0.1033
```

Checking Imbalance in R

```
pre.imbalance <- imbalance(group=LL$treated,
  data=LL, drop=c("treated","re78"))

pre.imbalance

Multivariate Imbalance Measure: L1=0.735
Percentage of local common support: LCS=12.4%

Univariate Imbalance Measures:

               statistic   type           L1 min 25%       50%
age         1.792038e-01 (diff) 4.705882e-03   0   1   0.00000
education   1.922361e-01 (diff) 9.811844e-02   1   0   1.00000
married     1.070311e-02 (diff) 1.070311e-02   0   0   0.00000
...
re74       -1.014862e+02 (diff) 5.551115e-17   0   0  69.73096
re75        3.941545e+01 (diff) 5.551115e-17   0   0 294.18457
...
```

Exact Matching

```
exact.match <- matchit(formula= treated ~ age + education
    + black + married + nodegree + re74 + re75 + hispanic +
    u74 + u75, data = LL, method = "exact")
```

```
exact.match
Call:
matchit(formula = treated ~ age + education + black + married +
    nodegree + re74 + re75 + hispanic + u74 + u75, data = LL,
    method = "exact")

Exact Subclasses: 36

Sample sizes:
          Control Treated
All           425     297
Matched        74      55
Unmatched     351     242
```

Exact Matching

```
exact.data <- match.data(exact.match)
```

```
head(exact.data)
          u74 u75   weights subclass
16110   1   1 1.0000000        1
16141   1   1 1.0000000        2
... 16148   1   1 1.0000000        3
16156   1   1 0.6727273        3
16164   1   1 1.0000000        4
16182   1   1 4.0363636       14
```

What is the Treatment Effect?

```
lm(re78 ~ treated+ age + education + black
+ married + nodegree + re74 + re75
   + hispanic + u74 + u75, data = exact.data,
   weights=exact.data$weights)
```

Selected output:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3951.7       670.1   5.897 3.14e-08 ***
treated       1306.1      1026.2   1.273    0.205
```

Treatment Effect Again

```
y.treat <-
weighted.mean(exact.data$re78[exact.data$treated == 1],
  exact.data$weights[exact.data$treated == 1])


y.cont <-
weighted.mean(exact.data$re78[exact.data$treated == 0],
  exact.data$weights[exact.data$treated == 0])


y.treat-y.cont
[1] 1306.075
```

Nearest Neighbor Matching

```
nearest.match <- matchit(formula = treated ~ age + education
  + black + married + nodegree + re74 + re75 + hispanic +
  u74 + u75, data = LL, method = "nearest", distance = "logit")
```

Check balance post-matching:

```
data.matched <- match.data(nearest.match)
imbalance(group=data.matched$D, data=data.matched,
drop=c("D", "Y", "distance", "weights"))
```

OR

```
pre.balance <- summary(nearest.match)$sum.all
post.balance <- summary(nearest.match)$sum.match
```

Nearest Neighbor Matching

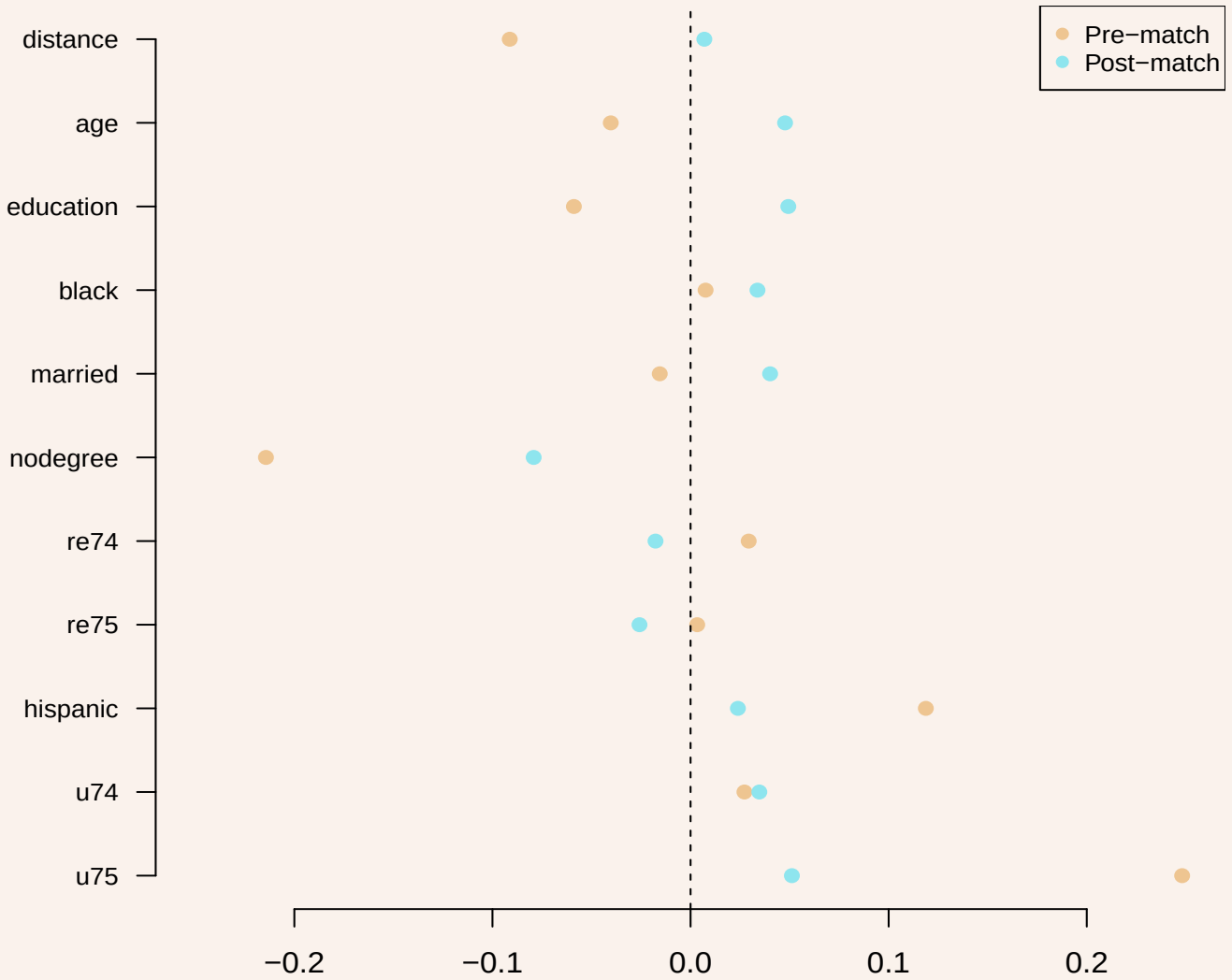
```
plot(x = (pre.balance[,1] - pre.balance[,2])/pre.balance[,3],
     y = 1:nrow(pre.balance), pch = 19, col = "burlywood2",
     xlab = "Standardized Imbalance", axes = FALSE,
     xlim = c(-.25,.25), ylab = "")

points(x = (post.balance[,1] - post.balance[,2])/post.balance[,3
     y = 1:nrow(post.balance), pch = 19, col = "cadetblue2")
abline(v = 0, lty = "dashed")


axis(1)
axis(2, at = 11:1, labels = rownames(pre.balance),
    las = 1, cex.axis = .8)


legend("topright", legend = c("Pre-match","Post-match"),
   pch = c(19,19),
   col = c("burlywood2","cadetblue2"), cex = .8)
```

Nearest Neighbor Matching



Nearest Neighbor Matching

```
nearest.data <- match.data(nearest.match)
```

```
head(nearest.data)
ack married nodegree      re74       re75
15993       1  33        12     0       1        0     0.000     0.0000
15994       1  20        12     0       1        0  8644.156  8644.1562
15995       0  39        12     1       1        0 19785.320  6608.1372
15996       1  49         8     0       1        1  9714.597  7285.9478
15997       0  26         8     0       1        1 37211.758 36941.2656
16000       0  27        12     1       1        0     0.000   752.3901
            re78 hispanic u74 u75  distance weights
15993 12418.0703        0   1   1 0.5237630       1
15994 11656.5059        0   0   0 0.5519635       1
15995   499.2572        0   0   0 0.4759660       1
15996 16717.1211        0   0   0 0.4460371       1
15997 30247.5000        0   0   0 0.3516366       1
16000     0.0000        0   1   0 0.5904726       1

## non-parametric estimate of the ATT
mean(nearest.data$re78[nearest.data$treated == 1]) - mean(nearest.data$re78[nearest.data$treated == 0])
[1] 1042.897

## A model-based estimate of the ATT

nearest.model <- lm(re78 ~ treated + age + education + black
  + married + nodegree + re74 + re75 + hispanic + u74 + u75,
  data = nearest.data)
```

CEM: Automatic Coarsening

```
auto.match <- matchit(formula = treated ~ age + education
  + black + married + nodegree + re74 + re75 + hispanic +
  u74 + u75, data = LL, method = "cem")
```

```
auto.match
Call:
matchit(formula = treated ~ age + education + black + married +
    nodegree + re74 + re75 + hispanic + u74 + u75, data = LL,
    method = "cem")

Sample sizes:
             Control Treated
All              425     297
Matched          222     163
Unmatched        203     134
Discarded          0       0
```

CEM: User Coarsening

```
re74cut <- hist(LL$re74, plot=FALSE)$breaks

re75cut <- seq(0, max(LL$re75)+1000, by=1000)

agecut <- c(20.5, 25.5, 30.5,35.5,40.5)

my.cutpoints <- list(re75=re75cut, re74=re74cut, age=agecut)

user.match <-  matchit(formula = treated ~ age + education
  + black + married + nodegree + re74 + re75 + hispanic +
  u74 + u75, data = LL, method = "cem", cutpoints
  = my.cutpoints)
```

CEM: User Coarsening

```
user.data <- match.data(user.match)
auto.data <- match.data(auto.match)

user.imb <- imbalance(group=user.data$treated, data=user.data,
drop=c("treated","re78","distance","weights","subclass"))

auto.imb <- imbalance(group=auto.data$treated, data=auto.data,
drop=c("treated","re78","distance","weights","subclass"))
```

CEM: User Coarsening

```
user.imb$L1

Multivariate Imbalance Measure: L1=0.437
Percentage of local common support: LCS=43.1%


auto.imb$L1

Multivariate Imbalance Measure: L1=0.592
Percentage of local common support: LCS=25.2%
```

CEM: Compare the Two

```
summary(user.match)$nn
           Control Treated
All            425     297
Matched        182     136
Unmatched      243     161
Discarded        0       0
```

```
summary(auto.match)$nn
           Control Treated
All            425     297
Matched        222     163
Unmatched      203     134
Discarded        0       0
```

CEM: Causal Effects

```
cem.match <- cem(treatment = "treated", data = LL,
    drop = "re78")
```

```
cem.match.att <- att(obj=cem.match, formula=re78 ~ treated,
    data = LL, model="linear")
```

```
cem.match.att # default is linear model
 G0  G1
All       425 297
Matched   222 163
Unmatched 203 134

Linear regression model on CEM matched data:

SATT point estimate: 550.962564 (p.value=0.368242)
95% conf. interval: [-647.777701, 1749.702830]
```

CEM: Causal Effects with a Model

```
cem.match.att2 <- att(obj=user.match, formula=re78 ~ treated +
age+education+black+married+nodegree+re74+re75
+hispanic+u74+u75,
data = LL, model="linear")


cem.match.att2
 G0  G1
All       425 297
Matched   222 163
Unmatched 203 134

Linear regression model on CEM matched data:

SATT point estimate: 550.962564 (p.value=0.368242)
95% conf. interval: [-647.777701, 1749.702830]

## Remaining imbalance? -> model with covariates
```