Linear mediation and moderation

Session 12

MATH 80667A: Experimental Design and Statistical Methods
HEC Montréal

Outline

Linear mediation model

Moderation

Linear mediation

Reminder: three types of causal associations

Confounding

Common cause

Causal forks $X \leftarrow Z \rightarrow Y$

Causation

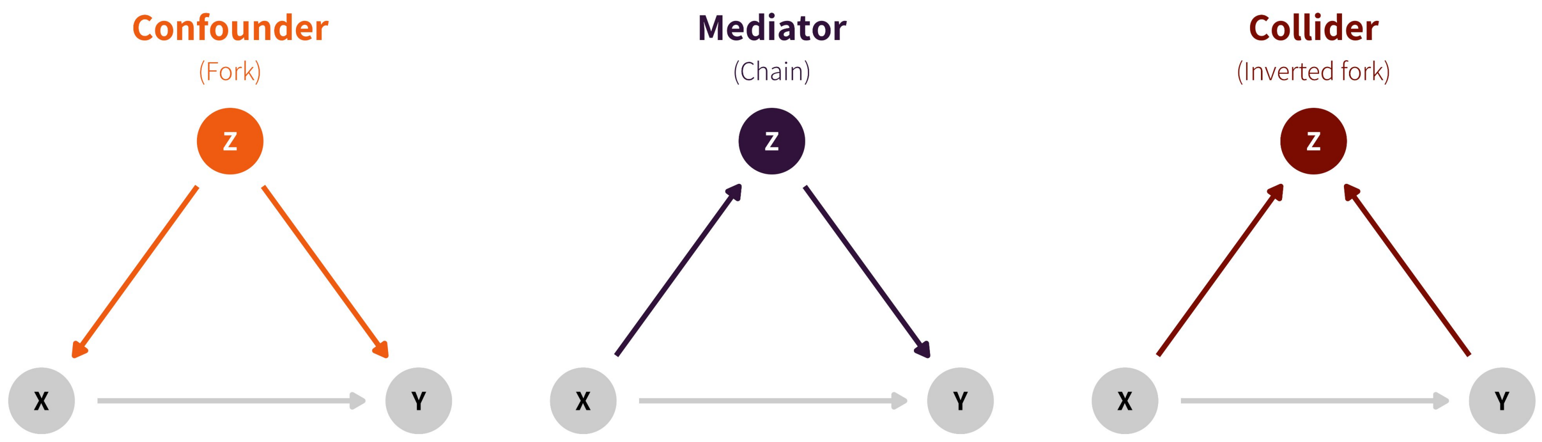
Mediation

Causal chain $X \rightarrow Z \rightarrow Y$

Collision

Selection / endogeneity

inverted fork $X \rightarrow Z \leftarrow Y$



Notation

Define

- treatment of individual i as X_i, typically binary with $X_i \in \{0, 1\}$ and
  - $X = 0$ (control), else $X = x_0$
  - $X = 1$ (treatment)
- 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.

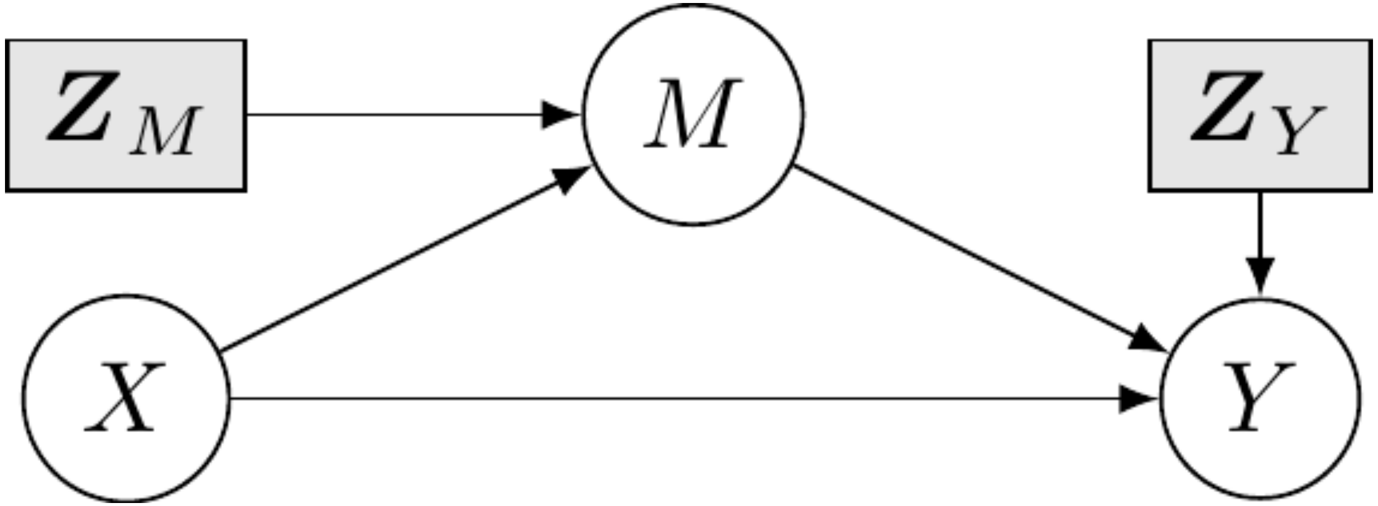
$$Y_i(x', m), M_i(x) \perp\!\!\!\perp X_i \mid \boldsymbol{Z}_i = \boldsymbol{z}$$

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

$\boldsymbol{Z}_M$ and $\boldsymbol{Z}_Y$ are controls for confounders, may or not be present in the model.

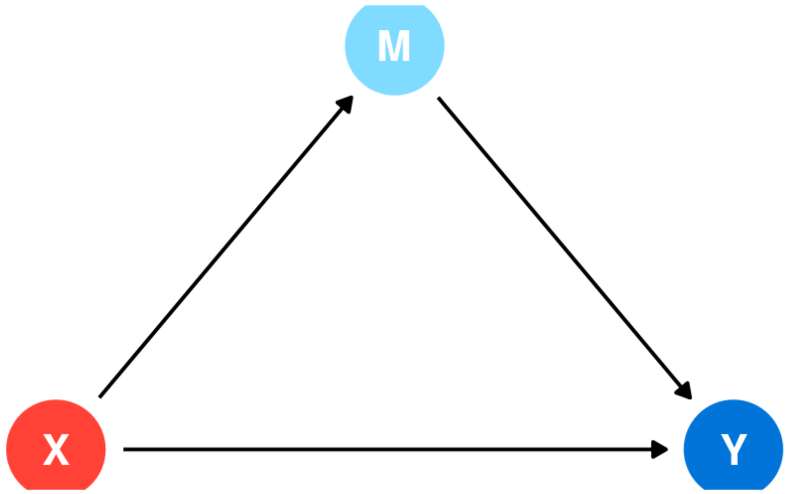
Total effect

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

$$\text{TE}(x, x_0) = \text{E}[Y \mid \mathrm{do}(X = x)] - \text{E}[Y \mid \mathrm{do}(X = x_0)]$$

This can be generalized for continuous X to any pair of values (x_1, x_2).

X → M → Y
plus
X → Y



Average controlled direct effect

$$
\begin{aligned}
\mathsf{ACDE}(m, x, x_0) &= \mathsf{E}\{Y_i(x, m) - Y_i(x_0, m)\} \\
&= \mathsf{E}\{Y \mid \mathrm{do}(X = x, m = m)\} - \mathsf{E}\{Y \mid \mathrm{do}(X = x_0, m = m)\}
\end{aligned}
$$

The average controlled direct effect (ACDE) is the expected change in response for the population when

- the experimental factor changes from x to x_0 and
- the mediator is set to a fixed value m

This typically requires experimental manipulation of both variables.

Direct and indirect effects

Natural direct effect: the expected change in Y under treatment x if M is set to whatever value it would take under control x_0

$$\mathsf{NDE}(x, x_0) = \mathsf{E}[Y\{x, M(x_0)\} - Y\{x_0, M(x_0)\}]$$

Natural indirect effect: the expected change in Y if we set X to its control value and change the mediator value which it would attain under x

$$\mathsf{NIE}(x, x_0) = \mathsf{E}[Y\{x_0, M(x)\} - Y\{x_0, M(x_0)\}]$$

Counterfactual conditioning reflects a physical intervention (experimentation), not mere conditioning.

Necessary and sufficiency of mediation

From Pearl (2014):

> The difference $\text{TE} - \text{NDE}$ quantifies the extent to which the response of Y is owed to mediation, while NIE quantifies the extent to which it is explained by mediation. These two components of mediation, the necessary and the sufficient, coincide into one in models void of interactions (e.g., linear) but differ substantially under moderation

The Baron–Kenny linear mediation model

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

$$\begin{aligned} \underbrace{M}_{\text{mediator}} &= \underbrace{c_M}_{\text{intercept}} + \alpha X + \underbrace{\varepsilon_M}_{\text{error term}} \\ \underbrace{Y}_{\text{response}} &= \underbrace{c_Y}_{\text{intercept}} + \underbrace{\beta X}_{\text{direct effect}} + \gamma M + \underbrace{\varepsilon_Y}_{\text{error term}} \end{aligned}$$

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

$$\begin{aligned} Y &= \underbrace{\nu}_{\text{average of control}} + \underbrace{\tau X}_{\text{total effect}} + \underbrace{\varepsilon_{Y'}}_{\text{error term}} \\ \tau &= \mathrm{E}\{Y \mid \mathrm{do}(X = 1)\} - \mathrm{E}\{Y \mid \mathrm{do}(X = 0)\} \end{aligned}$$

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. ”

Old method

This approach has been discontinued, but still appears in older papers.

Baron and Kenny recommended running three linear regressions and testing

1. whether $\mathscr{H}_0 : \alpha = 0$
2. whether $\mathscr{H}_0 : \tau = 0$ (total effect)
3. whether $\mathscr{H}_0 : \gamma = 0$

The average conditional mediation effect (ACME) in the linear mediation model is $\alpha\gamma$ and we can check whether it's zero using Sobel's test statistic.

Problems with Baron–Kenny approach

- We conduct three tests, so this inflates the Type I error.
- The total effect can be zero because $\alpha\gamma = -\beta$, even if there is mediation.
- The method has lower power to detect mediation when effect sizes are small.

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, $S \stackrel{\cdot}{\sim} \mathsf{No}(0,1)$, 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
  - can be added to model, then use NID definition
- 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.

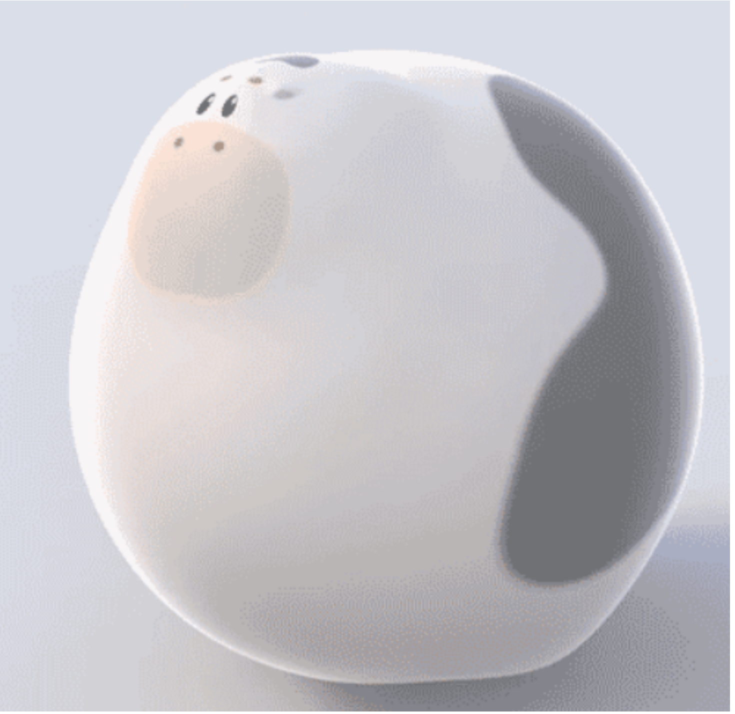
One way to assess the robustness of the conclusions to this is to consider correlation between errors, as (e.g., Bullock, Green and Ha, 2010)

$$\mathsf{E}(\widehat{\gamma}) = \gamma + \mathsf{Cov}(\varepsilon_M, \varepsilon_Y)/\mathsf{Va}(\varepsilon_M)$$

- We vary $\rho = \mathsf{Cor}(\varepsilon_M, \varepsilon_Y)$ to assess the sensitivity of our conclusions 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)
- It is possible to weaken assumptions (at the expense of more complicated models)
- Most papers do not consider confounders, or even check for assumptions
- Generalizations to interactions, multiple mediators, etc., requires care



Keenan Crane

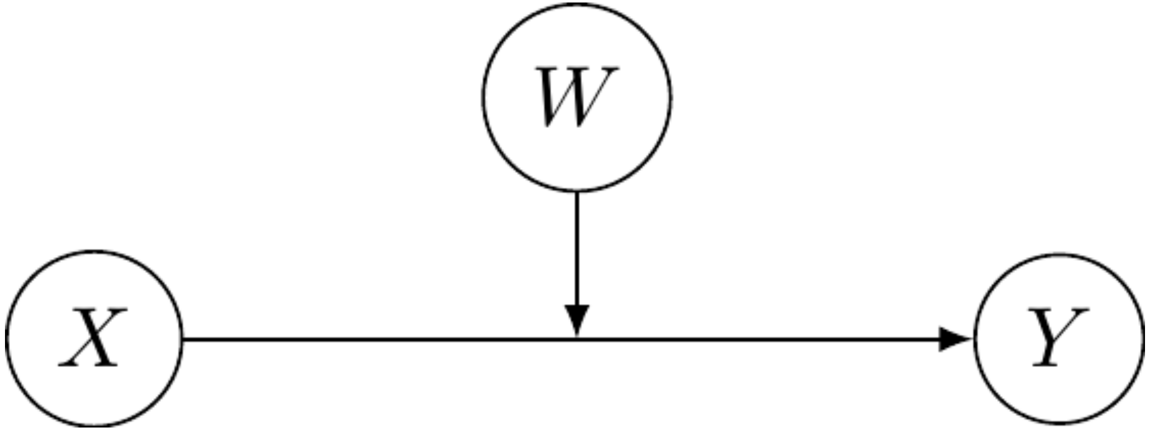
Key references

- Baron and Kenny (1986), The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations, *Journal of Personality and Social Psychology*
- 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
- 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.
- David Kenny's website

Moderation

Moderator

A **moderator** W modifies the direction or strength of the effect of an explanatory variable X on a response Y (interaction term).



Directed acyclic graph of moderation

Interactions are not limited to experimental factors: we can also have interactions with confounders, explanatories, mediators, etc.

Moderation in a linear regression model

In a regression model, we simply include an **interaction** term to the model between W and X.

For example, if X is categorical with K levels and W is binary or continuous, imposing sum-to-zero constraints for $\alpha_1, \ldots, \alpha_K$ and $\beta_1, \ldots, \beta_K$ gives

$$\underbrace{\mathrm{E}(Y \mid X = k, W = w)}_{\text{average response of group } k \text{ at } w} = \underbrace{\alpha_0 + \alpha_k}_{\text{intercept of group } k} + \underbrace{(\beta_0 + \beta_k)}_{\text{slope of group } k} w$$

Testing for the interaction

Test jointly whether coefficients associated to XW are zero, i.e.,

$$\beta_1 = \cdots = \beta_K = 0.$$

The moderator W can be continuous or categorical with $L \geq 2$ levels

The degrees of freedom (additional parameters for the interaction) in the F test are

- $K - 1$ for continuous W
  - are slopes parallel?
- $(K-1) \times (L-1)$ for categorical W
  - are all subgroup averages the same?

Example

We consider data from Garcia et al. (2010), a study on gender discrimination. Participants were given a fictional file where a women was turned down promotion in favour of male colleague despite her being clearly more experimented and qualified.

The authors manipulated the decision of the participant, with choices:

- not to challenge the decision (no protest),
- a request to reconsider based on individual qualities of the applicants (individual)
- a request to reconsider based on abilities of women (collective).

The postulated moderator variable is `sexism`, which assesses pervasiveness of gender discrimination.

Model fit

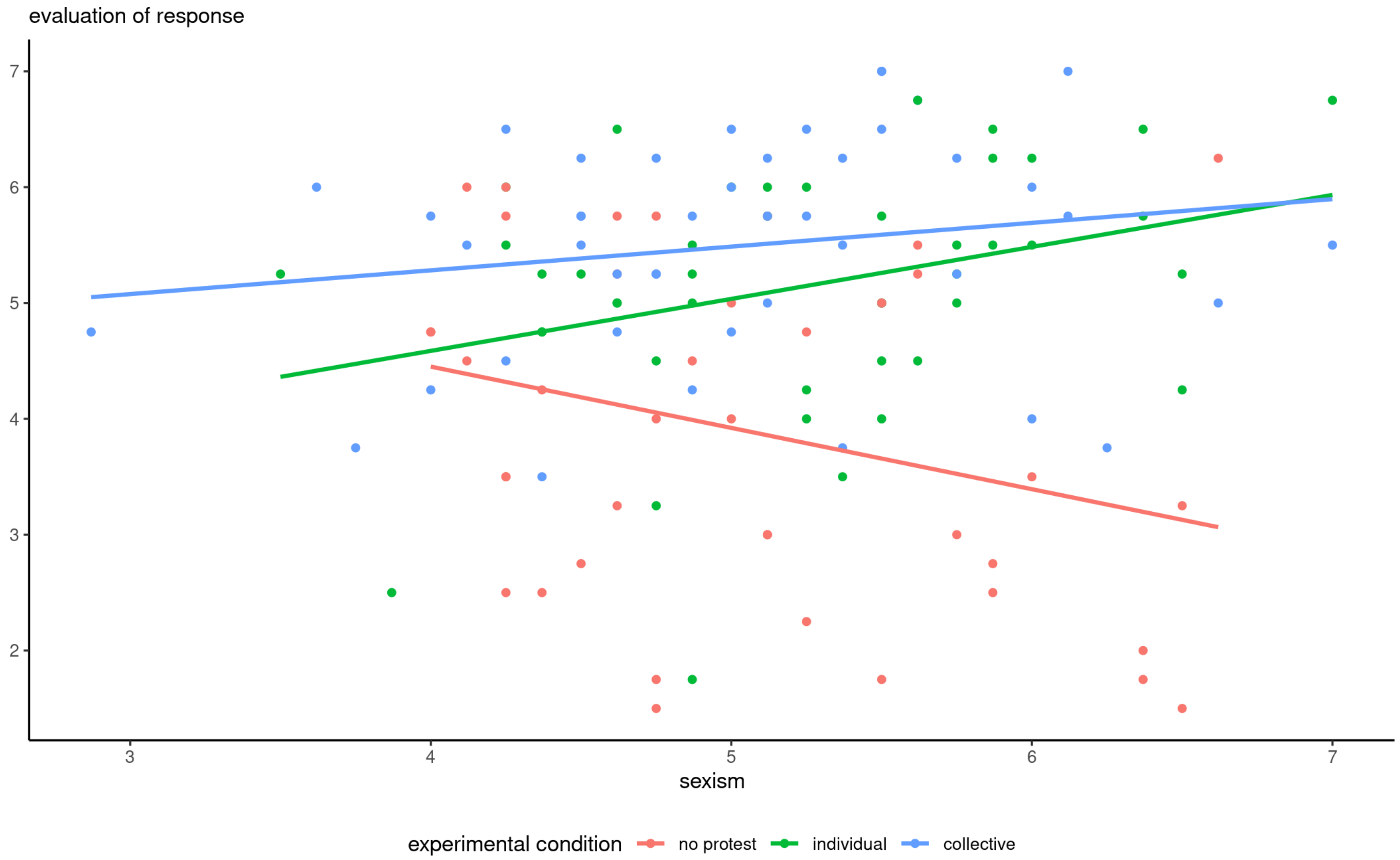
We fit the linear model with the interaction.

```
data(GSBE10, package = "hecedsm")
lin_moder <- lm(respeval ~ protest*sexism,
               data = GSBE10)
summary(lin_moder) # coefficients
car::Anova(lin_moder, type = 3) # tests
```

ANOVA table

term	sum of squares	df	stat	p-value
protest	6.34	2	2.45	.091
sexism	6.59	1	5.09	.026
protest:sexism	12.49	2	4.82	.010
Residuals	159.22	123		

Effects



Results won't necessarily be reliable outside of the range of observed values of sexism.

Comparisons between groups

Simple effects and comparisons must be done for a fixed value of sexism (since the slopes are not parallel).

The default value in `emmeans` is the mean value of `sexism`, but we could query for averages at different values of sexism (below for empirical quartiles).

```
quart <-  quantile(GSBE10$sexism, probs = c(0.25, 0.5, 0.75))
emmeans(lin_moder,
        specs = "protest",
        by = "sexism",
        at = list("sexism" = quart))
```

With moderating *factors*, give weights to each sub-mean corresponding to the frequency of the moderator rather than equal-weight to each category (`weights = "prop"`).

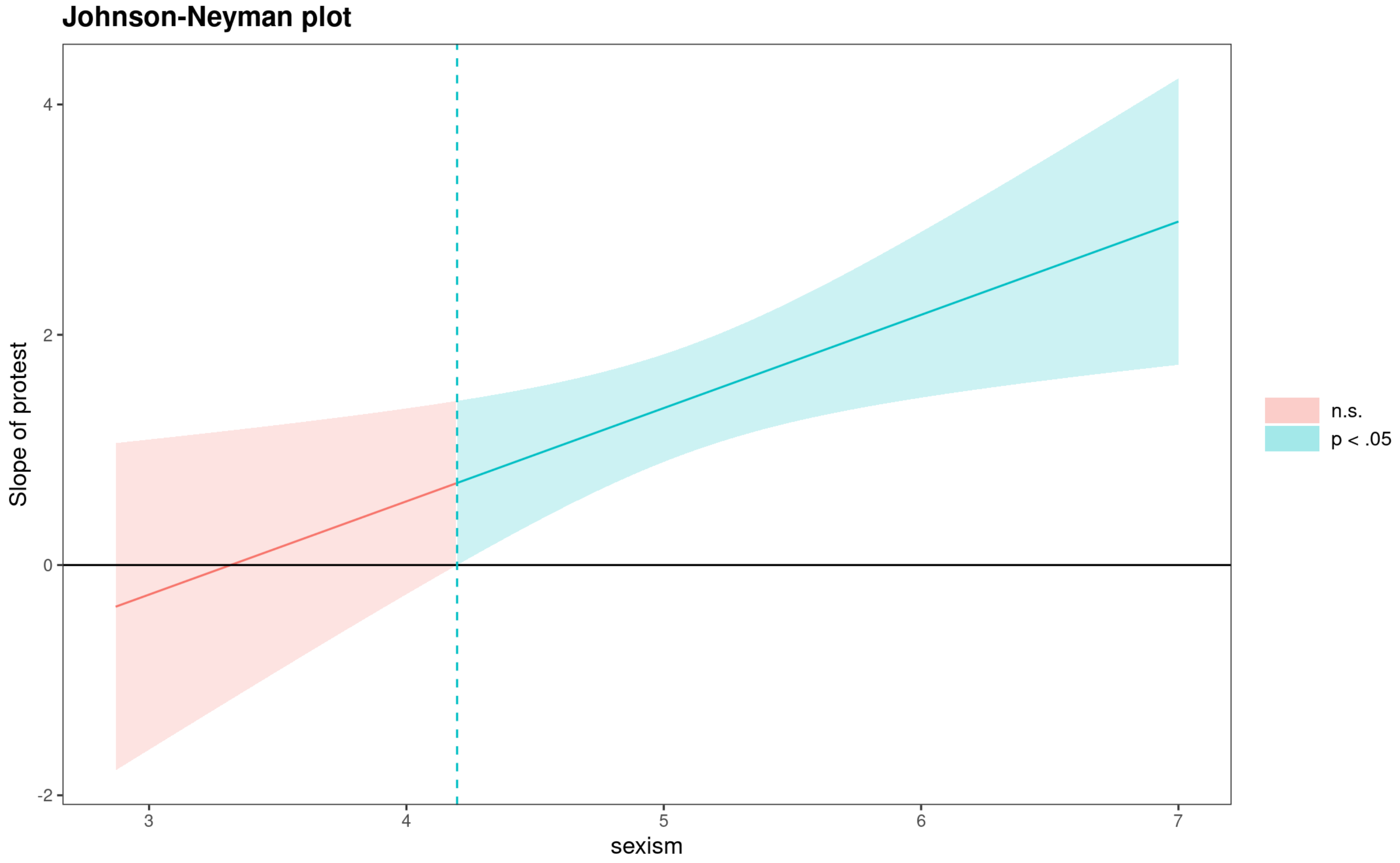
Sensitivity analysis

The Johnson and Neyman (1936) method looks at the range of values of moderator W for which difference between treatments (binary X) is not statistically significant.

```
lin_moder2 <- lm(
  respeval ~ protest*sexism,
  data = GSBE10 |>
  # We dichotomize the manipulation, pooling protests together
  dplyr::mutate(protest = as.integer(protest != "no protest")))
# Test for equality of slopes/intercept for two protest groups
anova(lin_moder, lin_moder2)
# p-value of 0.18: fail to reject individual = collective.
```

Syntax for plot

```
jn <- interactions::johnson_neyman(
  model = lin_moder2, # linear model
  pred = protest, # binary experimental factor
  modx = sexism, # moderator
  control.fdr = TRUE, # control for false discovery rate
  mod.range = range(GSBE10$sexism)) # range of values for sexism
jn$plot
```



Johnson–Neyman plot for difference between protest and no protest as a function of sexism.