Experimental Methods: Lecture 4

Raymond Duch

May 30, 2018

Director CESS Nuffield/Santiago

Road Map

- Mediation
- Heterogeneity

Introduction to Mediation

- Mediation is concerned with the study of intervening or mediating variables that transmit the influence of an experimental intervention
- Begin with a treatment Z_i on outcome Y_i
- Does X_i induce a change in mediating variable M_i?
- Does the induced change in M_i lead to a change in Y_i?
- This is not as straightforward to determine as it may first appear

Example: Local Government Representation in India

- Bhavnani (2009) studies local government representation in India
- Before 2002, a randomly selected portion of local council seats were reserved for women
- In 2002 the reservations were lifted, but constituencies where women held reserved seats in 1997 were still more likely to elect women representatives in 2002. Why?
    1. Reservations create/select a cohort of female incumbents whose experience in office makes them more appealing to voters
    2. Reservations give voters an opportunity to change their views about women and learn that women make capable representatives
    3. Female representatives increase voter participation, and a surge of new voters might continue to improve the chances of electing a woman after reservations expire
- Concludes evidence mostly for first hypothesis based on selection
- But the key point is that each hypothesis posits a different mediator that is influenced by reservations

Regression-based approaches to mediation

- Regression-based analyses typically rely on some form of three-equation system like the following:

$$M_i = \alpha_1 + aZ_i + e_{1i} \tag{1}$$

$$Y_i = \alpha_2 + bZ_i + e_{2i} \tag{2}$$

$$Y_i = \alpha_3 + dZ_i + bM_i + e_{3i} \tag{3}$$

- where e_{xi} are unobserved disturbances representing cumulative effects of missing variables, Z_i randomly assigned, and M_i a pre-treatment covariate
- Substitute equation (1) in (3) and compare to (2):

$$Y_i = \alpha_3 + (d + ab)Z_i + (\alpha_1 + e_{1i})M_i + e_{3i}$$

- Total effect of Z_i on Y_i is $c = d + ab$
- consists of direct effect d and mediated effect ab
- But what if coefficients vary from observation to observation (i.e. non-constant treatment effect)?

$$E[a_i b_i] = E[a_i]E[b_i] + cov(a_i, b_i)$$

- So we cannot just estimate $E[a_i]$ and $E[b_i]$ separately and multiply them together to get this product

Regression-based approaches to mediation

- What if we assumed constant treatment effects?
- Random assignment of Z_i means (1) and (2) can be estimated without bias, i.e. $E \perp e_{1i}, e_{2i}$
- But what about $Y_i = \alpha_3 + dZ_i + bM_i + e_{3i}$?
- M_i is not randomly assigned!
- M_i could be systematically related to unmeasured causes of Y_i and correlated with e_{3i}
- Can you think of an unmeasured cause of Y_i that is correlated with e_{3i} in the Bhavnani case?

Regression-based approaches to mediation

- In Bhavnani, Z_i is previous reservation, Y_i the election of female representative in 2002, M_i the number of women candidates running for office in 2002 (H1)
- Factors other than randomly assigned reservations cause women candidates to run for office
- One idea: What if some districts were more egalitarian than others?
- Female candidates more likely to run in districts that are more egalitarian
- This unmeasured disturbance would be correlated with e_{3i}, the unmeasured factors that affect the election of a woman in 2002

The Direction of the Bias

- Assume $M_i = \alpha_1 + aZ_i + e_{1i}$ and $Y_i = \alpha_3 + dZ_i + bM_i + e_{3i}$
- as N goes to infinity (Gerber and Green 2009),

$$\hat{b}_{n\to\infty} = b + \frac{cov(e_{1i}, e_{3i})}{var(e_{1i})}$$

$$\hat{d}_{n\to\infty} = b + a\frac{cov(e_{1i}, e_{3i})}{var(e_{1i})}$$

- $cov(e_{1i}) > 0$ is likely: Even controlling for Z_i, if women were more likely to run for office in 2002 in district i, they were more likely to win there (because of egalitarianism)
- Bias thus inflates the estimate of b and deflates the estimate of d
    - Exactly the bias that researchers look for find for mediation
    - Could add covariates such that $cov(e_{1i} = 0)$
- To reinforce these points, we will do a quick detour to a Monte Carlo experiment that illustrates these points more clearly

Mediation and Potential Outcomes

- Define $M_i(z)$ as the potential value of M_i when $Z_i = z$
- Define $Y_i(m, z)$ as potential outcome when $M_i = m$ and $Z_i = z$
- $Y_i(M_i(1), 1)$ thus expresses potential outcome when $Z_i = 1$ and M_i takes on potential outcome that occurs when $Z_i = 1$
- Total effect of Z_i on Y_i is $Y_i(M_i(1), 1) - Y_i(M_i(0), 0)$
- What is the direct effect of Z_i on Y_i controlling for M_i?
    - There is more than one definition
    - $Y_i(M_i(0), 1) - Y_i(M_i(0), 0)$ is direct effect of Z_i on Y_i holding m constant at $M_i(0)$
    - $Y_i(M_i(1), 1) - Y_i(M_i(1), 0)$ is direct effect of Z_i on Y_i holding m constant at $M_i(1)$
    - $Y_i(M_i(0), 1)$ and $Y_i(M_i(1), 0)$ are complex potential outcomes, so named because they are purely imaginary and never occur empirically

Mediation and Potential Outcomes

- What is the direct effect of Z_i on Y_i through M_i?
  - This is the effect on Y_i of changing from $M_i(0)$ to $M_i(1)$ while holding Z_i constant
  - So again, depending on Z_i, we get two definitions of the indirect effect
  - $Y_i(M_i(1), 1) - Y_i(M_i(0), 1) | Z_i = 1$ and $Y_i(M_i(1), 0) - Y_i(M_i(0), 0) | Z_i = 0$
  - Again $Y_i(M_i(0), 1)$ and $Y_i(M_i(1), 0)$ are the earlier complex potential outcomes
- Each of these four equations involve a term that is fundamentally unobservable
- True even if we assume that both indirect effects are equal
- There is thus a fundamental limitation on what we can learn from an experiment while manipulating only Z_i without making further assumptions

Ruling Out Mediators

- What if the sharp null hypothesis $M_i(0) = M_i(1)$ is true?
- $Y_i(M_i(1), 1) - Y_i(M_i(0), 1)|Z_i = 1$ and $Y_i(M_i(1), 0) - Y_i(M_i(0), 0)|Z_i = 0$
- Then both indirect effects equal 0. Experiments may indicate when mediation does not occur, but sometimes difficult to do in practice:
    - Need tight estimate around 0
    - Need sharp null to be true, not just $ATE = 0$
- Although sharp null cannot be proven, we can cite evidence suggesting whether this conjecture is a reasonable approximation
- We thus learn something useful about mediation when discovering a lack of causal relationship between Z_i and proposed mediator
- Conversely, if Z_i and M_i have a strong relationship, we cannot rule out M_i as a possible mediator

Manipulating the Mediators

- A fundamental problem is that M_i is not independently manipulated via random intervention
- Could we manipulate M_i as well to build the case for mediation? $\rightarrow$ In principle, yes, but difficult in practical situations
- Example Y_i is scurvy, Z_i is lemon, M_i is vitamin C
  - We want indirect effect $Y_i(M_i(1), 0) - Y_i(M_i(0), 0)$
  - $M_i(1)$ is vitamin C level of lemon, we feed pills without lemons
  - Still not perfect: Vitamin C in lemons consumed differently from pills, pills might have other effects on Y_i
- Manipulations of M_i are therefore instructive, but ability to provide empirical estimates inevitably requires additional assumptions
- In the Bhavnani example, possible M_i are number of female incumbents, voters' sense of whether it is appropriate or desirable to have women representatives, and turnout rate in local elections

Implicit Mediation

- Consider a treatment Z_i that contains multiple elements in it
- Rather than manipulating M_i, change the treatment to isolate particular elements of Z_i (i.e. Z^1, Z^2, Z^3) whose attributes affect M_i along the way
- Focus is not on demonstrating how a Z_i-induced change in M_i changes Y_i, but on the effect of different isolated treatments on Y_i
- In particular, no attempt to estimate the effects of observed changes in M_i at all

Example: Conditional Cash Transfers

- Interest in conditional cash transfers on poor to keep children in school and attend health clinics
- Field experiments find improved educational outcomes for children in developing countries from these transfers (Baird, McIntosh, and Ozler 2009)
- What could the causal mechanism be?
  1. Cash subsidies allow greater investment in childrens welfare
  2. Imposed conditions improve childrens welfare
- Baird, McIntosh and Ozler (2009) designed experiment with three groups
  - Control group with no subsidy, instructions, or conditions
  - One treatment group gets cash without conditions
  - Another treatment group gets cash with conditions
  - Finding: Null hypothesis of no difference between treatment groups cannot be rejected

Benefits of Implicit Mediation

1. Simple: Never strays from the unbiased statistical framework of comparing randomly assigned groups
2. By adding and subtracting elements from treatment, this approach lends itself to exploration and discovery of new treatments
   - Facilitates the process of testing basic propositions about what works by providing clues about the active elements that cause a treatment to work particularly well
3. Can gauge treatment effects on a wide array of outcome variables
   - Allows manipulation checks for establishing the empirical relationship between intended and actual treatments
   - Example: Does discussion in the classroom improve performance? Check if treatment increases discussion

Voter Turnout Example

- Gerber, Green, and Larimer (2008) interested in the effect of communication on turnout
- U.S. has voters files, anyone know what they are?
- 180,000 Michigan households in experiment
- 100,000 in control group (no postcards), other groups 20,000 each
- **Civic duty**: "It's your civic duty to vote"
- **Hawthorne**: "It's your civic duty to vote, we're doing a study and will check public records"
- **Self**: "You should vote, here's your recent voting record"
- **Neighbors**: "You should vote, here's your neighbors' voting records and your own"

Results

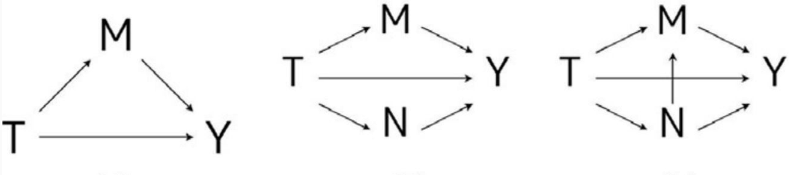
	Control	Civic	Hawthorne	Self	Neighbors
Pct Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N	191,243	38,218	38,204	38,218	38,201

Anyone here know how Gerber followed up on this study?

Summary

- We are often curious about the mechanisms by which an experimental treatment transmits its influence
- Adding mediators as right-hand variables to determine this is a flawed strategy that generally provides bias in favor of mediation
- Main issue here is that the mediator is not experimentally manipulated
- In theory we could manipulate mediators experimentally, but this is difficult for two reasons
  1. We never observe complex potential outcomes
  2. Manipulation of mediators directly is often impractical
- However, two lines of inquiry seem promising:
  1. We can rule out mediators easier than we can find them
  2. We can implicitly manipulate mediators

Causal Mechanisms



(a) (b) (c)

Potential Outcomes

Total unit effect:

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

Indirect effect:

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$$

Direct effect:

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

Chain Fallacy

Population Proportion	$M_i(1)$	$M_i(0)$	$Y_i(t,1)$	$Y_i(t,0)$	Treatment Effect on Mediator $M_i(1) - M_i(0)$	Mediator Effect on Outcome $Y_i(t,1) - Y_i(t,0)$	Causal Mediation Effect $Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
0.3	1	0	0	1	1	-1	-1
0.3	0	0	1	0	0	1	0
0.1	0	1	0	1	-1	-1	1
0.3	1	1	1	0	0	1	0
Average	0.6	0.4	0.6	0.4	0.2	0.2	-0.2

General Estimator Algorithm

- Model outcome and mediator
  - Outcome model: $p(Y_i | T_i, M_i, X_i)$
  - Mediator model: $p(M_i | T_i, X_i)$
- These models can be of any form (linear or nonlinear, semi- or nonparametric, with or without interactions)
- Predict mediator for both treatment values $M_i(1), M_i(0)$
- Predict outcome by first setting $T_i = 1$ and $M_i = M_i(0)$ and then $T_i = 1$ and $M_i = M_i(1)$
- Compute the average difference between two outcomes to obtain a consistent estimate of ACME
- Monte-Carlo or bootstrapping to estimate uncertainty

Example: Continuous Mediator and Binary Outcome

- Estimate the following models:

$$M_i = \alpha_2 + \beta_2 T_i + X_i + e_{2i}$$
$$Pr(Y_i = 1) = \Phi(\alpha_3 + b_3 T_i + \gamma M_i + X_i + e_{3i})$$

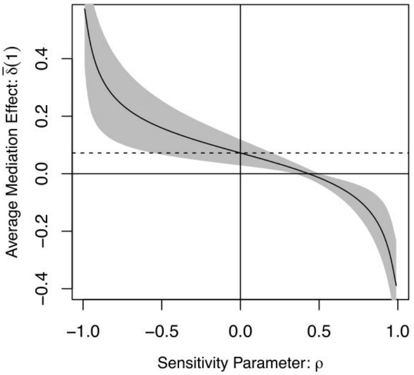
- Predict M_i for $T_i = 1$ and $T_i = 0$. This gives you $\hat{M}_i(1)$ and $\hat{M}_i(0)$
- Predict Y_i with $T_i = 1$ and $\hat{M}_i(0)$ and vice versa
- Take average of these two predictions

Media Cues and Immigration Attitudes

Brader et al. experiment:

- Subjects read a mock news story about immigration
- Treatment: immigrant in story is a Hispanic, and the news story emphasized the economic costs of immigration
- They measured a range of different attitudinal and behavioral outcome variables:
  - Opinions about increasing or decrease immigration
  - Contact legislator about the issue
  - Send anti-immigration message to legislator
- They want to test whether the treatment increases anxiety, leading to greater opposition to immigration

Sensitivity: Interpreting ρ



- ACME > 0 as long as the error correlation is less than 0.39 (0.30 with 95% CI)

Sensitivity: R Code

1. Fit models for the mediator and outcome variable and store these models.

```
> m <- lm(Mediator ~ Treat + X)
> y <- lm(Y ~ Treat + Mediator + X)
```

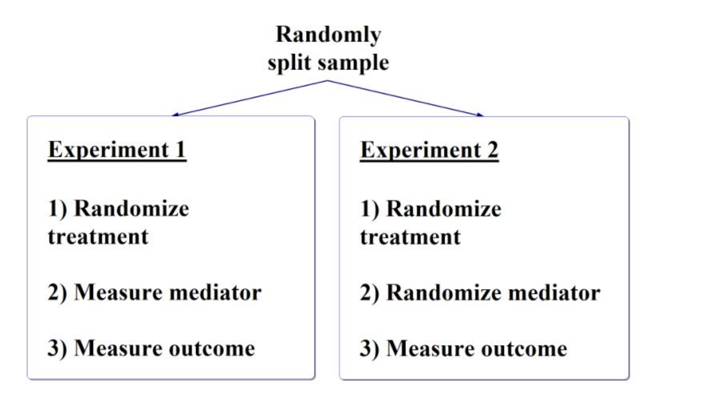
2. Mediation analysis: Feed model objects into the `mediate()` function. Call a summary of results.

```
> m.out<-mediate(m, y, treat = "Treat",
                 mediator = "Mediator")
> summary(m.out)
```

3. Sensitivity analysis: Feed the output into the `medsens()` function. Summarize and plot.

```
> s.out <- medsens(m.out)
> summary(s.out)
> plot(s.out, "rho")
> plot(s.out, "R2")
```

Parallel Design



Example from Behavioral Neuroscience

- Why study brain? Social scientists' search for causal mechanisms underlying human behavior $\rightarrow$ Psychologists, economists, and even political scientists
- Question: What mechanism links low offers in an ultimatum game with "irrational" rejections?
  - A brain region known to be related to fairness becomes more active when unfair offer received (single experiment design)
- Design solution: manipulate mechanisms with TMS
  - Knoch et al. use TMS to manipulate turn off one of these regions, and then observes choices (parallel design)legislator

Encouragement Design

- Randomly *encourage* subjects to take particular values of the mediator M_i
- Standard *instrumental variable* assumptions (Angrist et al.)
- Use a 2×3 factorial design:
    - Randomly assign T_i
    - Also randomly decide whether to positively encourage, negatively encourage, or do nothing
    - Measure mediator and outcome
- Informative inference about the "complier" ACME
- Reduces to the parallel design if encouragement is perfect
- Application to the immigration experiment: Use autobiographical writing tasks to encourage anxiety

Cross-over Design

- Recall *ACME* can be identified if we observe $Y_i(t_0; M_i(t))$
- Get $M_i(t)$, then switch T_i to t_0 while holding $M_i = M_i(t)$
- Crossover design:
    - Round 1: Conduct a standard experiment
    - Round 2: Change the treatment to the opposite status but fix the mediator to the value observed in the first round
- Very powerful identifies mediation effects for each subject
- Must assume *no carryover effect*: Round 1 must not affect Round 2
- Can be made plausible by design

Example from Labor Economics

Bertrand & Mullainathan (2004, AER)

- Treatment: Black vs. White names on CVs
- Mediator: Perceived qualifications of applicants
- Outcome: Callback from employers
- Quantity of interest: Direct effects of (perceived) race
- Would Jamal get a callback if his name were Greg but his qualifications stayed the same?
- Round 1: Send Jamals actual CV and record the outcome
- Round 2: Send his CV as Greg and record the outcome
- Assumptions are plausible

Cross-over Encouragement Design

- Cross-over encouragement design:
  - Round 1: Conduct a standard experiment
  - Round 2: Same as crossover, except encourage subjects to take the mediator values
- Example: Hainmueller & Hiscox (2010, APSR)
  - Treatment: Framing immigrants as low- or high-skilled
  - Possible mechanism: Low income subjects may expect higher competition from low skill immigrants
  - Manipulate expectation using a news story
  - Round 1: Original experiment but measure expectation
  - Round 2: Flip treatment, but encourage expectation in the same direction as Round 1

Heterogeneity: Kosuke and Ratkovic 2013

- Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. **The Annals of Applied Statistics, 7(1):443–470, 2013**
- FindIt package (CRAN: https://cran.r-project.org/web/packages/FindIt/index.html)

Heterogeneity: Example

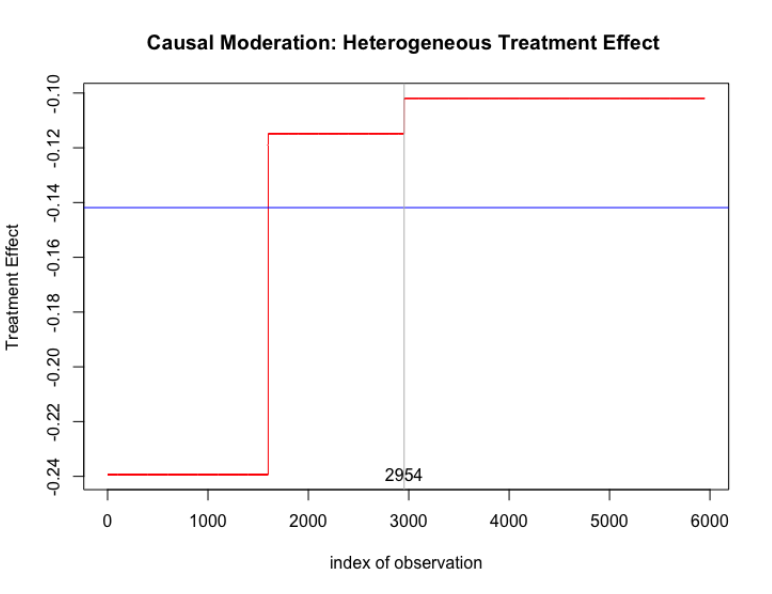
- Data: Lab and online experiments
- Treatment: $>$ mean on RET task
- Outcome: Report rate of RET earnings
- Covariates: Age, Gender
- Potential heterogeneity: mode of experiment (online vs lab), subject pool (student vs non-student).

Basic model: mode and pool interactions

Table 1: Heterogeneous treatment coefficients and interactions

Variable	Coefficient
Student	-0.115
Student $\times$ Online	-0.059
Treatment	-0.141
Treatment $\times$ Student	-0.015
Treatment $\times$ Online	0.121
Treatment $\times$ Student $\times$ Online	0.007
Intercept	0.591
ATE	-0.142

Basic model: heterogeneous effects



Het. effects including covariate interactions

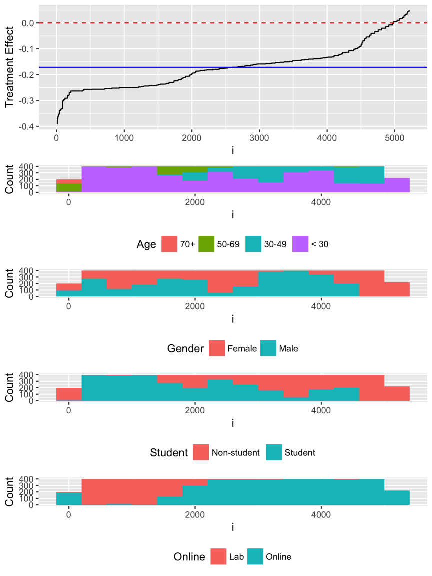
```
Call:
 Treatment Model: report.rate ~ treat.het
 Main Model : ~student + online + Age + Gender
 Interaction Covariates: ~student + online + Age + Gender
 Treatment type: [1] "single"
 Outcome type: [1] "continuous"

ATE:
[1] -0.1715897

Coefficients:
          Intercept             student                 Age              Gender       student:online         student:Age
              0.584              -0.043             0.00588              0.0311              -0.0781             0.00124
     student:Gender       online:Gender          Age:Gender               Age.2                treat       treat:student
              0.225              0.0567             0.00288           -3.37e-05               -0.135             -0.0429
       treat:online        treat:Gender   treat:student:Age treat:student:Gender treat:online:Gender    treat:Age:Gender
             0.0893              0.0316              0.0076              -0.202              -0.0317             -0.0044
        treat:Age.2
          -6.24e-05

---------
Model Fit Statistics:
GCV:
  Null:  0.205  Model:  986.984
Percent Misclassified:
   Null:  0   Model:  0.28
   Percent Improvement, vs. NULL:  -Inf %
Percent Outside Margin:
  50.134 %, n = 0
```

Het. effects including covariate interactions



References

[1] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.