Linear mediation

Session 12

MATH 80667A: Experimental Design and Statistical Methods
HEC Montréal

Linear mediation

Reminder: three types of causal associations

Confounding

Common cause

Causal forks X ← Z → Y

Causation

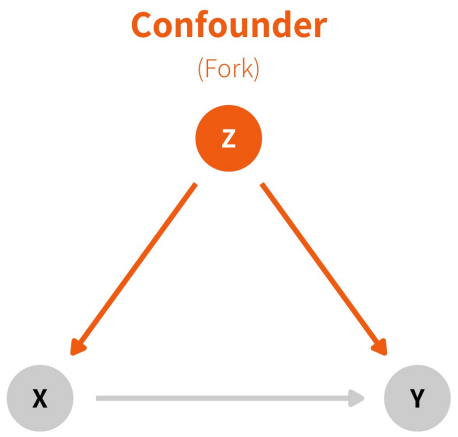
Mediation

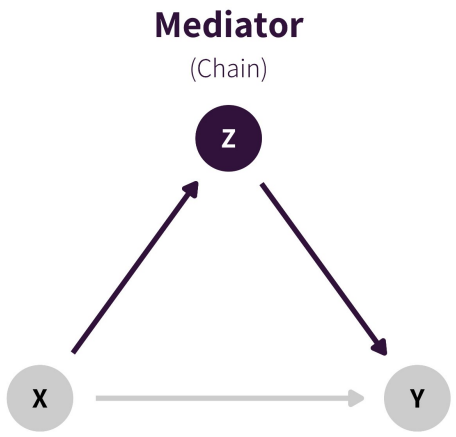
Causal chain X → Z → Y

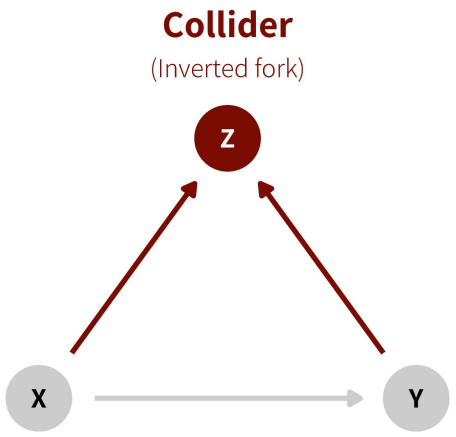
Collision

Selection / endogeneity

inverted fork X → Z ← Y







Notation

Define

- treatment of individual i as X_i, typically binary or categorical with $X_i \in \{0, \ldots, K\}$ and
  - $X = 0$ (control)
  - $X = k$ (treatment) for $k = 1, \ldots, K$
- potential mediation given treatment x as $M_i(x)$ and
- potential outcome for treatment x and mediator m as $Y_i(x, m)$.

Sequential ignorability assumption

1. Given pre-treatment covariates $\boldsymbol{Z}$, potential outcomes for mediation and treatment are conditionally independent of treatment assignment.

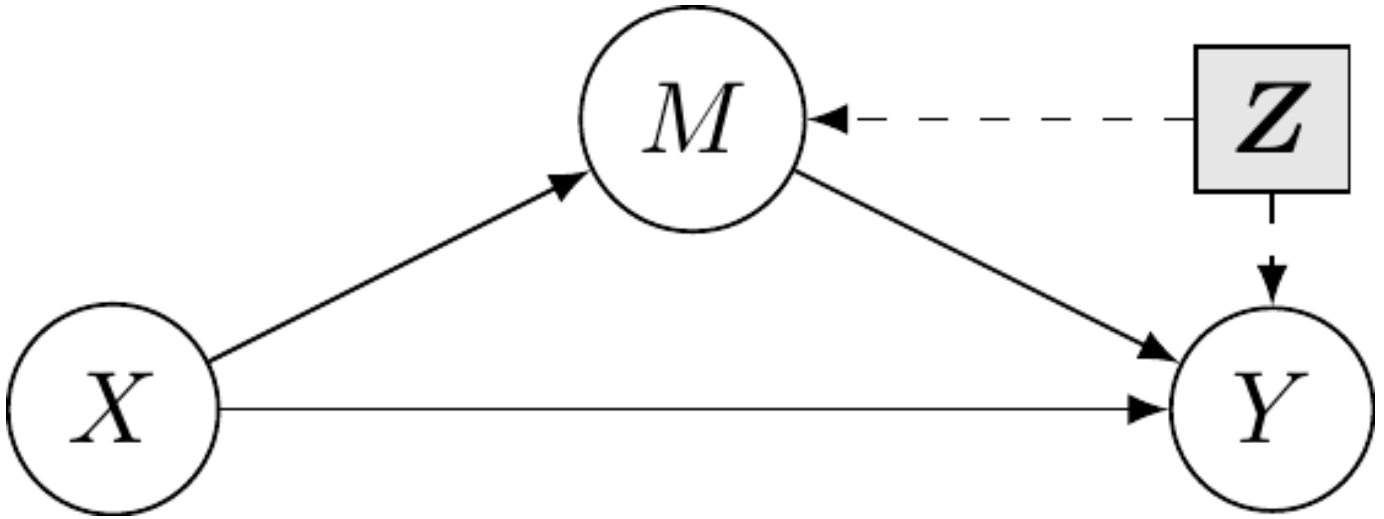
$$Y_i(x', m), M_i(x) \perp\!\!\!\perp X_i \mid \boldsymbol{Z}_i = \boldsymbol{z}$$

2. Given pre-treatment covariates $\boldsymbol{Z}$ and observed treatment x, potential outcomes for the response are independent of mediation.

$$Y_i(x', m) \perp\!\!\!\perp M_i(x) \mid X_i = x, \boldsymbol{Z}_i = \boldsymbol{z}$$

- Assumption 1 holds under randomization of treatment.
- Assumption 2 implies there is no confounder affecting both Y_i, M_i.

Directed acyclic graph



Directed acyclic graph of the linear mediation model

Z are potential confounders. If we randomly allocate X (experiment), then all incoming arrows vanish and we have no confounder for X vs M or X vs Y.

Average directed effect

The ADE measures the flow along $X \to Y$, disabling the pathway $X \to M \to Y$ by fixing the mediator value: it is

$$\mathsf{ADE}(x) = \mathsf{E}[Y_i(1, M_i(x)) - Y_i(0, M_i(x))\}$$

This measures the expected change in response when the experimental factor changes from treatment to control, while the mediator is set to a fixed value $M_i(x)$ uniformly over the population. Fixing the mediator may or not be feasible experimentally.

If there is no interaction between the treatment and the mediator, then $\mathsf{ADE}(0) = \mathsf{ADE}(1)$.

Average causal mediation effect

Also called indirect effect, obtained for a fixed intervention due to changing the values of the mediator to those it would take under the treatment and control group, respectively $M_i(1)$ and $M_i(0)$.

$$\mathsf{ACME}(x) = \mathsf{E}[Y_i\{x, M_i(1)\} - Y_i\{x, M_i(0)\}]$$

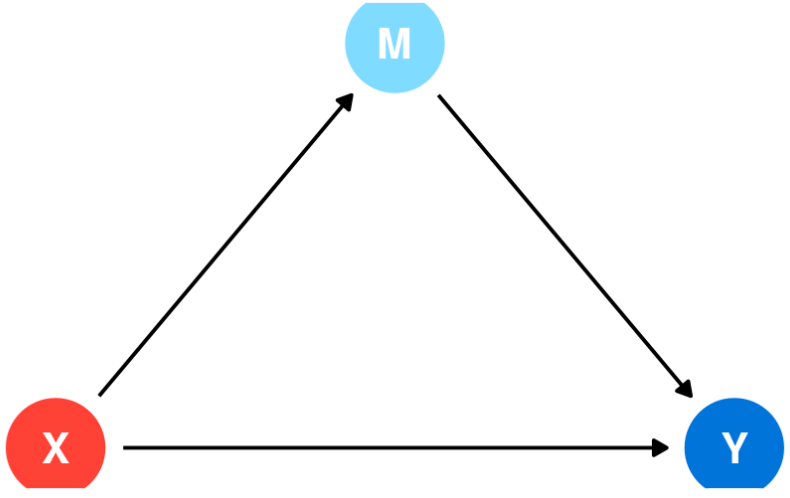
If there is no interaction between the treatment and the mediator, then $\mathsf{ACME}(0) = \mathsf{ACME}(1)$

Total effect

Total effect: overall impact of X (both through M and directly)

$$\mathsf{TE} = \mathsf{E}[Y \mid \mathrm{do}(X = 1)] - \mathsf{E}[Y \mid \mathrm{do}(X = 0)]$$

X → M → Y
plus
X → Y



The total effect is the average change in response if we randomize treatment assignment and consider the difference treatment vs control.

Total effect

The **total effect** measures the average overall impact of changes in outcome Y (both through M and directly) when experimentally manipulating X,

$$\mathsf{TE} = \mathsf{E}[Y_i\{1, M(1)\}] - \mathsf{E}[Y_i\{0, M(0)\}],$$

- The total effect is obtained as

$$\mathsf{TE} = \mathsf{ACME}(X) + \mathsf{ADE}(1 - X), \qquad X = 0, 1.$$

Linear mediation model

Consider the following two linear regression models with a binary treatment $X \in \{0, 1\}$ and M binary or continuous:

$$\underset{\text{mediator}}{M} = \underset{\text{intercept}}{c_m} + \alpha X + \underset{\text{error term}}{\varepsilon_m}$$

$$\underset{\text{response}}{Y} = \underset{\text{intercept}}{c_y} + \underset{\text{direct effect}}{\beta X} + \gamma M + \underset{\text{error term}}{\varepsilon_y}$$

We assume that zero-mean error terms ε_m and ε_y are **uncorrelated**.

- This is tied to the *no confounders* assumption.

Total effect decomposition

Plugging the first equation in the second, we get the marginal model for Y given treatment X

$$\mathsf{E}(Y \mid X = x) = \underbrace{(c_y + \gamma c_m)}_{\text{intercept}} + \underbrace{(\beta + \alpha\gamma)}_{\text{total effect}} \cdot x$$

In an experiment, we can obtain the total effect via the ANOVA model, with

$$Y = \underbrace{\nu}_{\text{average of control}} + \underbrace{\tau X}_{\text{total effect}} + \underbrace{\varepsilon_{Y'}}_{\text{error term}}$$

$$\tau = \mathrm{E}\{Y \mid \mathrm{do}(X = 1)\} - \mathrm{E}\{Y \mid \mathrm{do}(X = 0)\}$$

Product of coefficient method

In the linear mediation model of Baron and Kenny, the quantities of interest are

$$\begin{aligned}
\mathsf{ACME}(x) &= \mathsf{E}[Y\{x, M(1)\} - Y\{x, M(0)\}] = \alpha\gamma \\
\mathsf{ADE}(x) &= \mathsf{E}[Y\{1, M(x)\} - Y\{0, M(x)\}] = \beta \\
\mathsf{TE} &= \mathsf{E}[Y\{1, M(1)\} - Y\{0, M(0)\}] = \beta + \alpha\gamma
\end{aligned}$$

Model assumptions

There are three main assumptions for this quantity to be a valid estimator of the causal mediation effect

- the effect of treatment is **linear**.
- there is no interaction between mediator M and treatment X.
- the sequential ignorability assumption holds.

Example from Preacher and Hayes (2004)

> Suppose an investigator is interested in the effects of a new cognitive therapy on life satisfaction after retirement.
>
> Residents of a retirement home diagnosed as clinically depressed are randomly assigned to receive 10 sessions of a new cognitive therapy $(X = 1)$ or 10 sessions of an alternative (standard) therapeutic method $(X = 0)$.
>
> After Session 8, the positivity of the attributions the residents make for a recent failure experience is assessed (M).
>
> Finally, at the end of Session 10, the residents are given a measure of life satisfaction (Y). The question is whether the cognitive therapy's effect on life satisfaction is mediated by the positivity of their causal attributions of negative experiences. "

Sobel's test

Based on estimators of coefficients $\widehat{\alpha}$ and $\widehat{\gamma}$, construct a test statistic

$$S = \frac{\widehat{\alpha}\widehat{\gamma} - 0}{\mathsf{se}(\widehat{\alpha}\widehat{\gamma})}$$

The coefficient and variance estimates can be extracted from the output of the regression model.

In large sample, $S \dot{\sim} \mathsf{Normal}(0, 1)$, but this approximation may be poor in small samples.

Other test statistics

Sobel's test is not the only test. Alternative statistics are discussed in

> MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7(1), 83–104. https://doi.org/10.1037/1082-989X.7.1.83

Alternative

An alternative to estimate *p*-value and the confidence interval is through the nonparametric **bootstrap** with the percentile method, popularized by Preacher and Hayes (2004)

Nonparametric bootstrap: repeat B times, say $B = 10\ 000$

1. sample n (same as original number of observations) tuples (Y_i, X_i, M_i) from the database **with replacement** to obtain a new sample.
2. recalculate estimates $\widehat{\alpha}^{(b)}\widehat{\gamma}^{(b)}$ for each bootstrap dataset

Bootstrap confidence intervals

Percentile-based method: for a equitailed $1 - \alpha$ interval

1. Run the nonparametric bootstrap and obtain estimates $\widehat{\alpha}^{(b)}$ and $\widehat{\gamma}^{(b)}$ from the bth bootstrap sample.
2. Compute the $\alpha/2$ and $1 - \alpha/2$ empirical quantiles of

$$\{\widehat{\alpha}^{(b)}\widehat{\gamma}^{(b)}\}_{b=1}^{B}.$$

Boostrap two-sided p-value

Compute the sample proportion of bootstrap statistics that are larger/smaller than zero.

1. Order bootstrap statistics $S^{(1)} \leq \cdots \leq S^{(B)}$ and let $S^{(0)} = -\infty$, $S^{(M+1)} = \infty$.
2. Find M ($0 \leq M \leq B$) such that $S^{(M)} < 0 \leq S^{(M+1)}$ (if it exists)
3. The p-value is

$$p = 2 \min\{M/B, 1 - M/B\}.$$

Model assumptions

Same assumptions as analysis of variance and linear models

- Linearity of the mean model
  - residual plots, fitted values $\hat{y}$ against m and x
- Independent/uncorrelated errors
  - no confounding, lack of serial correlation (e.g., cross-panels)
- Equal variance of errors in each model (homoskedasticity)
- Large samples

Causal assumptions

Conclusions about mediation are valid only when causal assumptions hold.

Assuming that X is randomized, we need

- Lack of interaction between X and M (moderated mediation)
- Causal direction: $M \rightarrow Y$, so M must be an antecedent cause
  - M must be measured before Y
- Reliability of M (no measurement error)
- No confounding between X and M
  - can be included, but not mediators/colliders + correct form
- effect constant over individuals/levels

Sensitivity analysis

The no-unmeasured confounders assumption should be challenged.

Check the robustness of the conclusions by considering potential correlation between errors, as

$$\mathsf{E}(\widehat{\gamma}) = \gamma + \mathsf{Cor}(\varepsilon_m, \varepsilon_y)\sqrt{\frac{\mathsf{Va}(\varepsilon_y)}{\mathsf{Va}(\varepsilon_m)}}$$

- We vary $\rho = \mathsf{Cor}(\varepsilon_m, \varepsilon_y)$ to assess the sensitivity to confounding.
- The `medsens` function in the **R** package `mediation` implements the diagnostic of Imai, Keele and Yamamoto (2010) for the linear mediation model.

Defaults of linear mediation models

- Definitions contingent on model
  - (even if causal quantities have a meaning regardless of estimation method)
- Most papers do not consider confounders, or even check for assumptions
- Generalizations to interactions, multiple mediators, etc., requires care.

Keenan Crane

Extension: moderated mediation

Consider a more complex setting where the effect of the experimental factor X depends on the mediator, a case termed **moderated mediator** Judd and Kenny (1982).

In this case, the equation for the response variable becomes

$$\mathsf{E}(Y \mid M = m, X = x, \boldsymbol{Z} = \boldsymbol{z}) = c_Y + \beta x + \gamma m + \kappa x m + \boldsymbol{z}\boldsymbol{\omega}$$

ACME for moderated mediation

Upon substituting the equations for both inside the definition of average causal mediation effect, we find that the latter equals

$$\mathsf{ACME}(x) = (\gamma + \kappa x)\{M(1) - M(0)\} = \alpha(\gamma + \kappa x).$$

and thus the value differs depending on experimental regime (treatment or control), due to the presence of the interaction.

Both the average direct and total effects now depend on the theoretical average values of the covariates $\boldsymbol{Z}$ added to the model to control for confounding.

General approach

Imai et al. (2010) suggest using simulation for general models that naturally account for nonlinear effects, different natures of the mediator (binary, categorical, etc.) and the response.

- [1] **Modelling**. Specify models to each relationship
  1. a mediator model $f_M(x, \boldsymbol{z}; \boldsymbol{\theta}_M)$ that includes X and potential confounders $\boldsymbol{Z}$
  2. an outcome model $f_Y(x, m, \boldsymbol{z}; \boldsymbol{\theta}_Y)$ with both treatment indicator X, mediator M and potential confounders $\boldsymbol{Z}$.
- [2] **Estimation uncertainty**: Either use nonparametric bootstrap or a large-sample approximation to the distribution of model parameters $\boldsymbol{\theta}_M$ and $\boldsymbol{Y}$ to get J copies.

General approach

- [3] For each copy of the parameters $j = 1, \ldots, J$, draw M realizations of the mediation model for each observation $i = 1, \ldots, n$, giving $M_i^{(jm)}(1)$ and $M_i^{(jm)}(0)$. Use these resulting to obtain one draw from the outcome model for each of

$$Y_i^{(jm)}\{x_1, M_i^{(jm)}(x_2)\}, \qquad x_1, x_2 \in \{0, 1\}.$$

- [4] Plug these quantities in the definitions of $\mathsf{ACME}(x)$, $\mathsf{ADE}(x)$ and TE, averaging over the nM replications, over both observations $i = 1, \ldots, n$ and replications $m = 1, \ldots, M$. This yields a collection of J values $\mathsf{ACME}_j(x)$, $\mathsf{ADE}_j(x)$ and TE_j for each parameter value.
- [5] Use the J replications to compute a 95% percentile interval and return the sample means over J for the average causal effects.

Key references

- Judd and Kenny (1981). Process analysis: Estimating mediation in treatment evaluations, Evaluation Review
- Baron and Kenny (1986), The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations, *Journal of Personality and Social Psychology*.
- Shrout and Bolger(2002), Mediation in experimental and nonexperimental studies: New procedures and recommendations, *Psychological Methods*.
- Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behavior Research Methods, Instruments & Computers.
- Imai, Keele and Tingley (2010), A General Approach to Causal Mediation Analysis, *Psychological Methods*.
- Imai, Tingley and Yamamoto (2013), Experimental designs for identifying causal mechanisms (with Discussion), Journal of the Royal Statistical Society: Series A.
- Pearl (2014), Interpretation and Identification of Causal Mediation, *Psychological Methods*.
- Bullock, Green, and Ha (2010), Yes, but what's the mechanism? (don't expect an easy answer)
- Uri Simonsohn (2022) Mediation Analysis is Counterintuitively Invalid