Segment 3: Causal Inference with Regression (in randomized studies)

Section 07: Posttreatment Variables

Do Not Adjust for Posttreatment Variables!!

- **General Recommendation:** When available, adjust for *pre-treatment* variables that predict the outcome
  - Pre-treatment covariate, X
  - Adjust for systematic imbalances
  - "Unlucky" randomizations
  - Increase precision of causal estimates
  - Randomization $\Rightarrow$ will not induce bias

- **Don't do this for variables measured post-treatment!**
  - Post-treatment (intermediate) variable, Q
  - Anything measured after treatment that can (in principle) be affected by the treatment
  - That is, anything that might have been different if a different treatment had been assigned
  - Even if the treatment is randomized!

The Setting: Posttreatment Variables

- Treatment, Z
- Outcome, Y
- Intermediate (or posttreatment) variable, Q

Key Points:

1. Q is an *outcome* because it is affected by treatment
   - I.e., should think about Q^0, Q^1, not just Q^{obs}
2. "Adjusting" comparisons of Y^0 and Y^1 for Q^{obs} will not estimate causal effects
   - E.g., with a regression model that adjusts for Q^{obs}

Adjusting for a Posttreatment Variable

"Adjusting" for Q

- Z affects $Q \Rightarrow$ treatment groups imbalanced on Q
  - Q^1 observed in treatment arm
  - Q^0 observed in control arm
- Q is related to $Y \Rightarrow Q$ is *unbalanced* $\Rightarrow$ distorts estimate of the effect on Y
- Adjusting for observed Q can mislead about causal effects
- This is known as **posttreatment selection bias**

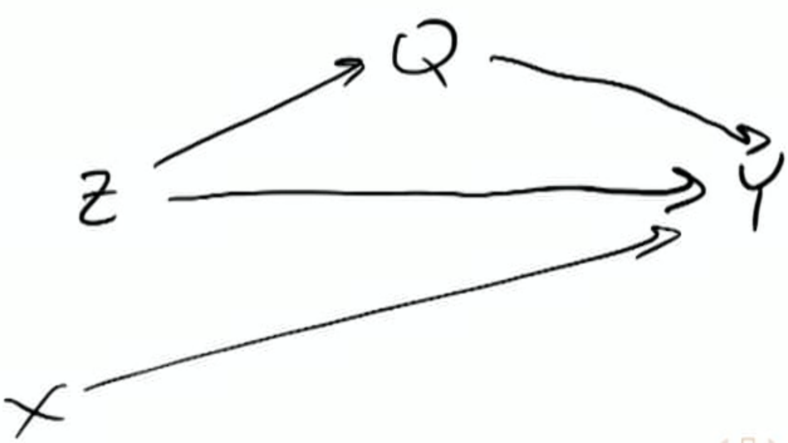
Key Idea: Need to Think About (Q_i^0, Q_i^1)

Z_i	X_i	Q_i^0	Q_i^1	Y_i^0	Y_i^1
0	$\star$	$\star$	?	$\star$	?
0	$\star$	$\star$	?	$\star$	?
0	$\star$	$\star$	?	$\star$	?
0	$\star$	$\star$	?	$\star$	?
...					
1	$\star$	?	$\star$	?	$\star$
1	$\star$	?	$\star$	?	$\star$
1	$\star$	?	$\star$	?	$\star$
1	$\star$	?	$\star$	?	$\star$

Example: Effect of Social Services on Child IQ

Z: Random assignment to social services (child care, home visits, $Z = 1$) vs. control ($Z = 0$)
Y: Child IQ score after 2 years
X: Parents' educational attainment
Q: continuous parenting quality measure $\in [0, 1]$

DAG



Social Services: Possible Regressions

1a) $Y_i = \alpha + \tau Z_i + \varepsilon_i \qquad Y^0, Y^1 \perp Z$

1b) $Y_i = \alpha + \tau Z_i + \beta X_i + \varepsilon \qquad Y^0, Y^1 \perp Z \mid X$

τ = ATE

2) $Y_i = \alpha^* + \tau^* Z_i + \beta^* X_i + \delta^* Q_i + \varepsilon_i$

$\tau^* \rightarrow \cancel{\text{ATE}} \qquad \cancel{Y^0, Y^1 \perp Z \mid X, Q}$

When we include Q, Tau does not retain the same meaning anymore, not going to represent the ATE. Also the ignorability assumption where the independence between potential outcomes in Z conditional on X and Q does not hold anymore.

Posttreatment Adjustment $\rightarrow$ Bias

Unit i	Pre-treat. Covariate X_i	Treatment Z_i	Observed intermediate outcome, Q_i	Potential intermediate outcomes Q_i^0	Q_i^1	Final outcome, Y_i
1	0	0	0.5	0.5	0.7	Y_1
2	0	1	0.5	0.3	0.5	Y_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$$Y_i = \alpha^* + \tau^* Z_i + \beta^* X_i + \delta^* Q_i + \varepsilon_i$$

Bias in the Social Services Example

- Regression adjusting for Q compares units with the same *observed* value of Q^{obs}
  - But these are really *potential outcomes* Q^0 or Q^1 depending on Z
- Two units with the *different* values of Z and the *same* value of Q^{obs} could be different in important ways
  - E.g., Innately higher (or lower) parenting ability
  - $\cancel{Y^0, Y^1 \perp\!\!\!\perp | X, Q}$
- If we could somehow know both (Q^0, Q^1), we could adjust for *that*
  - "Principal Stratification"

Causal Effects (Review, Summary)

The causal effect of assignment to Z on an outcome Y is defined to be a comparison between the sets of potential outcomes on a common set of units

$$\left\{Y_i^1; i \in \mathrm{set}_1\right\} \;\; vs. \;\; \left\{Y_i^0; i \in \mathrm{set}_2\right\}$$

given that the groups set_1 and set_2 are comparable or *balanced*

Covariate Adjustment

With randomized Z

1b) $Y_i = \alpha + \tau Z_i + \beta X_i + \varepsilon_i$

Compare observed outcomes among units in treatment groups *within strata of covariates*:

$$\underset{Y^1}{pr\{Y_i^{obs}}|X_i = x, Z_i = 1\} \quad vs. \quad \underset{Y^0}{pr\{Y_i^{obs}}|X_i = x, Z_i = 0\}$$

- If X are those required to satisfy the assumption of *ignorability*:
  - $Y^0, Y^1 \perp\!\!\!\perp Z|X$
  - $\Rightarrow$ This is a comparison between comparable sets of units
  - I.e., those with $X_i = x$
  - $\Rightarrow$ causal effect
- This is what regression does when we "adjust" for X
- (Estimates a causal effect if Z is randomized)

Posttreatment Variable Adjustment

With randomized Z

2) $Y = \alpha^* + \tau^* Z + \beta^* X + \delta^* Q$

Applying the same logic to the posttreatment variable, Q^{obs}:

$$pr\{Y_i^{obs}|Q_i^{obs} = q, Z_i = 1\} \quad vs. \quad pr\{Y_i^{obs}|Q_i^{obs} = q, Z_i = 0\}$$

$$pr\{Y_i^1|\underbrace{Q_i^1 = q}_{Set_1}\} \quad vs. \quad pr\{Y_i^0|\underbrace{Q_i^0 = q}_{Set_2}\}$$

$\neq$

- Problematic if treatment has any effect on the posttreatment variable
- $\{i : Q_i^0 = q\}$: Units with posttreatment value q under $Z = 0$
- $\{i : Q_i^1 = q\}$:Units with posttreatment value q under $Z = 1$
- Not the same types of units!
- **Comparison is not a causal effect**

"Standard" Method for Covariate Adjustment

With randomized Z

- Key point is that Q_i^{obs} encodes information about:
  - The treatment: how it affects Q
  - The unit: how it responds to treatment
- Compare against pre-treatment covariates, X, which only encode information about the unit
- Thus, the comparison:

$$pr\{Y_i^{obs}|Q_i^{obs} = q, Z_i = 1\} \quad vs. \quad pr\{Y_i^{obs}|Q_i^{obs} = q, Z_i = 0\}$$

estimates *net effects* of Z on Y.

  - Effect of Z on Y
  - Effect of Z on Q
  - (may be impossible to disentangle)

Example: Treatment Noncompliance

- Z: Randomized assignment to treatment
  - Random assignment to RIF/MMF
- Q: Indicator of which treatment was actually received
  - Receipt of RIF/MMF according to surgeon preference
- Y: Primary outcome of interest
  - Recovery complications

Unit i	Pre-treat. Covariate X_i	Treatment Z_i	Observed intermediate outcome, Q_i	Potential intermediate outcomes Q_i^0	Q_i^1	Final outcome, Y_i
1	0	0	0	0	1	Y_1
2	0	1	0	0	0	Y_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Example: Eelworms

- Z: Soil fumigants to kill eelworms applied to plots of land
- Q: Number of eelworm cysts per plot
- Y: The crop yield of the plot
- Part of the effect of Z on Y is "explained" by the effect on Q

Unit i	Pre-treat. Covariate X_i	Treatment Z_i	Observed intermediate outcome, Q_i	Potential intermediate outcomes Q_i^0	Q_i^1	Final outcome, Y_i
1	0	0	Low	Low	Low	Y_1
2	0	1	Low	High	Low	Y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Example: COVID Vaccine Trial

Randomly assign n=1000 patients

- Treatment: vaccine $(Z = 1)$ vs. placebo $(Z = 0)$
- Outcome: COVID infection $(Y = 1)$ or not $(Y = 0)$ at 6 months follow-up
- Antibody Response Q: low (L) or high (H) Ab response measured at 2 weeks follow up

Example: COVID Vaccine Trial

Control Arm
Treatment Arm
Unobserved

Z	Q^0	Q^1	$\bar{Y}^0$	$\bar{Y}^1$	
0	L	L	0.3	0.3	No change in $Q \Rightarrow$ no change in Y
0	L	H	0.2	0.06	
1	L	L	0.3	0.3	No change in $Q \Rightarrow$ no change in Y
1	L	H	0.2	0.06	

Vaccine does not affect infection when it does not increase Ab

Example: COVID Vaccine Trial

Control Arm
Treatment Arm
Unobserved

Z	Q^0	Q^1	$\bar{Y}^0$	$\bar{Y}^1$	
0	L	L	0.3	0.3	
0	L	H	0.2	0.06	Change in $Q \Rightarrow$ change in Y
1	L	L	0.3	0.3	
1	L	H	0.2	0.06	Change in $Q \Rightarrow$ change in Y

Vaccine prevents infections when it raises Ab

Example: COVID Vaccine Trial

Control Arm
Treatment Arm
Unobserved

Adjust for Q^{obs}

Z	Q^0	Q^1	$\bar{Y}^0$	$\bar{Y}^1$
0	L	L	0.3	0.3
0	L	H	0.2	0.06
1	L	L	0.3	0.3
1	L	H	0.2	0.06

$$\bar{Y}^0 = \frac{0.3+0.2}{2} = 25.0\%$$

$$\bar{Y}^1 = \frac{0.3}{1} = 30.0\%$$

$$p(Y=1|Z=1, Q^{obs}=L) \neq p(Y=1|Z=0, Q^{obs}=L)$$

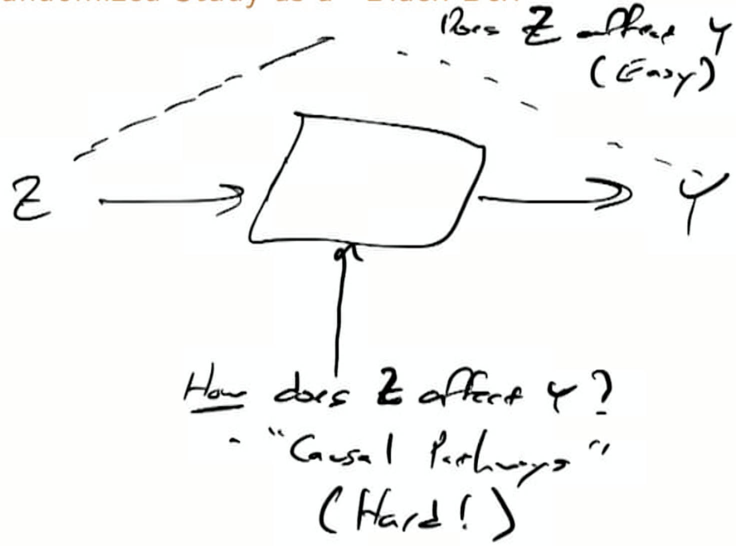
After adjusting for Ab, the vaccine causes more infections (!?)

Adjusting for Q

Vaccine $\Rightarrow$ more COVID?

- Vaccine appears to cause high Ab, on average
  - More $Q^{obs} = H$ in the $Z = 1$ group
- Vaccine/Placebo groups not comparable within observed level of Q
- "Adjusting" for Q leads to comparisons between **different** patients in each treatment group
- **Posttreatment selection bias**

Randomized Study as a "Black Box"



Example: Effect of Social Services on Child IQ

Z: Random assignment to social services (child care, home visits, $Z = 1$) vs. control ($Z = 0$)
Y: Child IQ score after 2 years
X: Parents' educational attainment
Q: continuous parenting quality measure $\in [0, 1]$

- Suppose Z increases Y by 10 points (on average)
  - Inferring this relatively easy since Z randomized

- "Causal Paths" Question: To what extent is the apparent treatment effect *because* of improved parenting (Q)?
  - Does the intervention *indirectly* affect IQ "through" improving parenting?
  - Does the intervention *directly* affect IQ "through" some other pathway?
  - To what extent is the intervention effect on IQ *mediated* through an effect on parenting?

Regression $\neq$ Pathways

There are suggestions in the literature for answering "causal pathways" questions using a sequence of regression models:

1. Fit a regression model that *does not* adjust for the intermediate parenting quality

$$Y_i = \alpha + \tau Z_i + \varepsilon_i$$

2. Fit a regression model that *does* adjust for the intermediate parenting quality

$$Y_i = \alpha^* + \tau^* Z_i + \delta^* Q_i + \varepsilon_i$$

Conclude something about pathways based on comparing estimate of τ, τ^*:

$$(\tau \neq 0) + (\tau^* = 0) \Rightarrow Z \text{ affects } Y \text{ "through" } Q$$
$$(\tau \neq 0) + (\tau^* \neq 0) \Rightarrow \text{Some "direct" effect of } Z$$

Generally not appropriate without many additional assumptions

Regression $\neq$ Pathways

Even in randomized experiments

We should recognize that a regression of the form:

$$Y_i = \alpha^* + \tau^* Z_i + \delta Q_i + \varepsilon_i$$

~~ATE~~

Will not necessarily answer causal pathways questions (posttreatment selection bias)

Makes comparisons of the form:

$$\{i; Z_i = 0, Q^0 = q\} \quad \text{vs.} \quad \{i; Z_i = 1, Q^1 = q\}$$

Hypothetical Underlying Data

Parenting potential	Parenting quality after assigned to control	treat	Child's IQ score after assigned to control	treat	Proportion of sample
Poor parenting either way	Poor	Poor	60	70	0.1
Good parenting if treated	Poor	Good	65	80	0.7
Good parenting either way	Good	Good	90	100	0.2
	Q^0	Q^1	Y^0	Y^1	

Adjusting for Q^{obs} "Breaks" Randomization

For example a regression model that adjusts for Q^{obs} entails comparisons such as:

$$\{i; Z_i = 0, Q^0 = \text{Good}\} \quad \text{vs.} \quad \{i; Z_i = 1, Q^1 = \text{Good}\}$$

Is a comparison between:

- $Z = 0, Q^0 =$"Good"
  - $Q^0 =$"Good", $Q^1 =$"Good" parents that would have good parenting regardless of treatment

vs.

- $Z = 1, Q^1 =$"Good"
  - $Q^0 =$"Good", $Q^1 =$"Good" parents that would have good parenting regardless of treatment
  - $Q^0 =$"Poor", $Q^1 =$"Good" parents that would have good parenting if treated

Summary: Don't Adjust for Posttreatment Variables

- Benefits of randomization don't extend to adjustment for posttreatment variables
  - Z randomized $\rightarrow$ adjusting for X in a regression does not change the *causal estimand*/interpretation of the causal effect
  - But adjusting for Q does....no guarantee that we are still estimating a causal effect of Z
  - (even more complicated when Z is not randomized)
- Generally, we cannot use regression to answer questions about "causal pathways" when Z is randomized
  - At least not without specialized assumptions
- There are causal methods designed to address these questions through formalizing such assumptions
  - Principal stratification
  - Instrumental variables
  - "Causal mediation analysis"