Statistics II

Week 8: **Final Exam Review**

General Questions

General Questions

- For the **application papers**, be prepared to answer:
  - What was the research question?
  - Which methods did the researchers use?
  - How can we interpret the model outputs (the ones shown on the slides)?

- Study up on all of the **assumptions** listed on the topic list.
  - IV assumptions, parallel trends, etc.

Vocabulary

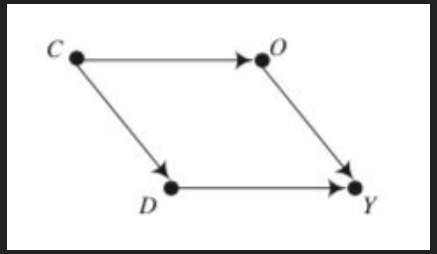
Vocabulary:

- **Parametric** vs. **non-parametric**: Parametric tests assume underlying statistical distributions in the data, while non-parametric tests do not.
- Types of studies: **observational** vs. **experiments** vs. **natural experiments**
  - With experiments treatment is in control of the researcher, with observational studies, it is not. Natural experiments are a special kind of observational study where treatment is as-if randomly assigned.
- **Exogeneity** = external vs. **endogeneity** = internal.
- **SUTVA**: The potential outcomes of individuals must be unaffected by changes in the treatment exposures of all other individuals.

DAGs

Back-door paths

- A **back-door path** is defined as any path between the causal variable and the outcome variable that begins with an arrow that points to the causal variable.



- Satisfying the **back-door criterion** just means you closed all the back-door paths.

POF

Basic Idea

- Each individual in a sample could **theoretically** be exposed to two states with **different potential outcomes**.
- Key assumption of the model is that each individual in the population of interest has a potential outcome under each treatment state, even though each individual can be observed in only one treatment state at any point in time. The other outcome will always be a **counterfactual**.
- Therefore, we focus on **average effects** across larger samples.
- We can apply the POF to basically any research question. It is a tool that is helpful for understanding our research question theoretically.

Notation

$$E(Y^1|D=1,\mathbf{Z}) = E(Y^1|D=0,\mathbf{Z}),\ E(Y^0|D=1,\mathbf{Z}) = E(Y^0|D=0,\mathbf{Z})$$

E: Expected outcome

Y1: Outcome under treatment

Y0: Outcome without treatment

D = 1: Treatment group

D = 0: Control group

Z: Set of control variables

The expected outcome of with treatment would be the same for those in the treatment group as those in the control group, conditioned on a set of control variables. Similarly, the expected outcome without treatment would be the same for those in the treatment or control groups, conditioned on a set of control variables.

Bias of NATE

Baseline Bias: The difference in average outcomes between the treatment and control group without treatment. Basically, how are these groups just naturally different before any treatment occurs?

Differential Treatment Effect Bias: The difference in average treatment effect between the treatment and control groups (weighted by the proportion of the population in the control group). How would each group respond differently to the treatment, if they both were to receive treatment?

Linear Regression

Mechanics of OLS

- We want to draw the “best” line that summarizes the relationship between X and Y.
- We define this as the line that **minimizes the sum of squared residuals**.
- General process:
  - Measure the distance between each data point and a hypothetical line. This is the residual.
  - Square the residual (to cancel out negatives and punish outliers)
  - Sum together all the squared residuals
  - Draw another line and repeat
  - Choose the line with the lowest total.

Btw: An **error** is the difference between the observed value and the true value (very often unobserved), while a **residual** is the difference between the observed value and the predicted value.

POF and Regression

If we put this structural model,

$$y_i = \beta_0 + \beta_1 D + e_i$$

in POF notation:

$$E(Y^0|D = 0) = \beta_0$$

$$E(Y^1|D = 1) = \beta_0 + \beta_1$$

$$\beta_1 = NATE$$

Estimating Bias in Coefficients when there is OVB

Suppose Y is test scores, A is attentiveness in class, and B is hours of study. The population regression equation is:

$$Y = \beta_0 + \beta_1 A + \beta_2 B$$

But we omit B:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 A$$

Estimating Bias in Coefficients when there is OVB

	A & B are positively correlated	A & B are negatively correlated
Y & B are positively correlated	**Positive Bias**	**Negative Bias**
Y & B are negatively correlated	**Negative Bias**	**Positive Bias**

Selecting Relevant Covariates

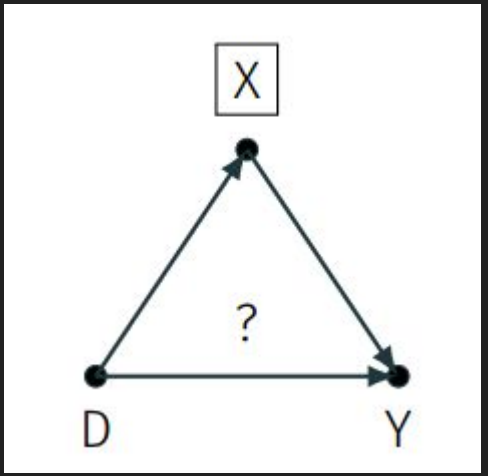
Selecting the variables to include is a theoretical, and not statistical problem.

- Start with theory to generate your covariates.
- Use DAGs and tricks like the backdoor criterion to narrow it down.
- Be cautious about conditioning on colliders (which can lead to **endogenous selection bias**) or mediators (which can lead to **post-treatment bias**).

Post-Treatment Bias

If we condition on a **mediator**, we “control away” the indirect effect of D on Y.

There is a **trade-off** between post-treatment control and OVB.



Matching

Randomization

Remember: Causal effects are identified if treatment is randomly assigned.

- **Unconditional randomization**: Randomly assign a fixed number of subjects from the study population to the treatment.
    - This can be inefficient if the outcome varies vastly for reasons unrelated to the treatment. For example, environmental factors may vary widely in treatment and control groups and have more of an effect than the treatment itself, making treatment effects challenging to identify.
- **Conditional randomization:** Choose relevant covariates pre-treatment (if they are knowable), build/stratify groups within particular combinations of these covariates, and then randomly assign treatment within these groups.
- **Paired randomization:** Two subjects per combination of covariate value, one of which is randomly assigned to treatment.

Stratification and Standardization

- **Stratification**: Covariate levels (or combinations thereof) define groups within which the treatment is randomly assigned.
  - Treatment effects are calculated separately for each group

- After we have these effects for each group, we can **standardize** to get the ATE.

$$ATE = \sum_{X=x}[E(Y^1|D=1, X=x) - E(Y^0|D=0, X=x)]P(X=x)$$

Conditional Independence Assumption & Selection on Observables

- The **conditional independence assumption** says that if we condition on the observed variables that affect treatment assignment and treatment-specific outcomes, the expected value for the treatment group under treatment is the same as for the control group under treatment (and same for both without treatment).
- This assumption is also known as a **selection-on-observables** assumption because its central tenet is the observability of the common variables that generate the dependence.

Exact Matching

Matching is the observational-study cousin of paired randomization.

1. Use theoretical and empirical knowledge to identify relevant confounder(s) *X*

2. Starting from treated subjects, select at least one match from the control group with exactly the same value(s) on *X*

3. Drop subjects off “common support” (unmatched subjects)

4. Estimate causal effect as the average difference in Y across pairs of matched subjects:

$$E(Y \mid D = 1, X) - E(Y \mid D = 0, X)$$

Common Support

Name	Y	D	X
Jake	10	1	3
Gina	12	1	3
Terry	8	1	2
Rosa	6	0	3
Charles	3	0	2
Ray	1	0	1

Common Support

Name	Y	D	X
Jake	10	1	3
Terry	8	1	2
Rosa	6	0	3
Charles	4	0	3
Ray	2	0	2

Non-exact Matching

Exact matching can lead to a very small final sample size, especially if there are lots of covariates to consider.

Instead, we can match based on **propensity scores:** a measure of the probability of a unit of being in the treatment group. Propensity scores are based on the idea that we want to match observations on treated units, and therefore want to find control observations that *look* as if they are treated.

> In other words, based on the characteristics of this unit, what is the likelihood that it has been treated?

This is usually modeled with logit/probit regression.

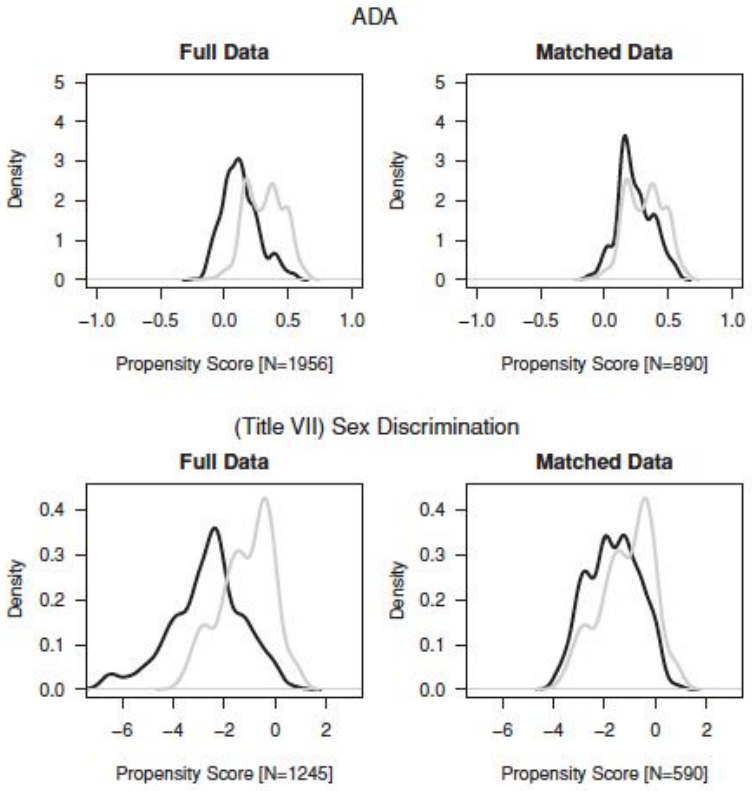
Logistic Regression

Similar to linear regression, except we're working with a binary categorical outcome variable.

Instead of fitting a line to the data, logistic regression fits an S-shaped curve (sigmoid) that goes from 0 to 1; It tells you the probability of outcome based on the covariate(s) - this is our propensity score!

Coefficients are presented in terms of log(odds), so be careful with interpretation: it is not as straightforward as with linear regression.

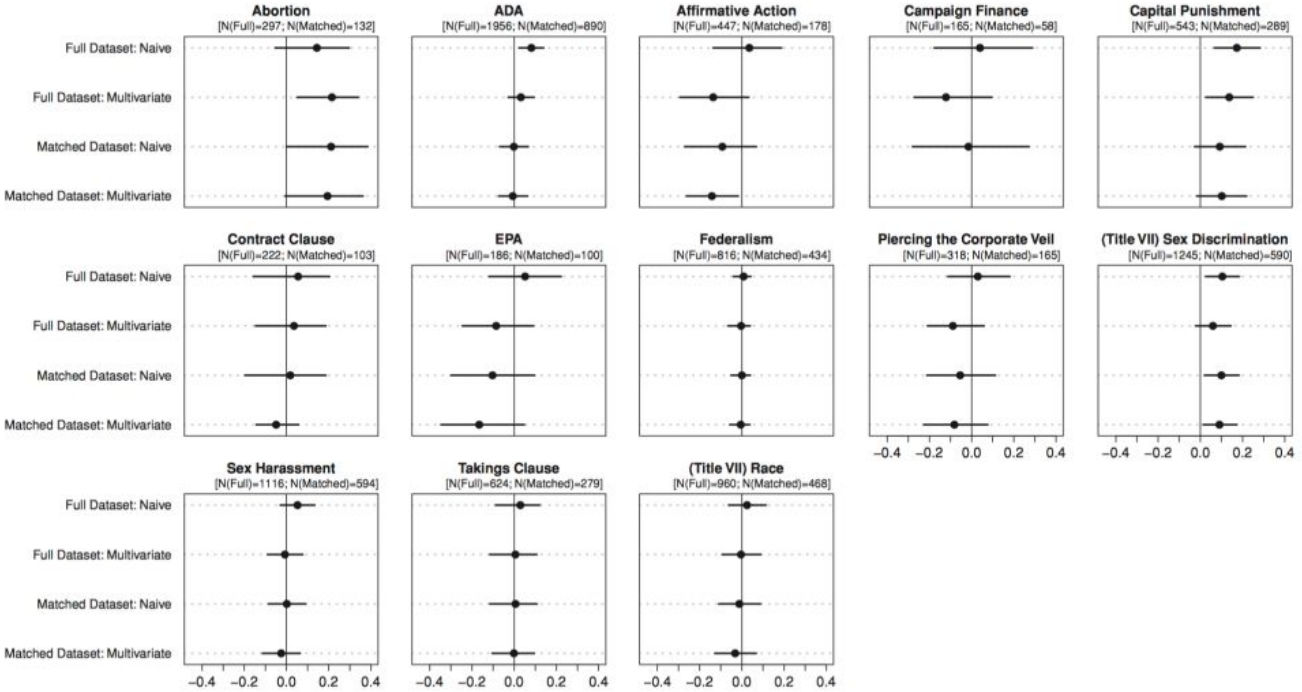
FIGURE 2 Kernel Density Plots of the Estimated Propensity Score for the ADA and Title VII Sex Discrimination Individual Effects Analyses



The black lines depict the density for all-male panels (control); the grey lines for mixed-sex panels (treatment). Each left-hand panel represents the full datasets while the right-hand panels display the propensity scores for only the matched data.

Propensity score distribution before and after the match.

FIGURE 4 **Dotplots of Average Treatment Effects (ATEs) for Individual Effects Across 13 Issue Areas**



The lines represent 95% confidence intervals for the average treatment effect. For every issue area, the first two models are logistic regression models fit to each full, unbalanced dataset. The naive model includes only the judge's sex as a covariate. The other model includes the judge's sex and a number of controls, including ideology. The next two models show the ATE after nearest-neighbor matching with replacement on the estimated propensity score. The first is for a difference of proportions analysis. The second is for a logistic regression model with the judge's sex and a number of controls including ideology.

Dotplot of effect coefficients

Balance Tests

Difference-in-means test (t-test) of our pretreatment variables *X* for the treatment and control groups.

> Reminder: Propensity score matching is done using the probability of a unit of being in the treatment group based on a set of variables *X*. Our balance tests are performed on these set of *X*.

Capitalize on the propensity score tautology: If treatment probability is identically distributed in treatment and control group, all *X* will be balanced.

- If *X* are balanced, then estimated propensity scores are ok
- If *X* are not balanced, re-specify treatment probability model

A balance table shows difference between unmatched and matched groups on various *X* variables.

TABLE 2 Matching Summary Statistics for the Individual Effects Analyses for ADA and Title VII Sex Discrimination Cases

	ADA Cases						
	Full Data ($N = 1956$)				Matched Data ($N = 890$)		
Variable	Mean Treated	Mean Control	eQQ Med	Percent Reduction	Mean Treated	Mean Control	eQQ Med
Propensity Score	0.32	0.13	0.19	94.89	0.32	0.31	0.09
Minority Judge	0.09	0.11	0.00	·	0.09	0.12	0.00
Judicial Experience	0.47	0.47	0.00	·	0.47	0.48	0.00
Judicial Common Space	−0.17	0.06	0.17	98.04	−0.17	−0.17	0.06
Confirmation Year	1991.14	1985.17	5.00	92.60	1991.14	1990.70	2.00
	(Title VII) Sex Discrimination Cases						
	Full Data ($N = 1245$)				Matched Data ($N = 590$)		
Variable	Mean Treated	Mean Control	eQQ Med	Percent Reduction	Mean Treated	Mean Control	eQQ Med
Propensity Score	−1.13	−2.75	1.58	91.67	−1.13	−1.27	0.57
Minority Judge	0.12	0.09	0.00	30.39	0.12	0.14	0.00
Judicial Experience	0.45	0.45	0.00	·	0.45	0.43	0.00
Judicial Common Space	−0.12	0.10	0.16	81.48	−0.12	−0.08	0.11
Confirmation Year	1990.38	1984.58	6.00	98.12	1990.38	1990.27	2.00

The left portion of each table provides results for the full, unmatched data, while the right portion displays results after matching has taken place. eQQ med is the median difference in the empirical quantile-quantile plot (an eQQ med of zero is ideal).

Balance statistics before and after the match.

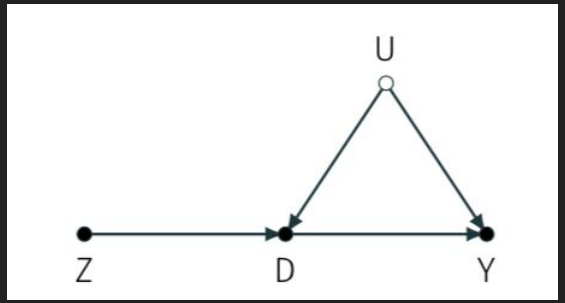
Instrumental Variables

Motivation

- Often we cannot force subjects to take a treatment, and those who choose to take a treatment may systematically differ from those who do not (selection bias).
- So, we can focus on **randomized encouragement** to take the treatment instead.
- **Instrumental variable studies** are the observational-study cousin of randomized encouragement.

Instrumental Variables

- **Basic idea**: If you're trying to estimate the effect of D on Y, but there are unobserved confounders U, we can use an exogenous variable Z that isn't affected by U to measure the unique variation in D from Z.
- To study this, we split the variation of D into **two parts**:
  - One **potentially related to** (potentially unobserved) **confounders U**
  - One **truly exogenous**
- To find the portion of D unrelated to confounders, one needs a variable Z (the instrumental variable) that is (as if) randomly assigned and related to D



Intent to Treat Effect (ITT)

- Effect of encouragement itself on the outcome regardless of actual treatment. It only considers assignment to the treatment or control groups.

- Because Z is randomized, ITT is identified by difference in means between the encouraged and unencouraged:

$$ITT = E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$$

Compliance Types

- Some people will always take the treatment, regardless of whether they are in treatment or control (**always-takers**),
- and some never will (**never-takers**).
- Some will always do as their told (**compliers**),
- and some will always do the opposite (**defiers**).

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Complier/Never-taker	Defier/Never-taker
$D_i = 1$	Defier/Always-taker	Complier/Always-taker

We cannot directly identify the group to which any particular respondent belongs.

IV Assumptions

1. Relevance, or nonzero average encouragement effect.
    - Encouragement needs to make a difference.
    - Testable - we can see if there is a difference between treatment and control groups.

$$E(D_i(1) - D_i(0)) \neq 0$$

IV Assumptions

2. Exogeneity/Ignorability of the instrument: Potential treatments and outcomes must be independent of Z (no OVB).
    - Hypothetical potential outcomes shouldn’t be related to Z.
    - Given by quasi-randomization of encouragement; matter of plausibility.
    - Not empirically testable.

$$\{Y_i(0), Y_i(1), D_i(0), D_i(1)\} \perp\!\!\!\perp Z_i$$

IV Assumptions

3. Exclusion restriction: Instrument affects outcome only via treatment.
    a. Potential outcome under encouragement should be same as potential outcome under control regardless of whether treated with encouragement or not.
    b. Implies that variation in the instrument does not change the potential outcomes other than through its effect on D
    c. Implies zero ITT effect for always-takers/never-takers.
    d. Hardly testable!

$$Y_i(z = 1, d) = Y_i(z = 0, d) \; for \; d = 0, 1$$

IV Assumptions

4. Monotonicity: No defiers.
    a. Also hardly testable; matter of plausibility

$$D_i(1) \geq D_i(0) \; for \; all \; i.$$

Decomposing the ITT effect

- ITT effect can be decomposed into combination of subgroup ITTs:

$$\begin{aligned}\text{ITT} = \text{ITT}_c \times \Pr(\text{compliers}) + \text{ITT}_a \times \Pr(\text{always-takers}) \\ + \text{ITT}_n \times \Pr(\text{never-takers}) + \text{ITT}_d \times \Pr(\text{defiers})\end{aligned}$$

where

$$\begin{aligned}\text{ITT}_c &= E[Y_i(1, D_i(1)) - Y_i(0, D_i(0)) | D_i(1) = 1, D_i(0) = 0], \\ \text{ITT}_a &= E[Y_i(1, D_i(1)) - Y_i(0, D_i(0)) | D_i(1) = D_i(0) = 1], \text{ etc.}\end{aligned}$$

- Under monotonicity and exclusion restriction, this simplifies as:

$$\begin{aligned}\text{ITT} &= \text{ITT}_c \times \Pr(\text{compliers}) + \text{ITT}_a \times \Pr(\text{always-takers}) \\ &\quad + \text{ITT}_n \times \Pr(\text{never-takers}) + 0 \quad \textbf{[monotonicity!]} \\ &= \text{ITT}_c \times \Pr(\text{compliers}) + 0 \times \Pr(\text{always-takers}) \\ &\quad + 0 \times \Pr(\text{never-takers}) \quad \textbf{[exclusion restriction!]} \\ &= \text{ITT}_c \times \Pr(\text{compliers})\end{aligned}$$

IV estimand and interpretation

- ITT_c can be identified as:

$$\text{ITT}_c = \frac{\text{ITT}}{\Pr(\text{compliers})} = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)}$$

- ITT_c can be interpreted as **Local Average Treatment Effect (LATE)** for compliers:

$$\text{ITT}_c = \text{LATE}_c = E(Y_i(1) - Y_i(0)|D_i(1) = 1, D_i(0) = 0]$$

- ITT_c is sometimes also called "Complier Average Causal Effect" (CACE)
- LATE has a clear causal meaning, but its **interpretation is often tricky**:
  - Compliers are defined in terms of principal strata, so we can never identify who they actually are
  - Different encouragement (instrument) yields different compliers

Calculating LATE

LATE can be calculated using the **Wald Estimator:**

$$LATE = \frac{cov(Y_i, Z_i)}{cov(D_i, Z_i)}$$

or **Two-Stage Least Squares (2SLS)**:

1. Regress cause D on instrument Z
2. Regress outcome Y on estimate of cause D from stage 1

2SLS only retains the variation in D that is generated by quasi-experimental variation in Z.

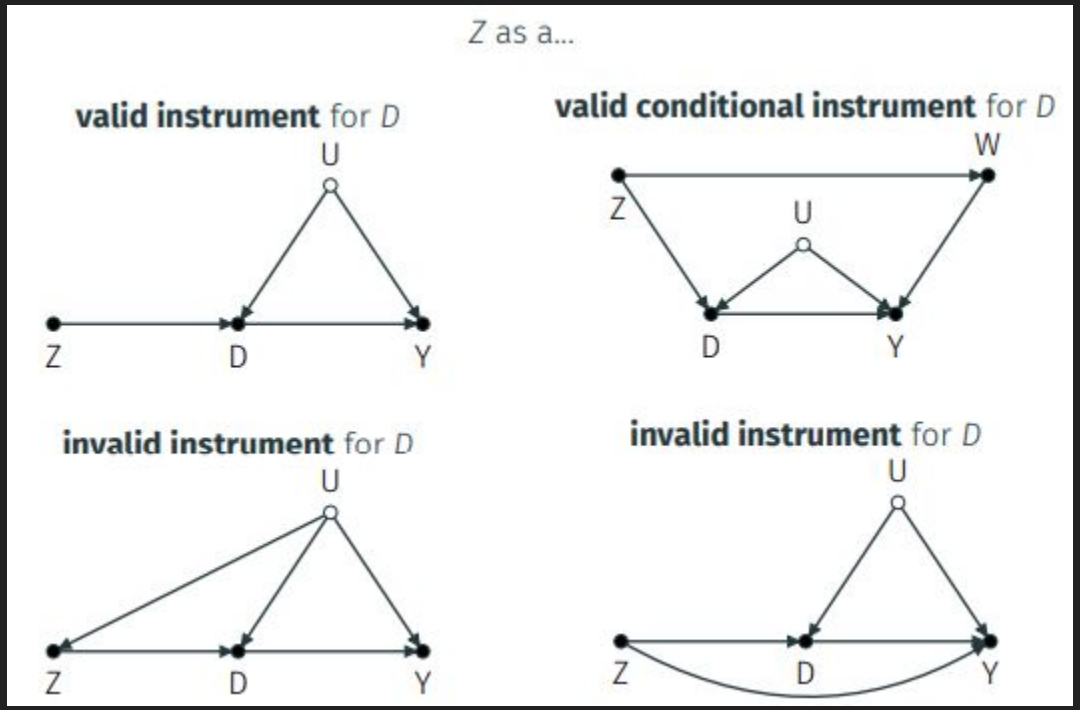
Calculating LATE

Treatment status by assignment

	Draft eligibility	
Military service	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	5,948	1,915
$D_i = 1$	1,372	865

- Average (log) earnings among eligibles is 5.411, among non-eligibles 5.440
- **So what's our estimate of the LATE?**
  - $E(D|Z = 1) = \frac{865}{(865+1915)} = 0.311$
  - $E(D|Z = 0) = \frac{1372}{(1372+5948)} = 0.188$
  - Denominator of LATE: $0.311 - 0.188 = 0.123$
  - LATE $= \frac{(5.411-5.440)}{0.123} = -0.236$

IVs from DAG perspective



Questions

- IV assumptions
- ITT and LATE
- Calculating LATE

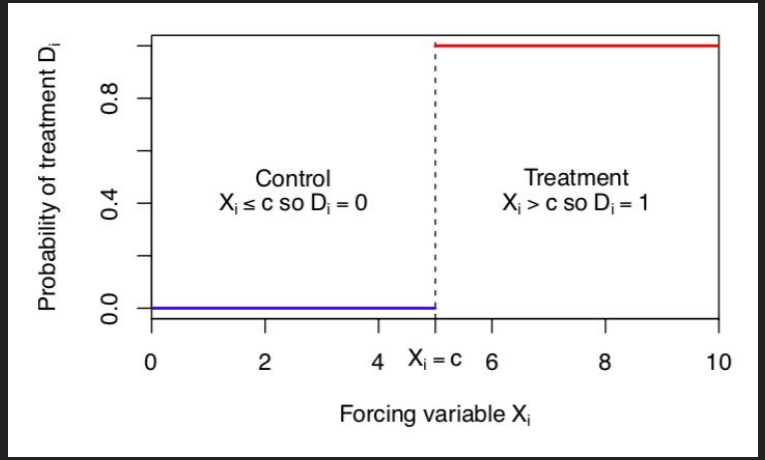
Regression Discontinuity Designs

Core Idea

- Treatment is assigned according to a rule based on another variable (called the **forcing** or **running** variable).
- Treated and untreated units may differ in their potential outcomes (non-random selection into treatment) based on the forcing variable. There might be other factors that determine the forcing variable outcome.
- However, whether units end up just below or just above the threshold is assumed to be a matter of chance (**local randomization**).
- Treatment effect is determined by comparing those just on either side of the cut-off.

Sharp RDD

- In sharp RDD, our forcing variable **X perfectly determines** which side of the cut-off you are on (treatment or control).
- For example, being over or under the age of 21 (in the US) determines whether or not you are eligible to legally buy alcohol.



Key Assumption

- **Continuity of average potential outcomes**: basically, units on one side of the threshold need to be essentially the same as units on the other side. Average potential outcomes should be continuous on both sides of the cut-off.
- The continuity assumption allows us to do a tiny bit of extrapolation and estimate **LATE at the threshold**.
- However, this assumption can easily be violated: It could be that the potential outcomes are actually not continuous and there is some other variable driving differences at the cutoff point.
    - For example, you may be incentivized to report your income just below a threshold for government support - this sorting violates our assumption.

Estimating LATE

- **Choose a window** around the threshold c to create a “discontinuity sample.”
  - The narrower the better, but can you afford losing many observations?
- **Recode forcing variable** X to deviations from threshold (centered on 0).
- **Decide which model** is the most appropriate given the nature of the data: linear with a common slope, linear with different slopes, or non-linear.

How to choose a model specification?

- A trade-off between **bias** and **variance**
  - When you go non-linear you might reduce variance because you can pick up every sensitivity in the data, but estimates will be biased due to following “noise.”
- Standard practice: Try and compare **different specifications** to show robustness
  - Ideally looking for similar results across different models.
- Do **local regression** (such as LOWESS) to guide choice

Falsification Checks

1. **Sensitivity**: Are results sensitive to alternative specifications?
    a. Nonlinear relation ≠ discontinuity
    b. If units start curving up near lower threshold and down near upper, it might just be non-linearity vs. a discontinuity jump.
2. **Balance checks**: Does any covariate Zi jump at the threshold?
    a. Aiming for scenario where we are comparing individuals that are pretty much identical except for what side of the cut off they ended up on. Only want to see a jump in Y, no other variables.
3. Do jumps occur at **placebo thresholds** c∗?
    a. If yes, this could mean something else is going on that could challenge our research design.
4. **Sorting**: Do units sort around the threshold?
    a. Sometimes there is an incentive to end up above or below a threshold. An agent's behavior can invalidate the continuity assumption

Identification of the threshold causal effect

- Key assumption: **Continuity of average potential outcomes**

$$E(Y_i(d)|X_i = x) \text{ is continuous in } x \text{ around } X_i = c \text{ for } d = 0, 1.$$

- Causal estimand: **local ATE at the threshold**

$$\tau_{SRD} = E(Y_i(1) - Y_i(0)|X_i = c)$$

- Identification result: If the continuity assumption holds, τ_{SRD} is nonparametrically identified as

$$\tau_{SRD} = \lim_{x \downarrow c} E(Y_i|X_i = x) - \lim_{x \uparrow c} E(Y_i|X_i = x)$$

- **Intuition:**
  - D_i is wholly determined by X_i, so conditional ignorability is trivially satisfied given X_i: $Y_i(1), Y_i(0) \perp\!\!\!\perp D_i|X_i$
  - However conditioning on X_i in a usual way (e.g., via regression, matching) won't work: no common support!
  - The continuity assumption allows us to do a tiny bit of extrapolation and compensate for the lack of common support at the threshold

Questions

- How to calculate in R?

- How to read results from the assignment?

Panel Data, Diff-in-Diff, and Fixed Effects

Panel Data

Time series data allow us to observe the same subject or unit in different causal states at different points in time. Data with multiple units over time are called panel data.

True panels, a.k.a. longitudinal data, record the same individuals over time in waves.

Independently pooled cross-sections are cross-sections repeated at different time periods. The observations (often respondents) are not identical over time.

Estimating Effects with Panel Data

- Assume we have a data set with random assignment into treatment and control, with **one outcome measurement before treatment** and **one outcome measurement after treatment.**
- As usual, we have a problem of not knowing **counterfactuals**: Ex. the average post-period outcome for the treated group in the absence of the treatment.
- To get around this we could:
  - Compare **before and after treatment for the treatment group**
    - This assumes no change in average potential outcome over time.
  - Compare the **treatment and control groups after treatment**
    - This assumes the PO of control group is the same as the counterfactual PO for those being treated.

Difference-in-Differences

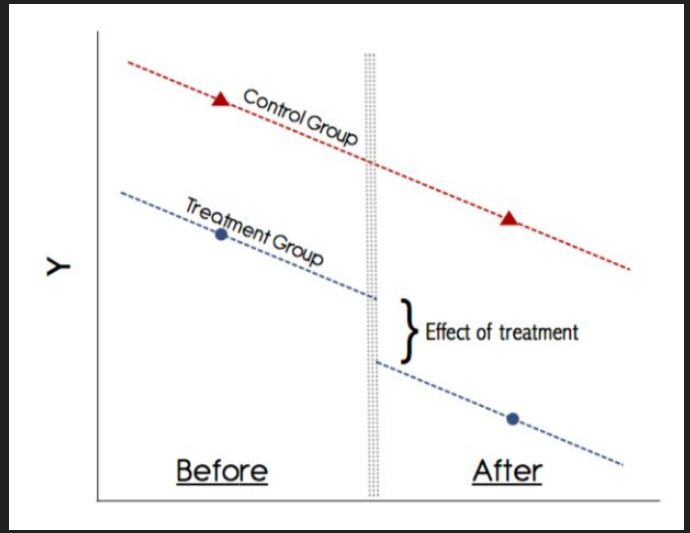
Or, compare the:

1. **Difference between the treatment and control group after treatment**

And subtract the

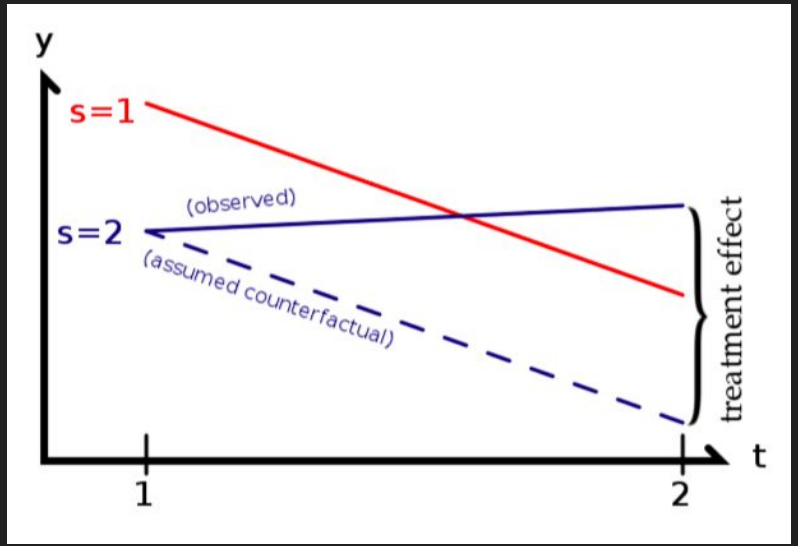
2. **Difference between the treatment and control group before treatment**

This approach uses overtime difference in control group as a counterfactual of what would have happened in the treatment group without the intervention.



Parallel Trends Assumption

Difference-in-differences estimation critically rests on the assumption that **observed overtime changes in the control group reflect, on average, unobserved changes in the treatment group** in the absence of treatment.



Fixed Effects

Panel estimators of causal effects exploit data with an additional time dimension to account for covariates (observed or unobserved) which are fixed:

- **Fixed within units:** Time-invariant traits of the units having an effect on the outcome. Something that affects the unit permanently (like racial identity).
- **Fixed within time**: Time-specific effects that affect all the units simultaneously (like changes in income due to national changes in the economy over time).

There are also **idiosyncratic factors** specific to a unit in a particular time. These are potential confounders.

Fixed Effects Estimation

Assume we have a common panel data setup that accounts for unit and time fixed effects:

- If we take the mean of the equation over time, the mean of the unit fixed effects is still just the original value, because it is **time-invariant.**
- If we subtract the averaged equation from the original equation, we have the **time-demeaned** equation where unit fixed effects drop out.
- If we estimate our model using this time-demeaned equation, we are left with the FE estimator (“**within estimator**”).
- Basically, we are left with a model where all the confounders that don’t vary over time just drop out.

Questions

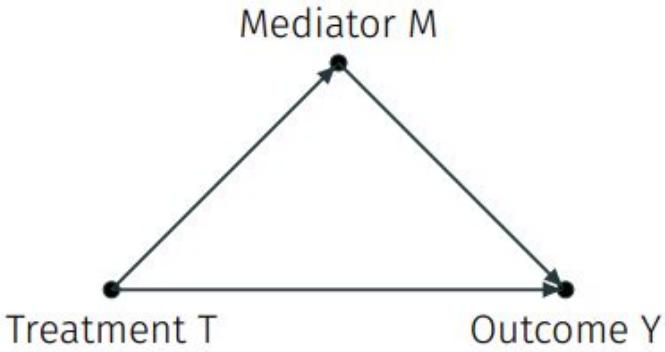
- How are DD and FE related?

- How is ignorability satisfied in DD and FE?

Causal Explanation: Mediation and Moderation

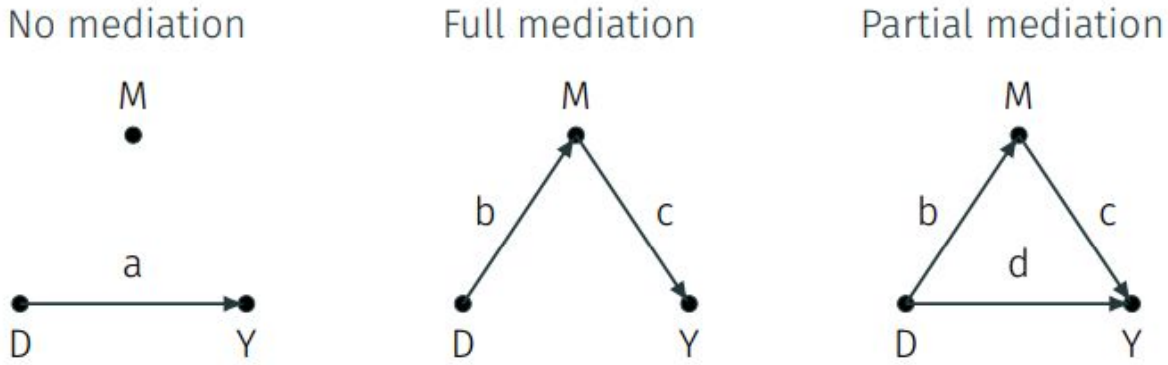
What is a *mediator*?

- Mediators are **intermediate variables** between the causal variable of interest and the outcome
- **Causal mediation analysis**:



- Break up causal pathways (the sum of which being the total effect) into **direct and indirect effects**

Traditional causal mediation analysis



Baron and Kenny's (1986) four steps of mediation analysis:

1. Regress Y on D to get $\hat{a}$ (total effect); if $a = 0$ → **no mediation**
2. Regress M on D to get $\hat{b}$; if $b = 0$ → **no mediation**
3. Regress Y on M adjusting for D to get $\hat{c}$; if $c = 0$ → **no mediation**
4. Regress Y on D adjusting for M to get $\hat{d}$; if $c \neq 0$ and $d = 0$ → **full mediation**; if $c \neq 0$ and $d \neq 0$ → **partial mediation**

Mediated effect: $a - d = b \cdot c$, direct effect: d

What is a *mediator*?

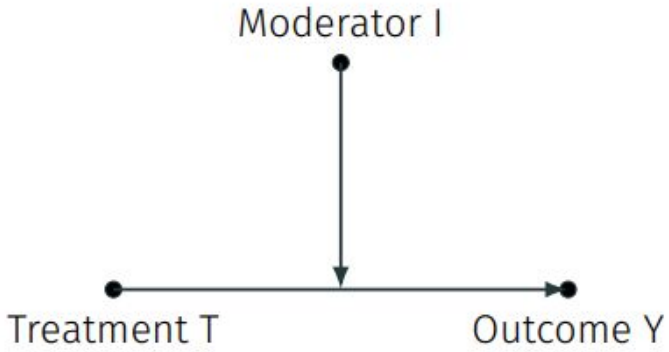
- Any treatment-outcome analysis (in the social sciences at least) can be converted into a causal mediation analysis when focusing on (often more micro) variables that mediate the relationship. For instance:

The effect of	**on**	**is mediated by**
alcohol consumption	intoxication	ethanol in the blood-stream
economy	incumbent vote share	retrospective voting
development aid	growth	infrastructure improve-ment

- Identifying causal mechanisms, i.e. the effects of mediators, can be very difficult even in the presence of experimental manipulation of treatment

What is a *moderator*?

- A moderator is a variable that affects the direction and/or strength of the relationship between the causal variable of interest and the outcome
- Such an effect is called **interaction effect**
- **Graphical representation**:



- For the record: This is not a "DAG" in the classical sense. Representing interaction effects in DAGs is not possible/necessary (see https://stats.stackexchange.com/questions/157775/representing-interaction-effects-in-directed-acyclic-graphs)

What is a *moderator*?

- **Conditional** (a.k.a. **interactive**, a.k.a. **moderated**) **relationships** are extremely common. For instance:

The effect of	**on**	**depends on**
alcohol consumption	intoxication	body weight
economy	incumbent vote share	clarity of responsibility
development aid	growth	corruption level of recipient government

Moderators

- In the social sciences, it's always worth thinking about conditional effects
- In the POF moderation can be conceptualized as a Conditional Average Treatment Effect (CATE)

$$\tau_{CATE}(x) = E(Y^1 - Y^0 | X = x) \tag{1}$$

- This answers the question: What is the ATE for a specific value x of the pre-treatment variable X?
- A moderating variable X shows that the ATE changes with different values of x.
- E.g. drinking two Beers will create a higher rate of intoxication for someone who weights 60 kg compared to someone who weights 90kg.

Estimation

- The most common way to estimate moderation is using Interaction Effects in Regression Models.
- Suppose we have a binary moderator $X_i \in (0,1)$ and a binary treatment $T_i \in (0,1)$ and a continuous outcome measure Y.
- We can estimate the CATE using a linear regression model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 X_i T_i + \epsilon_i \tag{2}$$

- Please note: To use this empirical estimate of the CATE the conditioning variable needs to be independent to the treatment. E.g. the moderater is not affected by the treatment, otherwise we condition on a post-treatment variable, which induces bias.

Your turn moderation: The effect of televised debates

	Model 1	Model 2
Intercept	5.0***	4.9***
	(12.6)	(11.6)
Treatment (Follow Debate)	0.25	0.1
	(0.23)	(0.56)
Non-partisans		0.1
		(0.5)
Non-partisans X Treatment		0.7***
		(4.02)
N	804	804

- What is the Conditional Average Treatment Effect (CATE) among Partisans and non-partisans?
- How can you interpret the CATE?
- Is the effect statistically significant?