Module 1: Hospital Pricing and Selection on Observables

Part 3: Matching and Weighting

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Goal

Find covariates X_i such that the following assumptions are plausible:

1. Selection on observables:

$$Y_{0i}, Y_{1i} \perp\!\!\!\perp D_i | X_i$$

2. Common support:

$$0 < \Pr(D_i = 1 | X_i) < 1$$

Then we can use X_i to group observations and use expectations for control as the predicted counterfactuals among treated, and vice versa.

Assumption 1: Selection on Observables

$$E[Y_1|D,X] = E[Y_1|X]$$

In words…nothing unobserved that determines treatment selection and affects your outcome of interest.

Assumption 1: Selection on Observables

- Example of selection on observables from *Mastering Metrics*

Assumption 2: Common Support

Someone of each type must be in both the treated and untreated groups

$$0 < \Pr(D = 1|X) < 1$$

Causal inference with observational data

With selection on observables and common support:

1. Subclassification
2. Matching estimators
3. Reweighting estimators
4. Regression estimators

Subclassification

Sum the average treatment effects by group, and take a weighted average over those groups:

$$ATE = \sum_{i=1}^{N} P(X = x_i) \left(E[Y|X, D = 1] - E[Y|X, D = 0] \right)$$

Subclassification

- Difference between treated and controls
- Weighted average by probability of given group (proportion of sample)
- What if outcome is unobserved for treatment or control group for a given subclass?

- This is the *curse of dimensionality*

Matching: The process

1. For each observation i, find the m "nearest" neighbors, $J_m(i)$.

2. Impute $\hat{Y}_{0i}$ and $\hat{Y}_{1i}$ for each observation:

$$\hat{Y}_{0i} = \begin{cases} Y_i & \text{if} \quad D_i = 0 \\ \frac{1}{m} \sum_{j \in J_m(i)} Y_j & \text{if} \quad D_i = 1 \end{cases}$$

$$\hat{Y}_{1i} = \begin{cases} Y_i & \text{if} \quad D_i = 1 \\ \frac{1}{m} \sum_{j \in J_m(i)} Y_j & \text{if} \quad D_i = 0 \end{cases}$$

3. Form "matched" ATE:
$\hat{\delta}^{\text{match}} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}_{1i} - \hat{Y}_{0i} \right)$

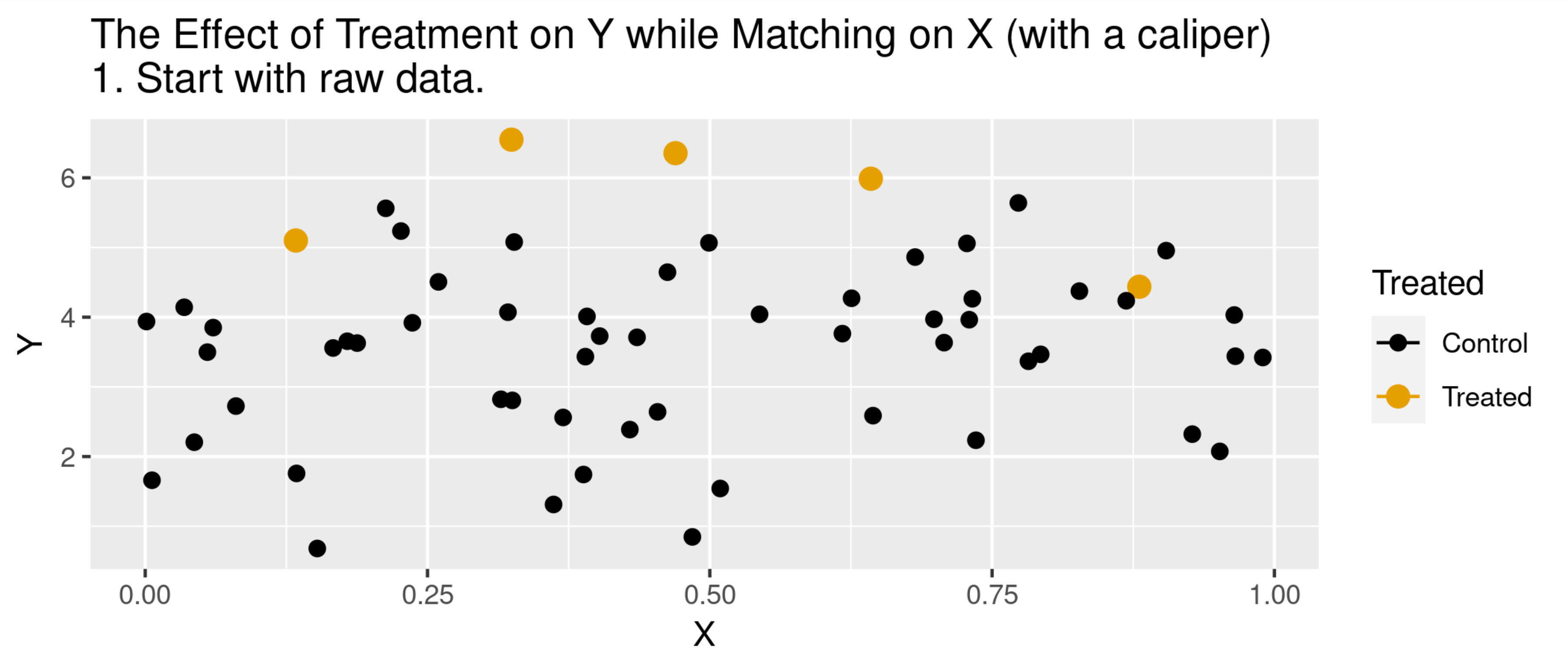
Matching: Defining "nearest"

1. Euclidean distance:
   $\sum_{k=1}^{K}(X_{ik} - X_{jk})^2$

2. Scaled Euclidean distance:
   $\sum_{k=1}^{K} \frac{1}{\sigma^2_{X_k}}(X_{ik} - X_{jk})^2$

3. Mahalanobis distance:
   $(X_i - X_j)'\Sigma_X^{-1}(X_i - X_j)$

Animation for matching



Matching: Defining "nearest"

- But are observations really the same in each group?
- Potential for "matching discrepancies" to introduce bias in estimates

- "Bias correction" based on

$$\hat{\mu}(x_i) - \hat{\mu}(x_{j(i)})$$

  (i.e., difference in fitted values from regression of y on x, with the difference between observed Y_{1i} and imputed Y_{0i})

Weighting

1. Estimate propensity score `ps ← glm(D~X, family=binomial, data)`, denoted $\hat{\pi}(X_i)$
2. Weight by inverse of propensity score

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N} \frac{Y_i D_i}{\hat{\pi}(X_i)}}{\sum_{i=1}^{N} \frac{D_i}{\hat{\pi}(X_i)}} \text{ and } \hat{\mu}_0 = \frac{\sum_{i=1}^{N} \frac{Y_i(1-D_i)}{1-\hat{\pi}(X_i)}}{\sum_{i=1}^{N} \frac{1-D_i}{1-\hat{\pi}(X_i)}}$$

3. Form "inverse-propensity weighted" ATE:

$$\hat{\delta}^{IPW} = \hat{\mu}_1 - \hat{\mu}_0$$

Regression

1. Regress Y_i on X_i among $D_i = 1$ to form $\hat{\mu}_1(X_i)$
2. Regress Y_i on X_i among $D_i = 0$ to form $\hat{\mu}_0(X_i)$
3. Form difference in predictions:

$$\hat{\delta}^{reg} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))$$

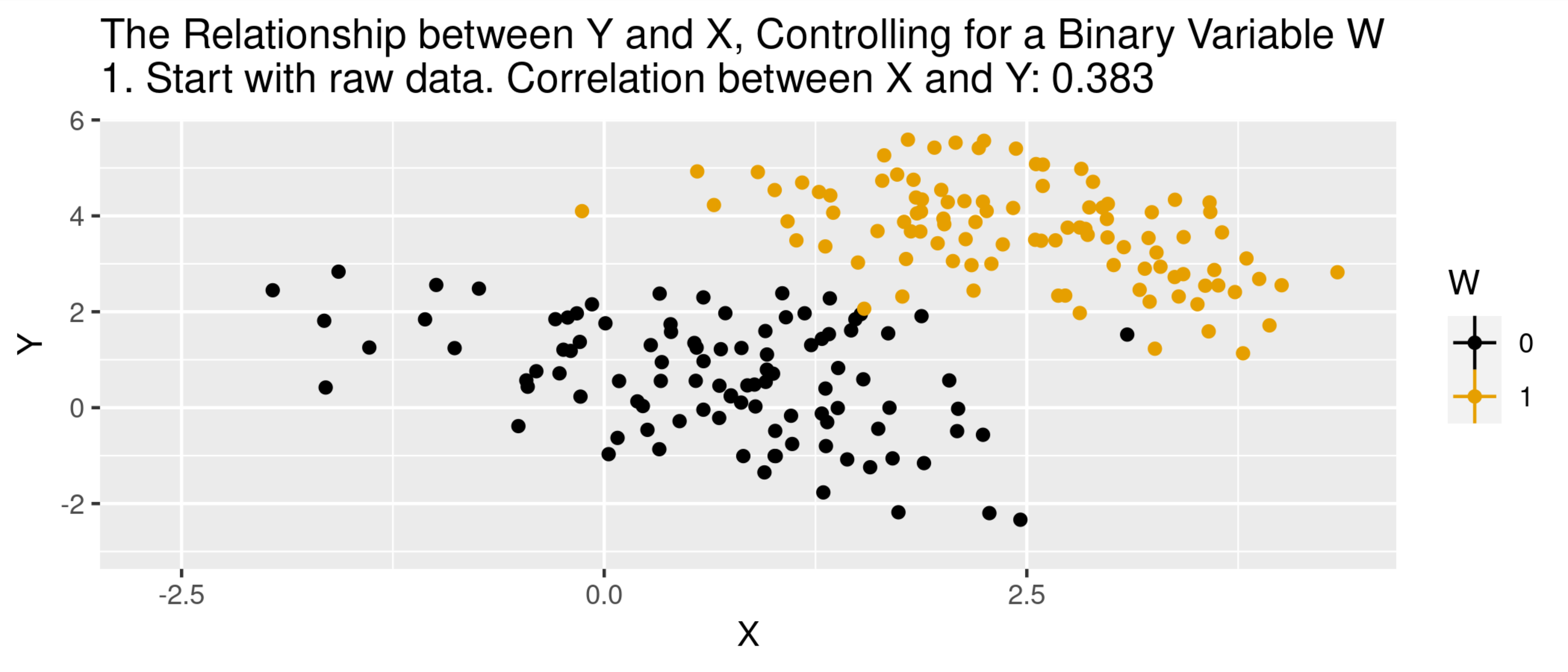
Regression

Or estimate in one step,

$$Y_i = \delta D_i + \beta X_i + D_i \times \left(X_i - \bar{X}\right)\gamma + \varepsilon_i$$

- Note the $\left(X_i - \bar{X}\right)$. What does this do?

Animation for regression



Simulated data

Now let's do some matching, re-weighting, and regression with simulated data:

```
n <- 5000
select.dat <- tibble(
  x = runif(n, 0, 1),
  z = rnorm(n, 0, 1),
  w = (x>0.65),
  y = -2.5 + 4*w + 1.5*x + rnorm(n,0,1),
  w_alt = ( x + z > 0.35),
  y_alt = -2.5 + 4*w_alt + 1.5*x + 2.25*z + rnorm(n,0,1)
)
```

Simulation: nearest neighbor matching

```r
nn.est1 <- Matching::Match(Y=select.dat$y,
                           Tr=select.dat$w,
                           X=select.dat$x,
                           M=1,
                           Weight=1,
                           estimand="ATE")
summary(nn.est1)
```

```
##
## Estimate...  3.8785
## AI SE......  0.53145
## T-stat.....  7.298
## p.val......  2.9199e-13
##
## Original number of observations..............  5000
## Original number of treated obs...............  1731
## Matched number of observations...............  5000
## Matched number of observations  (unweighted).  5013
```

Simulation: nearest neighbor matching

```
nn.est2 <- Matching::Match(Y=select.dat$y,
                           Tr=select.dat$w,
                           X=select.dat$x,
                           M=1,
                           Weight=2,
                           estimand="ATE")
summary(nn.est2)
```

```
##
## Estimate...  3.8785
## AI SE......  0.53145
## T-stat.....  7.298
## p.val......  2.9199e-13
##
## Original number of observations..............  5000
## Original number of treated obs...............  1731
## Matched number of observations...............  5000
## Matched number of observations  (unweighted).  5013
```

Simulation: regression

```
reg1.dat <- select.dat %>% filter(w==1)
reg1 <- lm(y ~ x, data=reg1.dat)

reg0.dat <- select.dat %>% filter(w==0)
reg0 <- lm(y ~ x, data=reg0.dat)
pred1 <- predict(reg1,new=select.dat)
pred0 <- predict(reg0,new=select.dat)
mean(pred1-pred0)
```

```
## [1] 4.126236
```

Violation of selection on observables

NN Matching

```r
nn.est3 <- Matching::Match(Y=select.dat$y_alt,
                           Tr=select.dat$w_alt,
                           X=select.dat$x,
                           M=1,
                           Weight=2,
                           estimand="ATE")
summary(nn.est3)
```

```
##
## Estimate...  7.6502
## AI SE......  0.053248
## T-stat.....  143.67
## p.val......  < 2.22e-16
##
## Original number of observations..............  5000
## Original number of treated obs...............  2756
## Matched number of observations...............  5000
## Matched number of observations  (unweighted).  22555
```

Regression

```r
reg1.dat <- select.dat %>% filter(w_alt==1)
reg1 <- lm(y_alt ~ x, data=reg1.dat)

reg0.dat <- select.dat %>% filter(w_alt==0)
reg0 <- lm(y_alt ~ x, data=reg0.dat)
pred1_alt <- predict(reg1,new=select.dat)
pred0_alt <- predict(reg0,new=select.dat)
mean(pred1_alt-pred0_alt)
```

```
## [1] 7.675315
```

What covariates to use?

- There are such things as "bad controls"
- We want to avoid control variables that are:

- Outcomes of the treatment
- Also endogenous (more generally)