Statistical Modeling and Causal Inference with R

Week 11: Causal Mediation

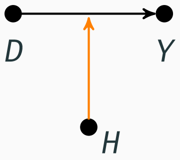
Manuel Bosancianu

Max Schaub

November 23, 2020

Hertie School of Governance

Recap



Moderation: the magnitude of the ATE *varies* systematically for sub-groups defined by H.

Example: effect of broadband availability on support for right-wing parties might be stronger for younger respondents (Schaub & Morisi, 2020).

In policy work, constitutes a check on whether specific groups are more exposed to benefits (or benefits) of intervention.

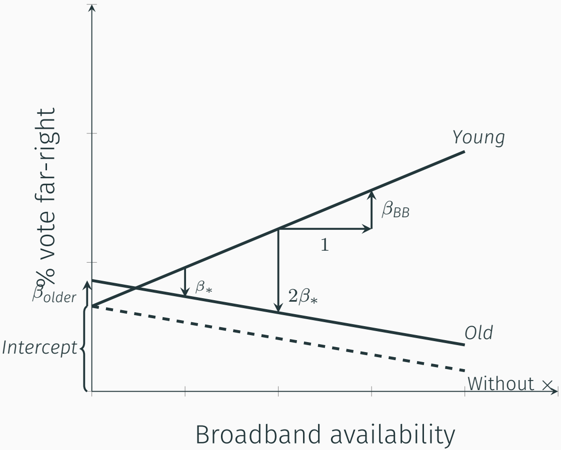
Interaction terms

Estimation: done through multiplicative interaction terms (Brambor, Clark, & Golder, 2006).

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 H_i + \beta_3 D_i \times H_i + \epsilon_i \tag{1}$$

- β_0: intercept;
- β_1: effect of treatment on outcome, *when* $H_i = 0$
- β_2: effect of mediator on outcome, *when* $D_i = 0$ (control group)
- β_3: how effect of treatment changes for each unit change in H_i (different subgroup)

Graphical presentation



Example with broadband Internet (graph adapted from Brambor et al., 2006). β_* means $\beta_{older*BB}$.

Today's session

Identifying *how* a treatment effect is transmitted:

- ✓ mediation in the classical LSEM tradition (Baron & Kenny, 1986)
- ✓ issues with the classical tradition (Bullock & Ha, 2011)
- ✓ the *causal mediation paradigm* (Imai, Keele, Tingley, & Yamamoto, 2011)
- ✓ more practice with DAGs and the POF!
- ✓ practical example with data from a job-seeking intervention

Today's example: JOBS II

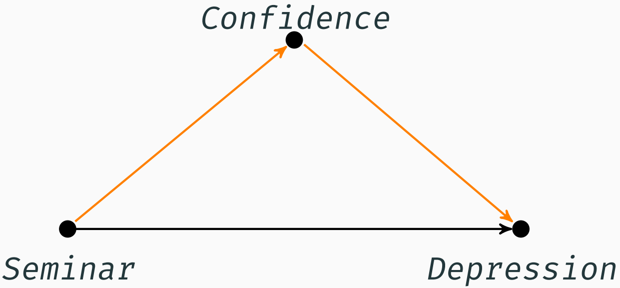
The JOBS II was a field experiment in Southeast Michigan in early 1990s.

Treatment was a job-search skills seminar administered to unemployed persons.
Control was a booklet with search tips.

Expected effects:

- ✓ increased employment
- ✓ increased self-confidence $\Rightarrow$ decreased depression

Today's example: JOBS II



How much of the *ATE* is *transmitted* through improved self-confidence?

Classical mediation framework

The Baron–Kenny model I

Very influential across disciplines (to this day).

The **moderator–mediator** variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.

RM Baron, DA Kenny - Journal of personality and social …, 1986 - psycnet.apa.org

In this article, we attempt to distinguish between the properties of moderator and mediator variables at a number of levels. First, we seek to make theorists and researchers aware of the importance of not using the terms moderator and mediator interchangeably by carefully …

☆ 99 Zitiert von: 81113 Ähnliche Artikel Alle 44 Versionen

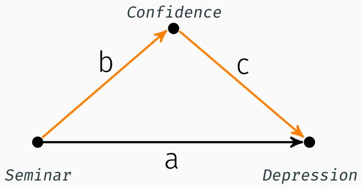
The **moderator–mediator variable** distinction in social psychological research: Conceptual, strategic, and statistical considerations.

RM Baron, DA Kenny - Journal of personality and social …, 1986 - psycnet.apa.org

In this article, we attempt to distinguish between the properties of **moderator** and **mediator** **variables** at a number of levels. First, we seek to make theorists and researchers aware of the importance of not using the terms **moderator** and **mediator** interchangeably by carefully …

☆ 99 Cited by 96559 Related articles All 49 versions Web of Science: 42452

The Baron–Kenny model II



We also have a set of *pre-treatment covariates*, X, which aren't included in the plot.

Seminar ⇒ *Depression*: direct path (a is the direct effect).

Seminar ⇒ *Confidence* ⇒ *Depression*: indirect path, AKA "front-door path" (in POF).

Indirect effect is a function of b and c.

3 regressions

$$\mathit{Depression}_i = \alpha_1 + \beta_1 \overbrace{\mathit{Seminar}_i}^{\mathit{treatment}} + \zeta_1 X_i + \epsilon_{i1} \tag{2}$$

$$\mathit{Confidence}_i = \alpha_2 + \beta_2 \mathit{Seminar}_i + \zeta_2 X_i + \epsilon_{i2} \tag{3}$$

$$\mathit{Depression}_i = \alpha_3 + \gamma \mathit{Seminar}_i + \beta_3 \underbrace{\mathit{Confidence}_i}_{\mathit{mediator}} + \zeta_3 X_i + \epsilon_{i3} \tag{4}$$

Total effect of *Seminar* on *Depression* is β_1 (Equation 2).

To see how it's decomposed, substitute Equation 3 into Equation 4.

Effect decomposition

$$
\begin{aligned}
Depression_i &= \alpha_3 + \gamma Seminar_i + \beta_3(\alpha_2 + \beta_2 Seminar_i + \zeta_2 X_i + \epsilon_{i2}) + \zeta_3 X_i + \epsilon_{i3} && (5)\\
&= \alpha_3 + \gamma Seminar_i + \beta_3\alpha_2 + \beta_3\beta_2 Seminar_i + \beta_3\zeta_2 X_i + \beta_3\epsilon_{i2} + \zeta_3 X_i + \epsilon_{i3} && (6)\\
&= \alpha_3 + Seminar_i(\gamma + \beta_3\beta_2) + X_i(\beta_3\zeta_2 + \zeta_3) + \beta_3(\alpha_2 + \epsilon_{i2}) + \epsilon_{i3} && (7)
\end{aligned}
$$

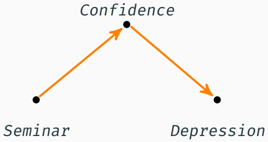
β_1 (total effect) is now a composite term: $\gamma + \beta_2\beta_3$.

γ is the direct effect of *Seminar* on *Depression*.

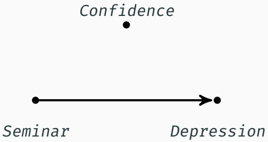
$\beta_2\beta_3$ is the indirect (or mediated) effect of *Seminar* on *Depression*, transmitted through *Confidence*.

Configurations

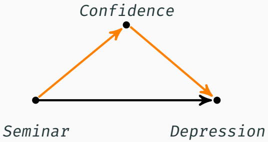
If $\gamma \approx 0$ and $\beta_2\beta_3 \neq 0$: full mediation.



If $\gamma \neq 0$ and $\beta_2\beta_3 \approx 0$: no mediation.



If $\gamma \neq 0$ and $\beta_2\beta_3 \neq 0$: partial mediation.



JOBS II data

Outcome: continuous, score on Hopkins Symptom Checklist, measured 6 months after treatment.

Mediator: job search self-efficacy, measured 2 months after treatment.

Covariates: age, education, gender, ethnic minority, depression measured during treatment, economic hardship, marital status, occupation, income.

$$Depression_i = \alpha_1 + \beta_1 \overbrace{Seminar_i}^{treatment} + \zeta_1 X_i + \epsilon_{i1} \tag{8}$$

$$Confidence_i = \alpha_2 + \beta_2 Seminar_i + \zeta_2 X_i + \epsilon_{i2} \tag{9}$$

$$Depression_i = \alpha_3 + \gamma Seminar_i + \beta_3 \underbrace{Confidence_i}_{mediator} + \zeta_3 X_i + \epsilon_{i3} \tag{10}$$

Results

Results from 3 specifications

	DV: Depression	DV: Confidence	DV: Depression
(Intercept)	0.895^{***}	3.870^{***}	1.499^{***}
	(0.133)	(0.159)	(0.158)
Seminar	-0.047	0.101^{*}	-0.032
	(0.035)	(0.042)	(0.035)
Confidence			-0.156^{***}
			(0.023)
R^2	0.244	0.116	0.271
Adj. R^2	0.230	0.100	0.256
Num. obs.	1285	1285	1285

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimates from pre-treatment covariates have been excluded from the table.

At this point, the manual calculations begin...

Computing effects

- Direct effect: -0.032
- Indirect effect: $\beta_2 \times \beta_3$ = 0.101×-0.156 = -0.016
- Total effect: *direct* + *indirect* = $-0.032 + (-0.016)$ = -0.047 (rounding)

Aside from point estimates, we also need estimates of uncertainty:

- *Seminar* $\Rightarrow$ *Confidence* = *b*
- *Confidence* $\Rightarrow$ *Depression* = *c*

Computing uncertainty

$$SE_{indirect} = \sqrt{c^2\sigma_b^2 + b^2\sigma_c^2 + \sigma_b^2\sigma_c^2} \tag{11}$$

	Direct	Indirect
β	-0.032	-0.016^{*}
SE	(0.035)	(0.007)

A case of pure mediation: a negative indirect effect of seminar attendance on depression (via confidence).

Challenges in classical framework

Cracks in the foundation

The approach established itself as *de facto* standard in social and political psychology, and connected fields.

Problems:

- ✓ bias in OLS estimates
- ✓ inability to cope with more complex causal structures
- ✓ mostly obscures identification assumptions
- ✓ does not extend to nonlinear models

Bias in estimates

$$Depression_i = \alpha_1 + \beta_1 Seminar_i + \zeta_1 X_i + \epsilon_{i1} \tag{12}$$
$$Confidence_i = \alpha_2 + \beta_2 Seminar_i + \zeta_2 X_i + \epsilon_{i2} \tag{13}$$
$$Depression_i = \alpha_3 + \gamma Seminar_i + \beta_3 Confidence_i + \zeta_3 X_i + \epsilon_{i3} \tag{14}$$

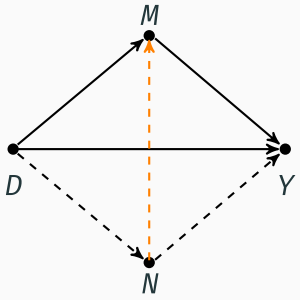
In Equation 14, β_3 is potentially estimated with bias:

$$\hat{\beta}_3 = \beta_3 + \frac{cov(\epsilon_{i2}, \epsilon_{i3})}{var(\epsilon_{i2})} \tag{15}$$

The only way $\hat{\beta}_3 = \beta_3$ and $\hat{\gamma} = \gamma$ is if $cov(\epsilon_{i2}, \epsilon_{i3}) = 0$.

This only holds if treatment and mediator are randomly assigned—not the typical design setup!

Complex causal structures + obscuring identification



N is another causal pathway from treatment (D) to outcome (Y).

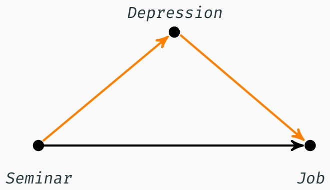
Even with D and M randomized, $cov(\epsilon_{i2}, \epsilon_{i3}) \neq 0$ if N is unobserved.

"Multiple-mediator" problem: this can still bias estimates (and poses measurement challenges).

Even when measuring all, "one-at-a-time" estimation of indirect effects is very likely inducing bias: omitted variables.

Focus is on model fit and required covariates, rather than highlighting assumptions needed for identification.

Problem with nonlinearity



Baron and Kenny (1986) don't say what kind of estimation to use, but everyone uses OLS in practice.

More difficult to use it for instances where moderators and outcomes are categorical.

Experimental approach to causal mediation

POF for causal mediation

For consistency, following the notation in Imai et al. (2011), though it differs from that of Angrist and Pischke (2015).

$Y_i(t)$: potential outcome under treatment status t, with $t \in \{0; 1\}$.

ITE is still $Y_i(1) - Y_i(0)$ (and still unobservable).

If treatment is randomized ($\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i$), ATE can be identified: $E[Y_i(1)] - E[Y_i(0)]$.

Total treatment effect

As before, we can write out the direct and the indirect effect.

- $M_i(t)$: mediator potential value (assume $\in \{0; 1\}$), if treatment status is t
- $Y_i(t, m)$: outcome potential value for treatment t and mediator m

$Depression_i(1, 1)$: depression level if attended seminar and is confident (versus not confident) in own job-search skills.

$$\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \tag{16}$$

Decomposition: causal mediation effect (ACME)

$$\delta_i(t) = Y_i(t, M_i(1)) - Y_i(t, M_i(0)),\ \textit{for each}\ t \in \{0;1\} \tag{17}$$

$\delta_i(t)$ (indirect effect): change in outcome if mediator changes from value under control ($M_i(0)$) to value under treatment ($M_i(1)$), keeping t constant.

If treatment has no impact on the mediator, then $M_i(0) = M_i(1)$, and $\delta_i(t)$ = 0.

- ✓ $Depression_i(1, Confidence_i(1))$: depression level for seminar participant
- ✓ $Depression_i(1, Confidence_i(0))$: depression level for seminar participant, with confidence level as if they had not participated
- ✓ $Depression_i(0, Confidence_i(0))$: depression level for non-participant
- ✓ $Depression_i(0, Confidence_i(1))$: depression level for non-participant, with confidence level as if they had participated

Decomposition: direct effect (ADE)

$$\zeta_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t)), \ \textit{for each } t \in \{0; 1\} \tag{18}$$

The total effect (τ) is a decomposition of δ_i and ζ_i:

$$\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0)) = \frac{1}{2}\sum_{t=0}^{1}(\delta_i(t) + \zeta_i(t)) \tag{19}$$

If direct and indirect effects don't vary based on treatment status ("no-interaction" assumption):

$$\tau_i = \delta_i + \zeta_i \tag{20}$$

Key assumption: sequential ignorability I

For standard estimation of ATE, only needed two: (1) non-interference (SUTVA), and (2) excludability.

For causal mediation, one more: **sequential ignorability**.

$$[Y_i(t', m), M_i(t)] \perp\!\!\!\perp T_i | X_i = x \tag{21}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \tag{22}$$

Equation 21: conditional on pre-treatment covariates, treatment is independent of potential outcomes and potential mediator values.

For JOBS II this is met, since treatment is randomly assigned.

Key assumption: sequential ignorability II

$$[Y_i(t', m), M_i(t)] \perp\!\!\!\perp T_i | X_i = x \tag{23}$$
$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \tag{24}$$

Equation 22: conditional on observed treatment *and* pre-treatment covariates, mediator is independent of potential outcomes.

For JOBS II this is **not met**, since people's level of confidence was not experimentally varied.

Very strong assumption, which usually doesn't hold in standard randomized experiments.

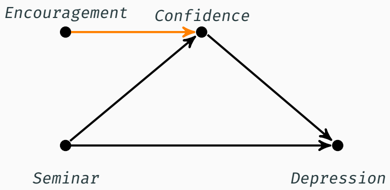
Standard experimental designs

JOBS II follows this set of steps:

1. randomly assign treatment
2. measure mediator after assigning treatment
3. measure outcome

1st part of sequential ignorability is met, but not the 2nd part: can't be sure mediator is independent of potential outcomes.

Encouragement experimental designs



Confounders not depicted here, but they might be at play as well.

Only *pre-treatment* confounders can be accommodated.

Stages:

1. randomly assign treatment
2. within treatment and control groups, randomize encouragement to low/high value of mediator
3. measure mediator and outcome values

Encouragement is like an instrument, meaning we identify the *complier* ACME.

Confidence could be built through an information treatment or a vignette.

Challenges in experimental approach

Though powerful, experimental designs for identifying ACME run into their own challenges.

- ✓ the instrument has to operate only on the mediator of interest
- ✓ estimated ACME is only for *compliers*, and complier group might be different for treatment and mediator
- ✓ if causal effects (δ_i and ζ_i) are not identical for all individuals, our ACME estimate is biased
- ✓ "weak instruments" persists as a problem

Summary

Problems with classical (LSEM) model

Baron and Kenny (1986) framework continues to be widely used.

Advantages: simplicity, speed, relative robustness (due to OLS estimation).

Indirect effect is a multiplication of 2 OLS coefficients (the SE can be computed as well).

Disadvantages: causal assumptions obscured, OLS estimates potentially biased*, difficulty with nonlinear models, challenge posed by multiple mediators.

Causal mediation framework

First and foremost, a way of emphasizing needed assumptions for causal interpretation of estimates (both for experimental and observational designs).

Key assumption: *sequential ignorability*.

Places high burden on experimental designs (standard ones don't meet the 2nd part of assumption).

Encouragement designs could be used fruitfully, though again care is needed in implementation.

Benefits

The proposed estimation method is non-parametric: works for a variety of statistical models (Imai, Keele, & Tingley, 2010) .

Comes with a toolbox of algorithms for sensitivity analysis (in case of assumption violations).

Offers a set of canned routines immediately available (the *`mediation`* package) to you.

Thank you for the kind attention!

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

Baron, R. M., & Kenny, D. A. (1986). The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, *14*(1), 63–82.

Bullock, J. G., & Ha, S. E. (2011). Mediation Analysis is Harder Than it Looks. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 508–521). Cambridge: Cambridge University Press.

Imai, K., Keele, L., & Tingley, D. (2010). A General Approach to Causal Mediation Analysis. *Psychological Methods*, *15*(4), 309–334.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *The American Political Science Review*, *105*(4), 765–789.

Schaub, M., & Morisi, D. (2020). Voter Mobilisation in the Echo Chamber: Broadband Internet and the Rise of Populism in Europe. *European Journal of Political Research*, *59*(4).