Casual Inference Introduction

Foundational Problem of Casual Inference

- Drawing Conclusions on Causality
- In factual world want to know counterfactual
- Example: How does textbook affect math scores

- A = treatment, 1 if treated, 0 if not treated
- Y(1) = outcome of treated unit
- Y(0) = outcome of untreated unit

Potential Outcomes

- Stable Unit Treatment Value Assumption (SUTVA)
- No interference with treatment units
- Consistency and definition of treatment
- Missing Values

Unit	A	Y(0)	Y(1)
1	1	?	80
2	0	50	?
3	1	?	50
4	1	?	50
5	0	80	?
6	0	50	?
7	0	80	?
8	1	?	80

Identification

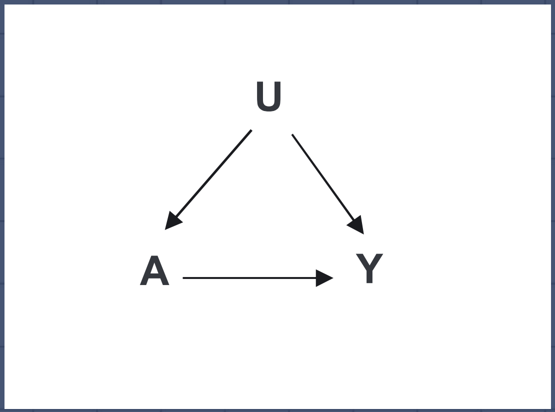
- Find casual estimand through statistical estimand
- Average Treatment Effect
- Can also use Regression

$$\frac{1}{\text{size of treatment group}} \sum_{i:A_i=1} Y_i - \frac{1}{\text{size of control group}} \sum_{i:A_i=0} Y_i$$

$$\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0]$$

Threat by Confounders

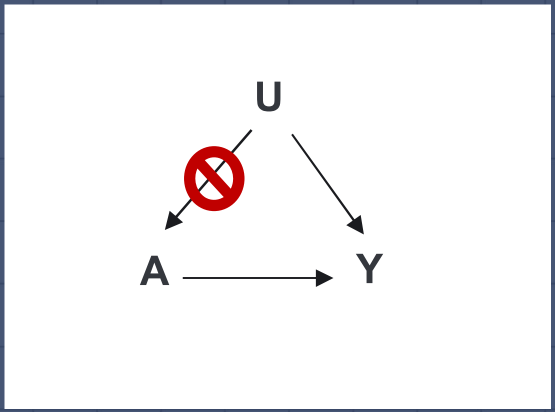
- 16 units of data collected
- U for confounder, baseline math score
- Lower probability of High Baseline Score being treated



Y(0)	Y(1)
80	80
80	80
80	50
80	50
50	50
50	50
50	50
50	50

Randomization for experimental study

- Assumption 1: SUTVA. If A = a, then Y = Y(a)
- Assumption 2: A is independent of U and Potential Outcomes
- Equal likelihood of treatment assignment



Y(0)	Y(1)
80	80
80	80
80	80
50	50
50	50
50	50
50	50
50	50

When Confounder is Known

- Standardization
  - Find average treatment effect within groups
  - Calculate average weighted mean
- Difference
  - Average treatment effect of high baseline math scores
  - Average treatment effect of low baseline math scores

$$\hat{ATE} = P(X = 1) * \hat{ATE}_{X=1} + P(X = 0) * \hat{ATE}_{X=0}$$

When confounder is known

- Inverse probability weighting (IPW)
  - Probability of being treated within group
  - Use weighted linear regression

$$\mathbb{E}[Y(a)] = \mathbb{E}\,\frac{I(A = a)Y}{P(A = a \mid X)}$$

Observational Studies

- Can’t assign assignment probabilities
- Higher dimensional data
- Ignorability

Matching R Package: Lalonde Dataset

- Job training program casual effect on real earnings in 1978
- Confounders
  - Age, years of education, race (Black or Hispanic), marital status, if they have a degree, real earnings in 1974 and 1975
  - Treatment is whether or not they did the job training program
  - re78 is real earnings in 1978.

age	educ	black	hisp	married	nodegr	re74	re75	treat	re78
37	11	1	0	1	1	0	0	1	9930.050
22	9	0	1	0	1	0	0	1	3595.890
30	12	1	0	0	0	0	0	1	24909.500
27	11	1	0	0	1	0	0	1	7506.150
33	8	1	0	0	1	0	0	1	289.790

Inverse Probability Weighting

- Find the propensity score
  - Probability of unit being treated
  - Use logistic regression to find propensity score

$$\mathbb{E}\frac{AY}{P(A=1|X)} - \mathbb{E}\frac{(1-A)Y}{1-P(A=1|X)}$$

```
ll.df = data.frame(lalonde)
ll.sub = ll.df[c(1:445), c(1:8,12)]

logreg = glm(treat ~ ., data = ll.sub, family = binomial)
ps = predict(logreg, type = "response")

ll.df$weights <- 1/(ws.df$treat*ps + (1-ws.df$treat)*(1-ps))
lm(re78 ~ treat, data = ws.df, weights = weights)
```

```
Call:
lm(formula = re78 ~ treat, data = ws.df, weights = weights)

Coefficients:
(Intercept)        treat
       4550         1641
```

Standardization or G-Computing

- Outcome regression of each group (treated and control)
- Apply coefficients to all units of data

$$\hat{E}[Y \mid A = 1, X] = \hat{\beta}_0 + \hat{\beta}_1^T X.$$

```r
ll.Nsub = ll.df[, c(1:9, 12)]
stzlin0 = lm(re78 ~ . -treat, data = ll.Nsub[ws.df$treat == 0,])
ps0 = predict(stzlin0, ll.df)

stzlin1 = lm(re78 ~ . -treat, data = ll.Nsub[ws.df$treat == 1,])
ps1 = predict(stzlin1, ll.df)

mean(ps1) - mean(ps0)
```

```
> mean(ps1) - mean(ps0)
[1] 1621.584
```

Bootstrapping to find Variance

- Take multiple (1000) samples from dataset
- Sample with replacement
- Go through each process (IPW and Standardization) multiple times
- Standard deviation of means

```
dim(ll.df)[1]
allipw = c()
fl_data = data.frame(lalonde)
for(i in 1:1000) {
  ll.df = fl_data[sample(1:445, 1000, replace = T), ]
  ll.Sub = ll.df[, c(1:8,12)]

  logreg = glm(treat ~ ., data = ll.Sub, family = binomial)
  ps = predict(logreg, type = "response")

  ll.df$weights <- 1/(ll.df$treat*ps + (1-ll.df$treat)*(1-ps))
  ipwreg = summary(lm(re78 ~ treat, data = ll.df, weights = weights))
  ipwC = ipwreg$coefficients[2,1]
  allipw = c(ipwC, allipw)
}
```

```
> mean(allipw)
[1] 1651.322
> sd(allipw)
[1] 450.3888
```

Bootstrapping to find Variance

- Take multiple (1000) samples from dataset
- Sample with replacement
- Go through each process (IPW and Standardization) multiple times
- Standard deviation of means

```
allstz = c()
for(i in 1:1000) {
  ll.df = fl_data[sample(1:445, 1000, replace = T), ]
  ll.Nsub = ll.df[, c(1:9, 12)]
  stzlin0 = lm(re78 ~ . -treat, data = ll.Nsub[ll.Nsub$treat == 0,])
  ps0 = predict(stzlin0, ll.Nsub)

  stzlin1 = lm(re78 ~ . -treat, data = ll.Nsub[ll.Nsub$treat == 1,])
  ps1 = predict(stzlin1, ll.Nsub)

  allstz = c(allstz, (mean(ps1) - mean(ps0)))
}
```

```
> mean(allstz)
[1] 1612.033
> sd(allstz)
[1] 468.5759
```

Comparison to Book Result

From Textbook:

	est	se
neyman	1794.343	670.9967
fisher	1676.343	677.0493
lin	1621.584	694.7217

```
> matchest.adj = Match(Y = y, Tr = z, X = x, BiasAdjust = TRUE)
> summary(matchest.adj)

Estimate...   2119.7
AI SE......   876.42
T-stat.....   2.4185
p.val......   0.015583
```

From Code:

- IPW result: 1641
- Standardization result: 1621.584

THANKS!

Any questions?