Introduction to Causal Mediation Analysis

Amanda Kay Montoya

Offered by

GLOBAL
SCHOOL
EMPIRICAL
RESEARCH
METHODS

June 2025

Example for the class

Garcia, D. M., Schmitt, M. T., Branscombe, N. R., & Ellemers, N. (2010). Women's reactions to ingroup members who protest discriminatory treatment: The importance of beliefs about inequality and response appropriateness. ***European Journal of Social Psychology, 49***, 733-745.

European Journal of Social Psychology
Eur. J. Soc. Psychol. **40**, 733–745 (2010)
Published online 6 July 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/ejsp.644

Research article

Women's reactions to ingroup members who protest discriminatory treatment: The importance of beliefs about inequality and response appropriateness

DONNA M. GARCIA[1]*, MICHAEL T. SCHMITT[2],
NYLA R. BRANSCOMBE[3] AND NAOMI ELLEMERS[4]
[1] *University of Guelph, Canada*
[2] *Simon Fraser University, Canada*
[3] *University of Kansas, USA*
[4] *Leiden University, The Netherlands*

Abstract

Our goal was to identify factors that shape women's responses to ingroup members who protest gender discrimination. We predicted and found that women who perceived gender discrimination as pervasive regarded a protest response as being more appropriate than a no protest response and expressed greater liking and less anger towards a female lawyer who protested rather than did not protest an unfair promotion decision. Further, beliefs about the appropriateness of the response to discrimination contributed to evaluations of the protesting lawyer. Perceptions that the complaint was an appropriate response to the promotion decision led to more positive evaluations of an ingroup discrimination protester.

Protest can be an effective means of improving the plight of a devalued group. Historically, there are many examples of protest, even from a single individual, that have advanced a group's social position (e.g. Meritor Savings Bank vs. Vinson, 477 US 57, 1986; Dekker vs. VJV-Centrum ECJ, 1992). Despite the potential gains to be obtained by protesting illegitimate treatment, protestors might not always be appreciated by members of their own group. Whether disadvantaged group members respond positively or negatively to ingroup protestors will likely depend upon the *perceived* implications that the protestor's action has for the ingroup. Unless protest is seen as justified by the social circumstances and an effective means of bringing about positive change, a protestor might be seen as making the ingroup look like complainers. Such threat to the ingroup's reputation could evoke the ire and disdain of the disadvantaged group towards the protestor. Hence, perceptions of the justification for and likely consequences of protest will be critical to others' reactions to an ingroup discrimination claimer. We propose that protest by an ingroup member will be seen as appropriate and thus appreciated to the extent that observers perceive that their ingroup is targeted by pervasive discrimination.

SOCIAL CONSEQUENCES OF CLAIMING DISCRIMINATION

Gender discrimination continues to be widespread throughout Western employment settings (see Charles & Grusky, 2004). The continuation of gender discrimination has substantive negative implications for women's economic and

*Correspondence to: Donna M. Garcia, Department of Psychology, University of Guelph Ontario, Guelph, Ontario, N1G 2W1Canada.
E-mail: donnagarcia3@gmail.com

Received 19 October 2008
Accepted 7 April 2009

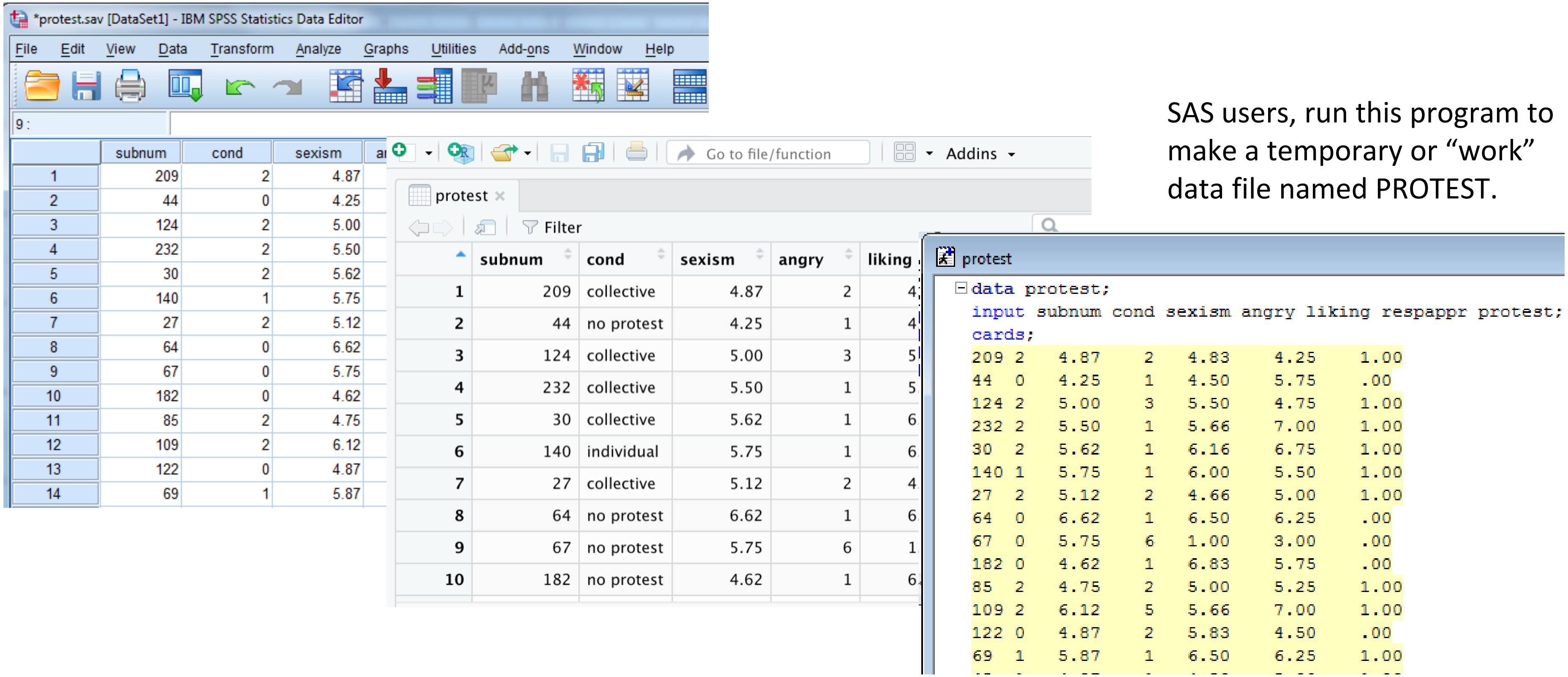
Participants (all female) read a narrative about a female attorney who lost a promotion at her firm to a much less qualified male through unequivocally discriminatory actions of the senior partners.

Participants assigned to the 'protest' condition were then told she protested the decision by presenting an argument to the partners about how unfair the decision was.

Participants assigned to the 'no protest' condition were told that although she was disappointed, she accepted the decision and continued working at the firm.

After reading the narrative, the participants evaluated **how appropriate they perceived her response to be**, and also evaluated the characteristics of the attorney, the responses of which were aggregated to produce a measure of **"liking."** Prior to the study, the participants filled out the Modern Sexism Scale.

The data: PROTEST


*protest.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
9 :
Go to file/function
Addins
protest
Filter
protest

	subnum	cond	sexism	ar
1	209	2	4.87	
2	44	0	4.25	
3	124	2	5.00	
4	232	2	5.50	
5	30	2	5.62	
6	140	1	5.75	
7	27	2	5.12	
8	64	0	6.62	
9	67	0	5.75	
10	182	0	4.62	
11	85	2	4.75	
12	109	2	6.12	
13	122	0	4.87	
14	69	1	5.87	

	subnum	cond	sexism	angry	liking
1	209	collective	4.87	2	4
2	44	no protest	4.25	1	4
3	124	collective	5.00	3	5
4	232	collective	5.50	1	5
5	30	collective	5.62	1	6
6	140	individual	5.75	1	6
7	27	collective	5.12	2	4
8	64	no protest	6.62	1	6
9	67	no protest	5.75	6	1
10	182	no protest	4.62	1	6

```
data protest;
input subnum cond sexism angry liking respappr protest;
cards;
209 2   4.87    2   4.83    4.25    1.00
44  0   4.25    1   4.50    5.75    .00
124 2   5.00    3   5.50    4.75    1.00
232 2   5.50    1   5.66    7.00    1.00
30  2   5.62    1   6.16    6.75    1.00
140 1   5.75    1   6.00    5.50    1.00
27  2   5.12    2   4.66    5.00    1.00
64  0   6.62    1   6.50    6.25    .00
67  0   5.75    6   1.00    3.00    .00
182 0   4.62    1   6.83    5.75    .00
85  2   4.75    2   5.00    5.25    1.00
109 2   6.12    5   5.66    7.00    1.00
122 0   4.87    2   5.83    4.50    .00
69  1   5.87    1   6.50    6.25    1.00
```

SAS users, run this program to make a temporary or “work” data file named PROTEST.

PROTEST: Experimental condition (1 = protest, 0 = no protest)

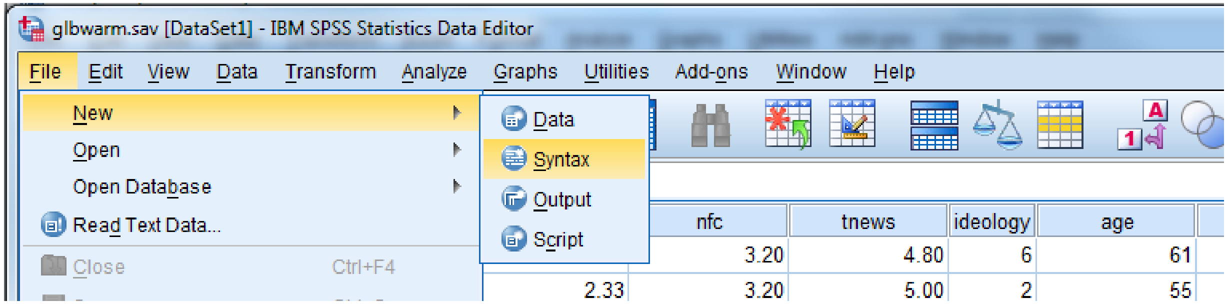
LIKING : Evaluation (liking) of the lawyer (higher = more positive evaluation, i.e. like more)

RESPAPPR: A measure of how appropriate the lawyer’s behavior in response to the action of the partners was perceived to be for the situation (higher = more appropriate)

Using SPSS syntax

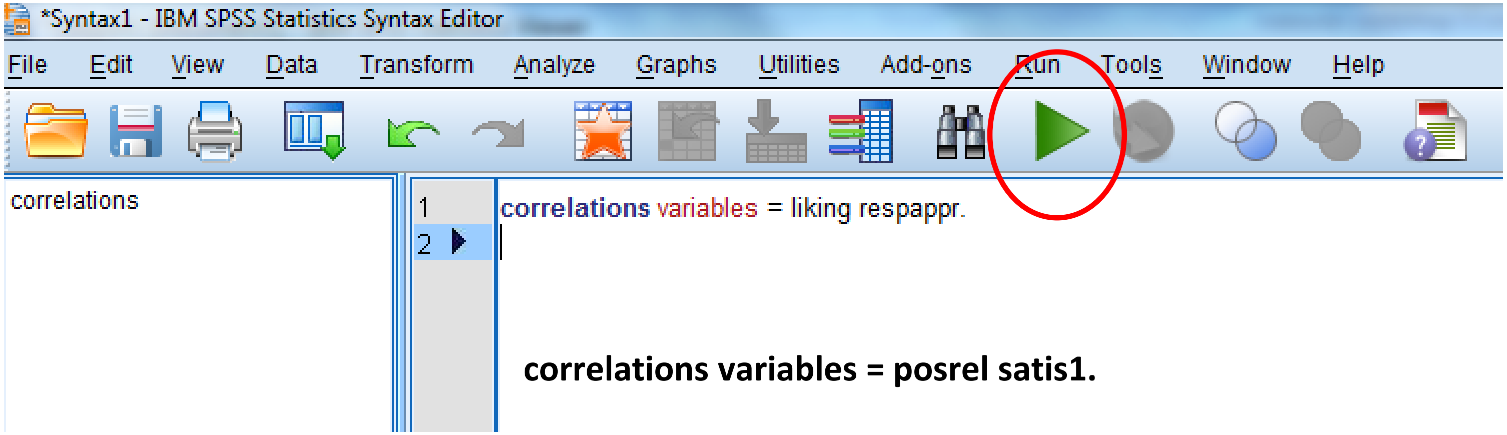
We will use syntax to instruct SPSS what to do in this class. There are many benefits of learning how to write SPSS syntax.

(1) Open a new syntax window (File > New > Syntax)



	nfc	tnews	ideology	age
	3.20	4.80	6	61
2.33	3.20	5.00	2	55

(2) Type your command(s) into the blank window that opens

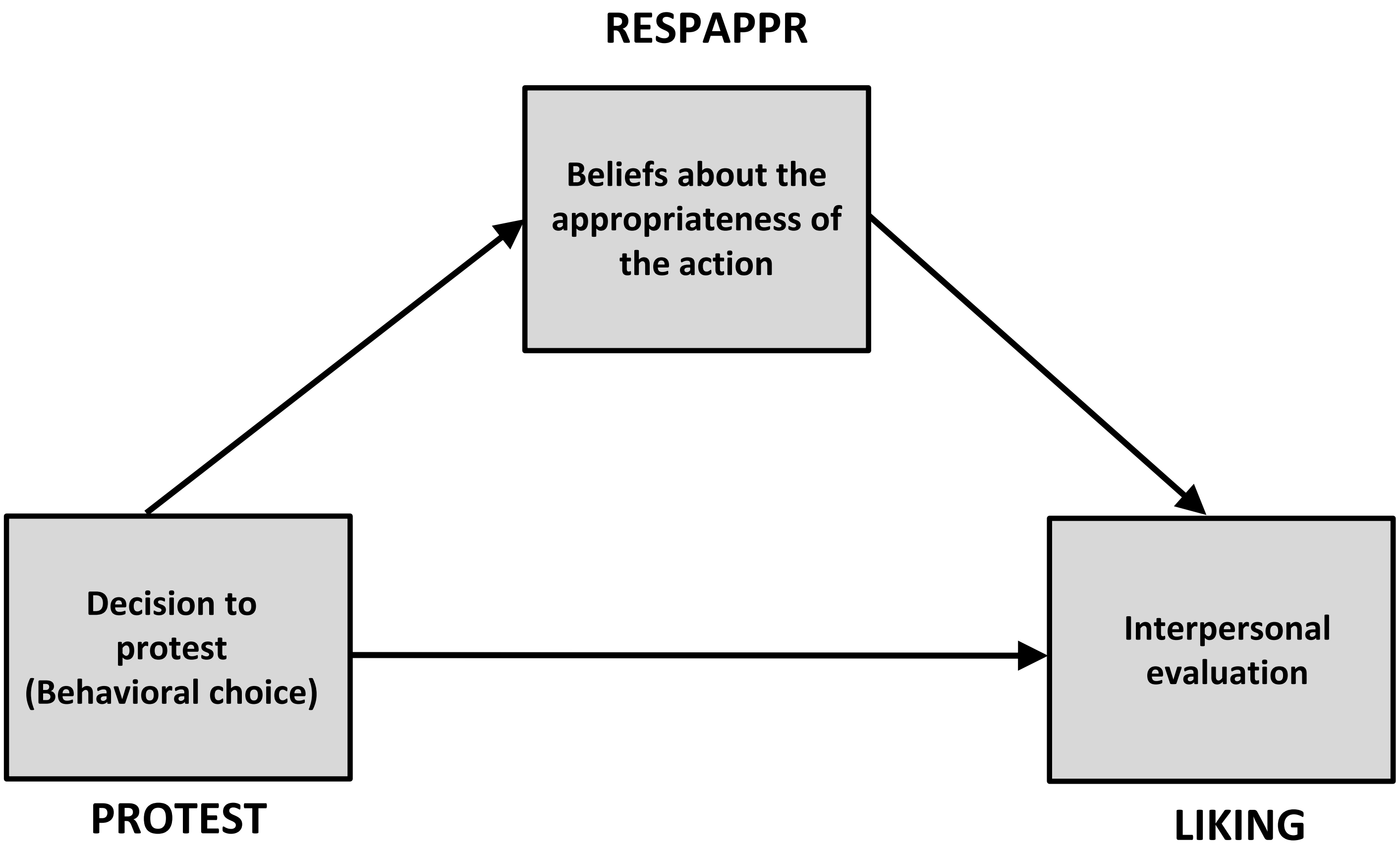


```
correlations variables = liking respappr.
```

correlations variables = posrel satis1.

(3) Click and drag to highlight code you want to execute and press the "play" button or select various options under "Run" in the syntax window menu. ▶

Our question

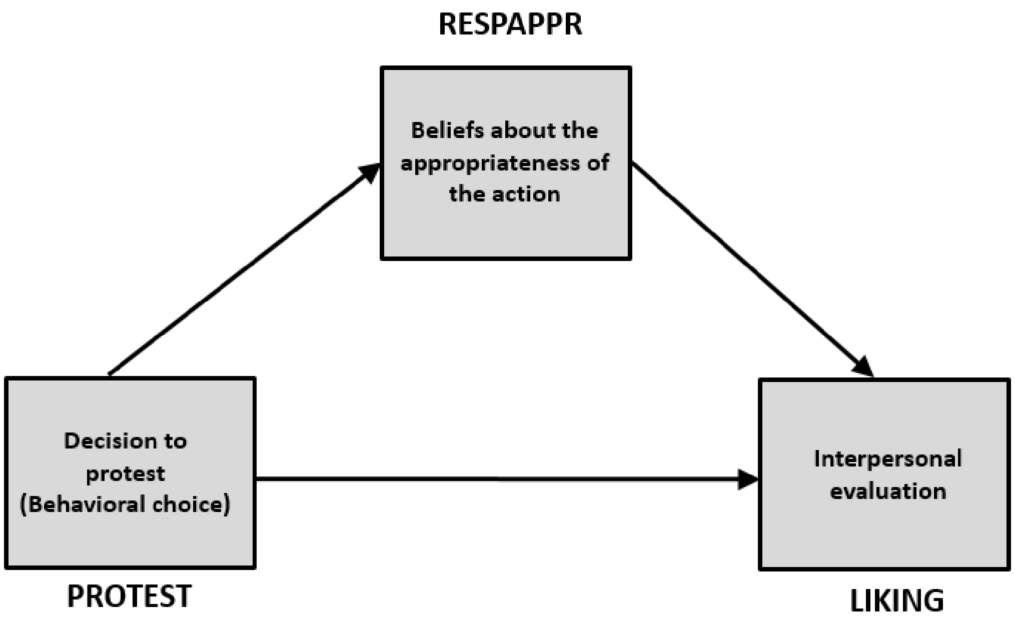

RESPAPPR
Beliefs about the appropriateness of the action
Decision to protest (Behavioral choice)
PROTEST
Interpersonal evaluation
LIKING

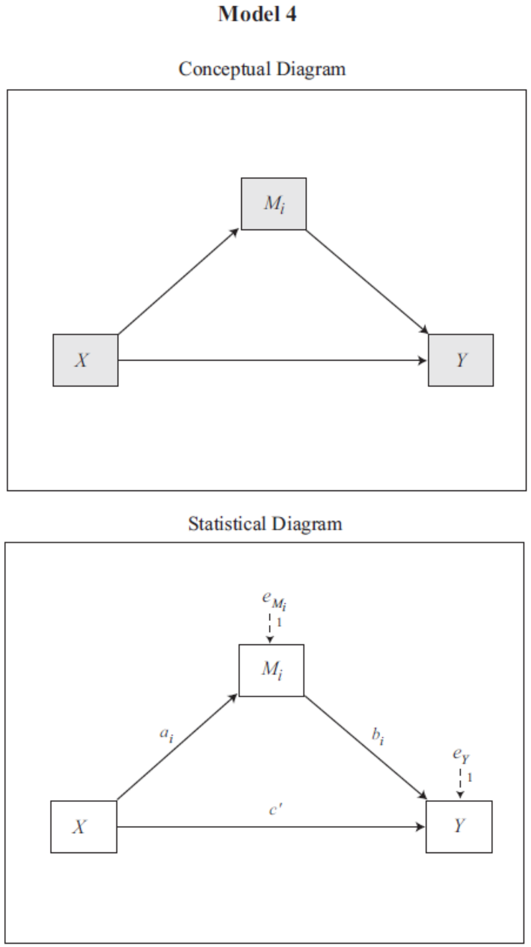
Do perceptions of the appropriateness of the response act as the mechanism through which that choice influences interpersonal evaluation?

Notice that this question is not asked contingent on evidence of simple association between the choice and the evaluation.

Estimation of the PROTEST model in PROCESS

PROCESS Model 4


RESPAPPR
Beliefs about the appropriateness of the action
Decision to protest (Behavioral choice)
PROTEST
Interpersonal evaluation
LIKING


Model 4

Conceptual Diagram

Statistical Diagram

What should the PROCESS command look like?

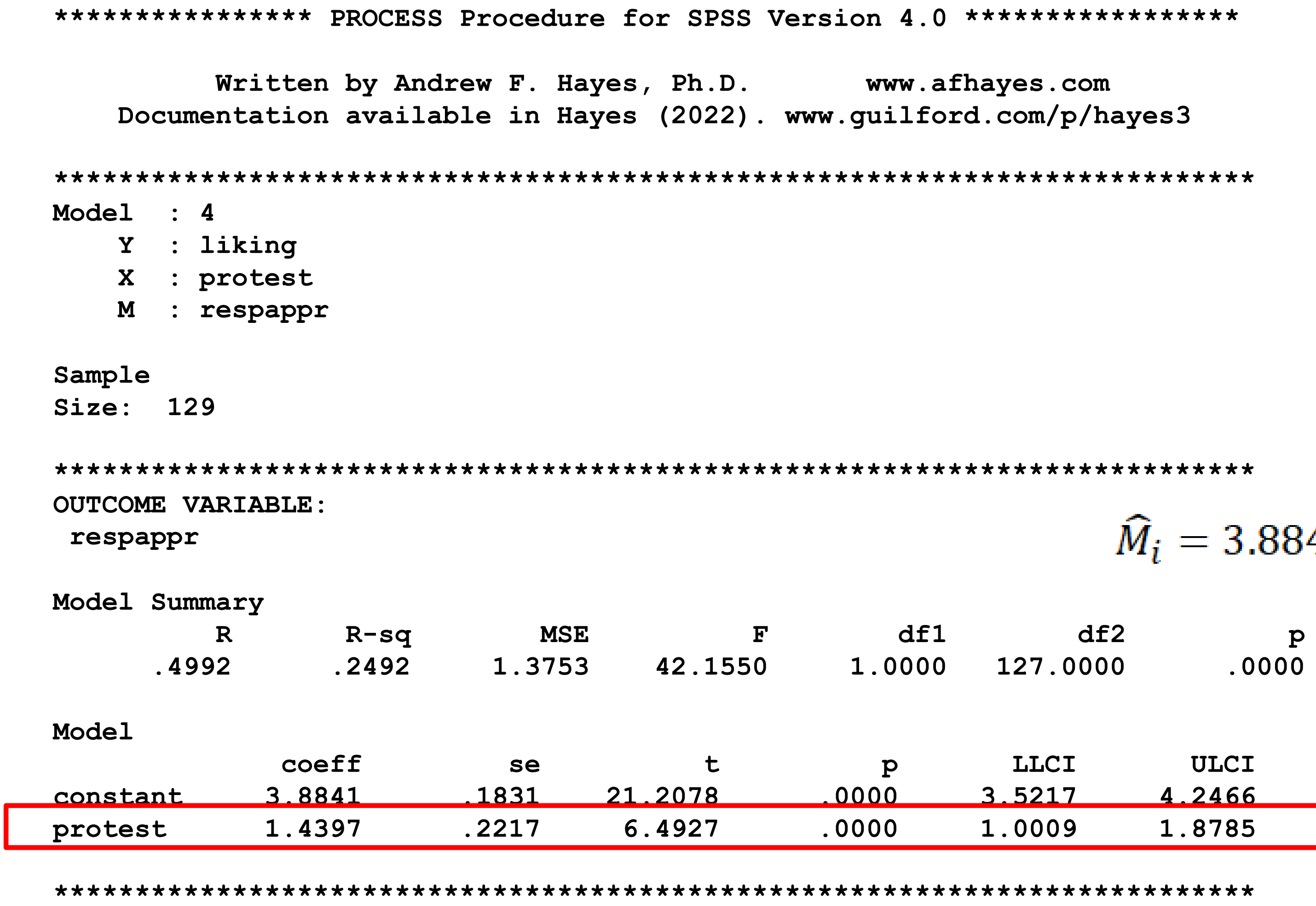
Red text not required.

```
process y=liking/x=protest/m=respappr/model=4/boot=10000/normal=1/total=1.
```

```
%process (data=protest, y=liking,x=protest,m=respappr,model=4,boot=10000,normal=1,total=1);
```

```
process(data=protest,y="liking",m="respappr",x="protest",model=4,boot=10000,normal=1,total=1)
```

PROCESS output



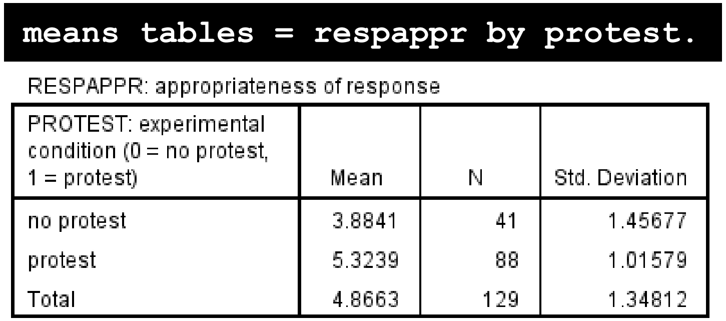
```
**************** PROCESS Procedure for SPSS Version 4.0 *****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2022). www.guilford.com/p/hayes3

**************************************************************************
Model  : 4
    Y  : liking
    X  : protest
    M  : respappr

Sample
Size:  129

**************************************************************************
OUTCOME VARIABLE:
 respappr

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4992      .2492     1.3753    42.1550     1.0000   127.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     3.8841      .1831    21.2078      .0000     3.5217     4.2466
protest      1.4397      .2217     6.4927      .0000     1.0009     1.8785

**************************************************************************
```

$$\widehat{M}_i = 3.884 + 1.440X_i$$

path *a*

Interpretation when *X* is dichotomous

$$\hat{M}_i = 3.884 + 1.440X_i$$



```
means tables = respappr by protest.
```

RESPAPPR: appropriateness of response

PROTEST: experimental condition (0 = no protest, 1 = protest)	Mean	N	Std. Deviation
no protest	3.8841	41	1.45677
protest	5.3239	88	1.01579
Total	4.8663	129	1.34812

When *X* = 1 (protest), $\hat{M}$ = 3.884 + 1.440(1) = 5.324
When *X* = 0 (no protest), $\hat{M}$ = 3.884 + 1.440(0) = 3.884

Notice that with X coded 0 and 1, the model yields the group means, *b* is the difference between the group means, and *the regression constant* is the mean for the group coded *X* = 0 (no protest condition).

More generally, if the two groups are coded by a difference of λ units, such that *X* = θ + λ for group 1 and X = θ for group 2,

$$b = (\bar{M}_1 - \bar{M}_2)/\lambda$$

If you get in the habit of coding a dichotomous variable such that the groups differ by one unit on *X*, *b* will always be a difference between group means.

PROCESS output

```
**************************************************************************
Outcome: liking
```

$$\widehat{Y}_i = 3.747 - 0.101X_i + 0.402M_i$$

```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4959      .2459      .8441    20.5483     2.0000   126.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     3.7473      .3058    12.2553      .0000     3.1422     4.3524
respappr      .4024      .0695     5.7884      .0000      .2648      .5400
protest      -.1007      .2005     -.5023      .6163     -.4975      .2960
```

path *b*

path *c'*

```
************************** TOTAL EFFECT MODEL ****************************
Outcome: liking
```

$$\widehat{Y}_i = 5.310 + 0.479X_i$$

```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .2131      .0454     1.0601     6.0439     1.0000   127.0000      .0153

Model
              coeff         se          t          p       LLCI       ULCI
constant     5.3102      .1608    33.0244      .0000     4.9921     5.6284
protest       .4786      .1947     2.4584      .0153      .0934      .8639
```

path *c*

```
**************************************************************************
```

Individual Coefficient Interpretations

$c = 0.479$ Because the two groups differ by one unit on X, this is the difference between the group means. The attorney was liked 0.479 units more when she protested than when she did not.

$a = 1.440$ Because the two groups differ by one unit on X, this is the difference between the group means. Protesting was seen as 1.44 units more appropriate for the situation than doing nothing.

$b = 0.402$ Holding constant what the attorney did, she was liked 0.402 more by those who saw her behavior as one unit more appropriate for the situation.

$c' = -0.100$ Because the two groups differ by one unit on X, -0.100 is the difference between the group means adjusted for differences between the groups in how appropriate her behavior was perceived as being for the situation (i.e., holding it constant)

PROCESS output

```
***************** TOTAL, DIRECT, AND INDIRECT EFFECTS ********************

Total effect of X on Y
     Effect         SE          t          p       LLCI       ULCI
      .4786      .1947     2.4584      .0153      .0934      .8639

Direct effect of X on Y
     Effect         SE          t          p       LLCI       ULCI
     -.1007      .2005     -.5023      .6163     -.4975      .2960

Indirect effect of X on Y
             Effect    Boot SE   BootLLCI   BootULCI
respappr      .5793      .1519      .3113      .9067

Normal theory tests for indirect effect
     Effect         se          Z          p
      .5793      .1350     4.2924      .0000
```

path *c*

path *c′*

ab with 95% bootstrap confidence interval

Sobel test

Her behavior was perceived as more appropriate if she protested relative to when she did not (*a* = 1.440), and the more appropriate her behavior, the more positively she was perceived (*b* = 0.402). Her choice to protest had a positive effect on how favorably she was perceived indirectly through perceived appropriateness of the response (point estimate: 0.579, 95% CI = 0.311 to 0.907). After accounting for this mechanism, there was no effect of her choice to protest on how she was evaluated (direct effect = -0.101, *p* = 0.62)

Interpretation of total, direct, and indirect effects

Generic

Total: Two people who differ by one unit on *X* are estimated to differ by *c* units on *Y* on average.

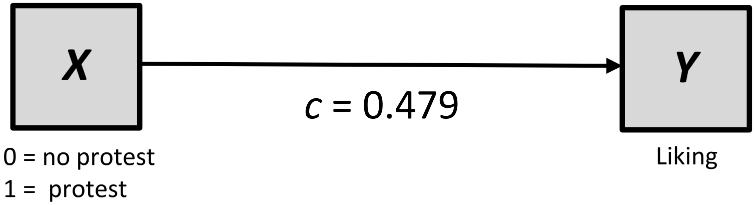
Indirect: They differ by *ab* units on average as a result of the effect of *X* on *M* which in turn affects *Y*.

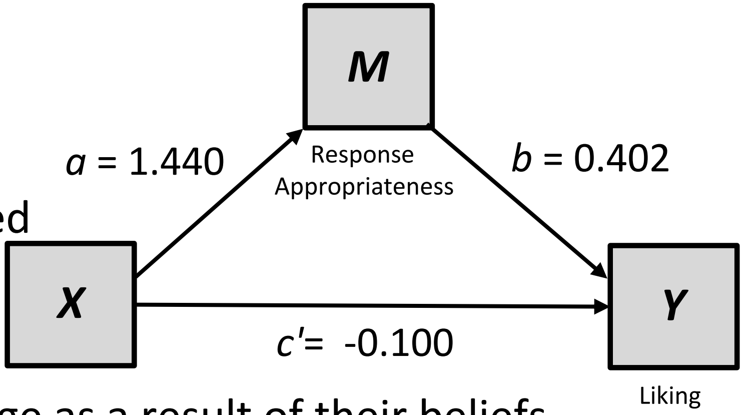
Direct: The rest of the difference, the difference of *c′* units, is due to the effect of *X* on *Y* independent of *M*.

Direct effect = -0.100

Indirect Effect = 1.440(0.402) = 0.579

Total effect = -0.100 + 0.579 = 0.479





Specific

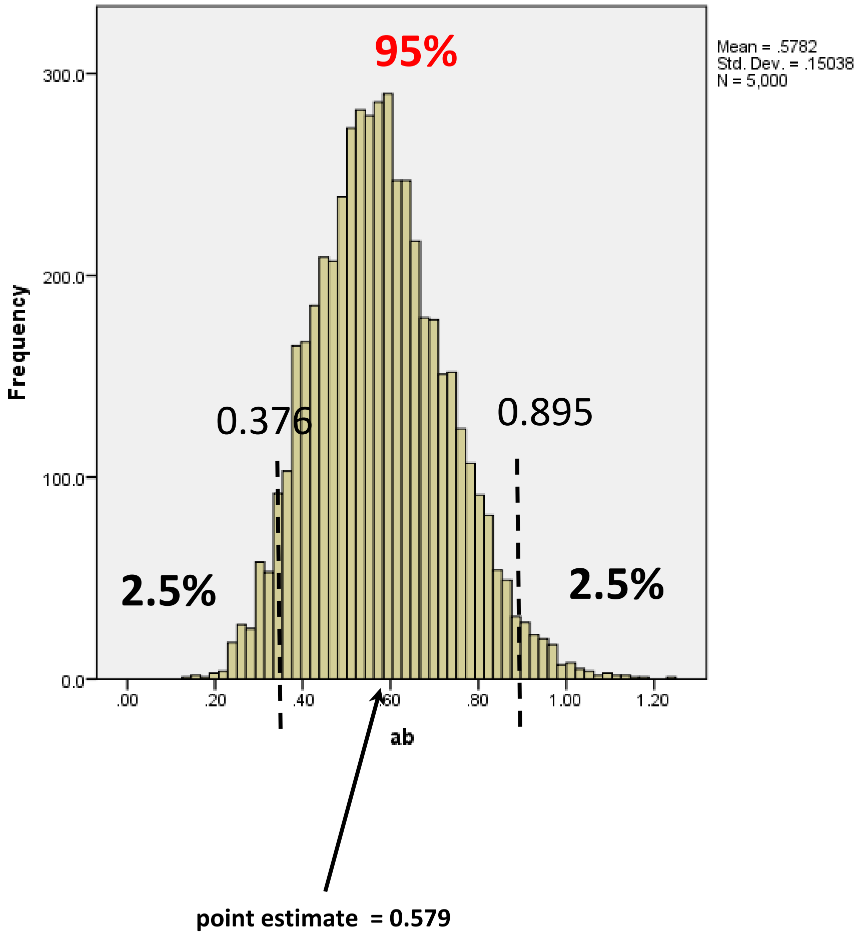
Total: Participants who were told the lawyer protested (*X* = 1) liked the lawyer 0.479 units **more**, on average, than those who were told she did not protest.

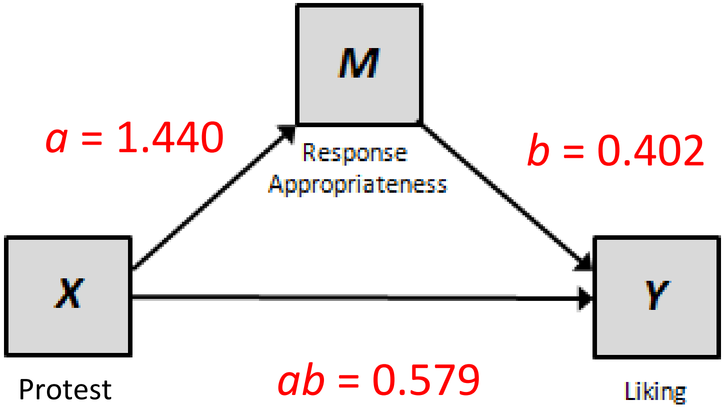
Indirect: They liked her by 0.579 units **more** on average as a result of their beliefs about the appropriateness of her response, which in turn affected their liking.

Direct: Among those equal in their beliefs about the appropriateness of her response, those who were told the lawyer protested liker her 0.100 units **less** (because the sign is negative) than those who were told she did not protest the decision.

5,000 bootstrap estimates of the indirect effect

95% of the 5,000 bootstrap estimates of the indirect effect were between 0.376 and 0.895. This is our 95% confidence interval.





Zero is not in the confidence interval, so we **can** claim an indirect effect different from zero with 95% confidence. This is akin to (though not exactly the same as) rejecting the null hypothesis of no indirect effect at the α = 0.05 level of significance.

Distribution of the Product

The distribution of the product method is similar to the MCCI. They are asymptotically equivalent.

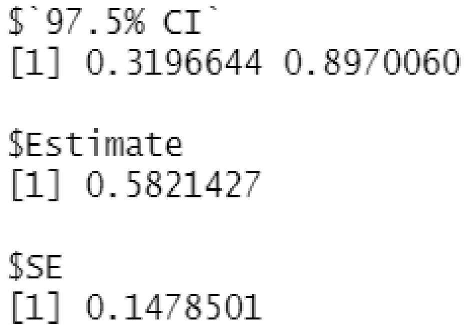
DP assumes that each regression coefficient has a normal distribution (just like MCCI)

DP derives the mathematical distribution of the product of the normal distribution, and generates p-values from that distribution

DP is not implemented in PROCESS, but is available in the RMediation package in R

```
medci(mu.x=1.44, mu.y=0.402, se.x=0.222,
se.y=0.069, rho=0.213,alpha=0.05,type="DOP")
```

```
medci(mu.x=1.44, mu.y=0.402, se.x=0.222,
se.y=0.069,rho=0.213,alpha=0.05,
type="prodclin")
```



```
$`97.5% CI`
[1] 0.3196644 0.8970060

$Estimate
[1] 0.5821427

$SE
[1] 0.1478501
```

Tofighi, D. and MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, **43**, 692–700. doi: 10.3758/s13428-011-0076-x

Confounding

Some effects in a mediation model are subject to 'confounding' even when *X* is based on random assignment, making causality harder to establish. Partialing out various confounders can help though won't solve the problem entirely.

$$\widehat{Y_i} = c_0 + cX_i + c_2U_i$$

$$\widehat{M_i} = a_0 + aX_i + a_2U_i$$

$$\widehat{Y_i} = c'_0 + c'X_i + bM_i + c'_2U_i$$

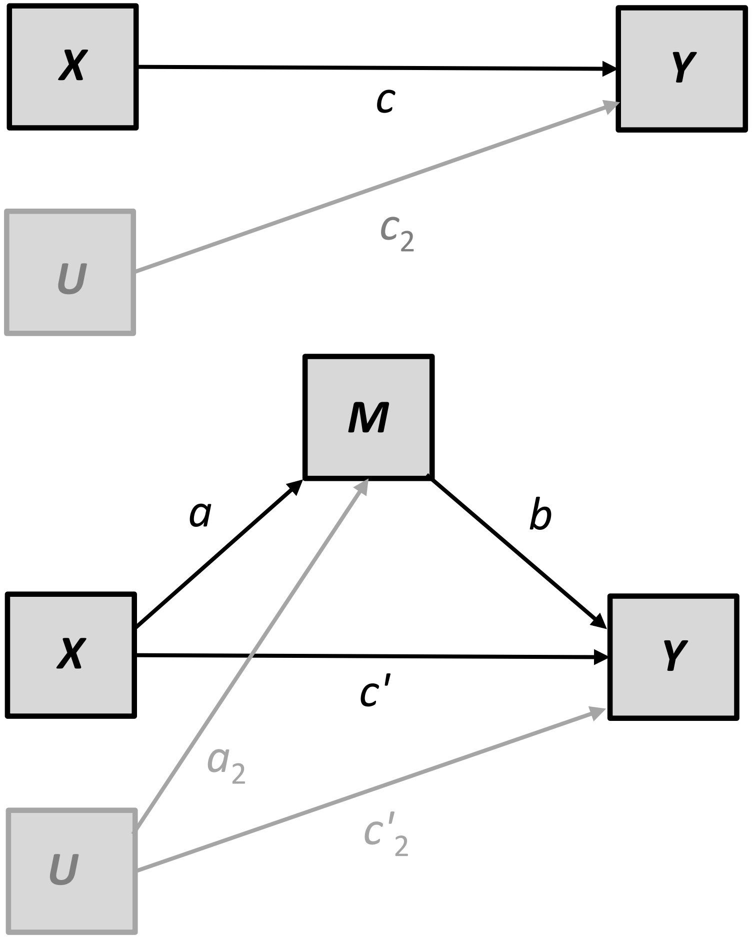
total effect = direct effect + indirect effect

$c = c' + (a \times b)$

indirect effect = total effect – direct effect

$a \times b = c - c'$

A simple mediation model, adjusting for a potential confounding variable (*U*)



Critiques of Statistical Mediation Analysis

Mediation Analysis is Very Popular

Mediation analysis is VERY popular in psychology (especially certain subfields)

- From 2005 to 2009, 59% of articles published in the Journal of Personality & Social Psychology and 65% in Personality & Social Psychology Bulletin included at least one mediation test (Rucker et al., 2011).
- In 2010 and 2011, the number of articles in a single issue of *Psychological Science* with such an analysis ranged between 1 and 6, with an average of 2.9 articles per issue. One cannot read an issue of *Psychological Science* without encountering at least one statistical mediation analysis. (Hayes & Scharkow, 2013)

The popularity of mediation analysis means that malpractices can have a large impact on research evidence in the field.

Concern about Cross-Sectional Studies

Concerns have been raise about the use of mediation analysis with cross-sectional data (in contrast to longitudinal or experimental)

Cross-sectional studies are studies which lack randomization to *X* and all measures are collected at the same time.

Experimental studies improve the causal inference of mediation (but are not sufficient for causal inference) because we can eliminate potential confounding on the *a*-path

Longitudinal studies improve causal inference of mediation (but are not sufficient for causal inference) because they can remove the possibility of reverse causation (if analyzed correctly).

Some journals have officially or unofficially banned mediation analysis with cross-sectional studies.

Rohrer JM, Hünermund P, Arslan RC, Elson M. That's a Lot to Process! Pitfalls of Popular Path Models. *Advances in Methods and Practices in Psychological Science*. 2022;5(2). doi:10.1177/25152459221095827

Making Causal Inference

Perhaps the gravest issue with mediation analysis is not the analysis itself but the way it is commonly interpreted.

Researchers make causal claims based on the results of mediation analysis, which is not founded.

These often manifest in the translation of the results to the discussion

- Using causal language in the discussion, as if causal direction is known
- Implying causality based on suggested future directions (e.g., implying that new interventions should be developed to intervene on the mediator)
- Recommending future studies that assume the causal direction in the model is correct, rather than further validation of the causal structure

These practices together have given mediation a bad name. Some recommend switching to "causal" mediation analysis frameworks.

I will argue this is not a "switch" as much as an integration of new techniques.

Causal Mediation Analysis

Counterfactual/Potential Outcomes Approach

A specific way of defining causal effects, which has become very popular is to consider specific questions using the idea of counterfactuals/potential outcomes.

If individuals with $X = x'$ instead had $X = x''$, how much would their value for Y have changed?

If the answer to this is zero, then X does not cause Y, but if it is non-zero, then X causes Y.

The outcome under each possible value of X ($Y_i(x')$ and $Y_i(x'')$) are called potential outcomes or counterfactuals.

We cannot observe multiple counterfactuals for each individual, and so we cannot directly examine the causal effect for each person.

Focal Estimates in Counterfactual Approach

Assume we have some kind of intervention with two levels (e.g., treatment ($X = 1$) and control ($X = 0$))

We can estimate the certain effects of interest…

ATE: Average Treatment Effect

If we consider that each person has a treatment effect:

$$Y_i(X = 1) - Y_i(X = 0)$$

If we take this value and average across all participants in the population that is the average treatment effect

$$ATE = E(Y_i(X = 1) - Y_i(X = 0))$$

If the population ATE is not zero, this would tell us that on average across the population there is an effect of the treatment on the outcome.

In an experimental study with random assignment to X, the mean difference on Y is an estimate of the ATE

Assumptions for Identification of ATE

Identification: defining the assumptions under which a specific estimand is truly estimating a causal effect

Consider our Protest data, the ATE of protesting on liking can be estimated by the difference in group means on liking. No assumptions are needed because we have random assignment to protest condition.

Consider the Harass data, the ATE of harassment on perceived academic failure could be estimated by the regression coefficient from the total effect model. But there are specific assumptions needed to this to be a causal effect.

> There are no variables which cause both harassment and perceived academic failure (no confounding)

In a model where we include potential confounders we must still assume that we have included **all possible confounders**

Focal Estimates in Counterfactual Approach

Knowing there is a treatment effect is useful, but also then begs some additional questions…
What is the mechanism? Why does X affect Y? How does X affect Y?

These questions have to do with **indirect effects** which are the focus on mediation analysis

> "What if the mediator was at the level we would expect under the other condition?"

The value of the outcome Y depends on both the value of X and M: $Y_i(X, M)$

The value of the mediator M depends on the value of X: $M_i(X)$

$$Y_i(X = x, M(X = x)) - Y_i(X = x, M(X = x'))$$

Indirect Effects

In the counterfactual approach there are two indirect effects through some mediator (M):

PNIE: Pure Natural Indirect Effect

$$Y_i(X = 0, M(X = 0)) - Y_i(X = 0, M(X = 1))$$

"Pure" is often used to refer to control conditions ($X = 0$)

The difference in the outcome if we could change each participant's mediator value to the value under the treatment, but everything else about them is consistent with the **control condition**.

TNIE: Total Natural Indirect Effect

$$Y_i(X = 1, M(X = 0)) - Y_i(X = 1, M(X = 1))$$

"Total" is used to refer to treatment conditions ($X = 1$)

The difference in the outcome if we could change each participant's mediator value to the value under the control, but everything else about them is consistent with the treatment condition.

PNIE with Linear Models

Let us assume the following generative models for *M* and *Y*

$$\widehat{M_i} = a_0 + aX_i$$

$$\widehat{Y_i} = c'_0 + c'_1X_i + bM_i$$

$$PNIE \;=\; Yi(X = 0, M(X = 0)) \;-\; Yi(X = 0, M(X = 1))$$

$$M(X = 0) = \; a_0 + \; a \times 0 \;=\; a_0$$
$$M(X \;=\; 1) = \; a_0 + \; a \times 1 \;=\; a_0 \;+\; a$$

$$Y\big(X \;=\; 0, M(X \;=\; 0)\big) = \; Y(X \;=\; 0, M \;=\; a_0)$$
$$= \; c'_0 + c' \times 0 \;+\; b \times (a_0)$$

$$Y\big(X \;=\; 0, M(X \;=\; 1)\big) = \; Y(X \;=\; 0, M \;=\; a_0 + a)$$
$$= \; c'_0 + c' \times 0 \;+\; b \times (a_0 + a)$$

$$Y(X \;=\; 0, M(X \;=\; 0)) - Y(X \;=\; 0, M(X \;=\; 1))$$
$$= \; (c'_0 + c' \times 0 \;+\; b \times (a_0)) - (c'_0 + c' \times 0 \;+\; b \times (a_0 + a)) \;=\; ab$$

TNIE with Linear Models

Let us assume the following generative models for *M* and *Y*

$$\widehat{M_i} = a_0 + aX_i$$

$$\widehat{Y_i} = c'_0 + c'_1 X_i + bM_i$$

$$TNIE = Yi(X = 1, M(X = 0)) - Yi(X = 1, M(X = 1))$$

$$M(X = 0) = a_0 + a \times 0 = a_0$$
$$M(X = 1) = a_0 + a \times 1 = a_0 + a$$

$$Y\big(X = 1, M(X = 0)\big) = Y(X = 1, M = a_0)$$
$$= c'_0 + c' \times 1 + b \times (a_0)$$

$$Y\big(X = 1, M(X = 1)\big) = Y(X = 1, M = a_0 + a)$$
$$= c'_0 + c' \times 1 + b \times (a_0 + a)$$

$$Y(X = 1, M(X = 0)) - Y(X = 1, M(X = 1))$$
$$= (c'_0 + c' \times 1 + b \times (a_0)) - (c'_0 + c' \times 1 + b \times (a_0 + a)) = ab$$

PNIE: The *XM* interaction

It is common in causal mediation analysis to allow the relationship between M and Y to differ by levels of X, by including an XM interaction in the equation for Y

$$\widehat{M_i} = a_0 + aX_i$$

$$\widehat{Y_i} = c'_0 + c_1'X_i + bM_i + c_2'X_iM_i$$

$$PNIE \; = \; Yi(X = 0, M(X = 0)) \; - \; Yi(X = 0, M(X = 1))$$

$$Y\big(X \; = \; 0, M(X \; = \; 0)\big) = \; Y(X \; = \; 0, M \; = \; a_0) = \; c_0' + c_1' \times 0 \; + \; b \times (a_0) + c_2'(0)(a_0)$$

$$Y\big(X \; = \; 0, M(X \; = \; 1)\big) = \; Y(X \; = \; 0, M \; = \; a_0 + a) = \; c_0' + c_1'(0) + \; b(a_0 + a) + c_2'(0)(a_0 + a)$$

$$Y(X \; = \; 0, M(X \; = \; 0)) - Y(X \; = \; 0, M(X \; = \; 1)) = \; (c_0' + c_1'(0) \; + \; b(a_0)) - (c_0' + c_1'(0) \; + \; b(a_0 + a)) \; = \; ab$$

TNIE: The *XM* interaction

It is common in causal mediation analysis to allow the relationship between M and Y to differ by levels of X, by including an XM interaction in the equation for Y

$$\widehat{M_i} = a_0 + aX_i$$

$$\widehat{Y_i} = c'_0 + c_1'X_i + bM_i + c_2'X_iM_i$$

$$TNIE \;=\; Yi(X=1, M(X=0)) \;-\; Yi(X=1, M(X=1))$$

$$Y\big(X \;=1, M(X \;=\; 0)\big) = \; Y(X \;=1, M \;=\; a_0) =\; c_0' + c_1'(1) \;+\; b(a_0) + c_2'(1)(a_0)$$

$$Y\big(X \;=1, M(X \;=\; 1)\big) = \; Y(X \;=1, M \;=\; a_0 + a) =\; c_0' + c_1'(1) +\; b(a_0 + a) + c_2'(1)(a_0 + a)$$

$$Y\big(X \;=1, M(X \;=\; 0)\big) - Y\big(X \;=1, M(X \;=\; 1)\big) =\; \big(c_0' + c_1'(1) +\; b(a_0)\big) - \big(c_0' + c_1'(1) +\; b(a_0 + a)\big) =\; ab + c_2'a$$

Direct Effects

In causal mediation analysis there are three types of direct effects.
Using the generative model with the interaction they are…

Pure Natural Direct Effect

The direct effect of *X* when the mediator is set to the value of the mediator predicted under the control condition

$Y\big(X = 1, M = M(0)\big) - Y(X = 0, M = M(0))$

Estimate: $c_1' + c_2' a_0$

Total Natural Direct Effect

The direct effect of *X* when the mediator is set to the value of the mediator predicted under the treatment condition

$Y\big(X = 1, M = M(1)\big) - Y(X = 0, M = M(1))$

Estimate: $c' + c_2' a_0 \ + a c_2'$

Controlled Direct Effect

The direct effect of *X* when the mediator is set to the overall mean of the mediator, without conditioning that mean on a value of X

$Y(X = 1, M = M_i) - Y(X = 0, M = M_i)$

Estimate: $c_1' + c_2' \overline{M}$

Direct and Indirect Effects

In statistical mediation analysis, the Total Effect = Direct + Indirect

In causal mediation analysis, we have two different indirect effects and two different direct effects, so what do we do?

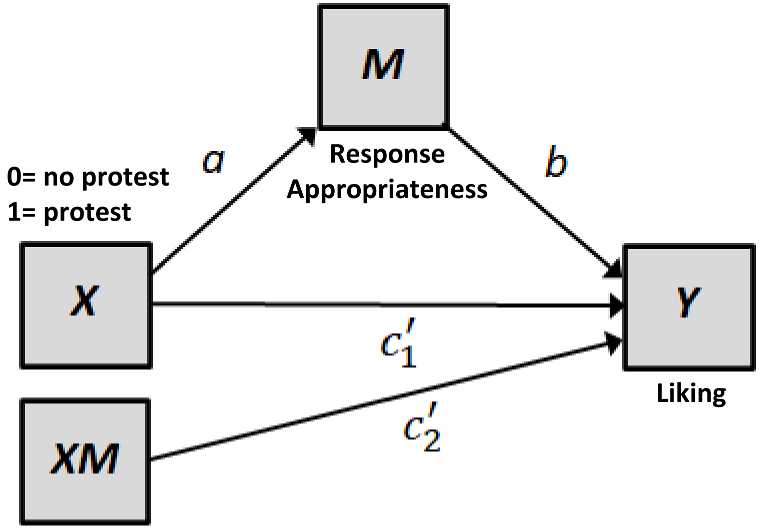
$$TE = TNIE + PNDE$$

$$TE = PNIE + TNDE$$

The total effect is partitioned into our total and pure natural effects, but if you use the total indirect it adds up with the pure direct effect and vice versa.

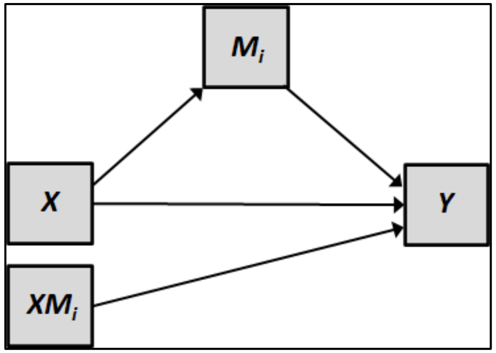
Estimation of the PROTEST model in PROCESS

PROCESS Model 4

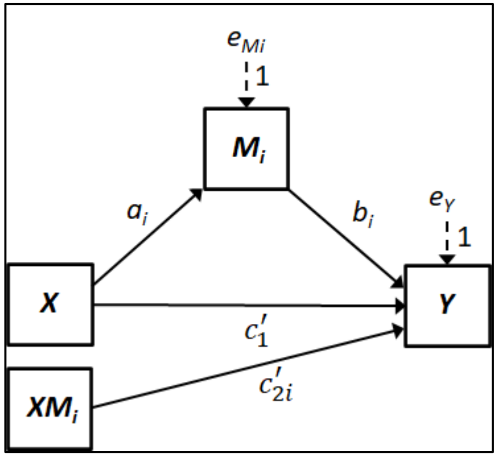


Model 4

Conceptual Diagram



Statistical Diagram



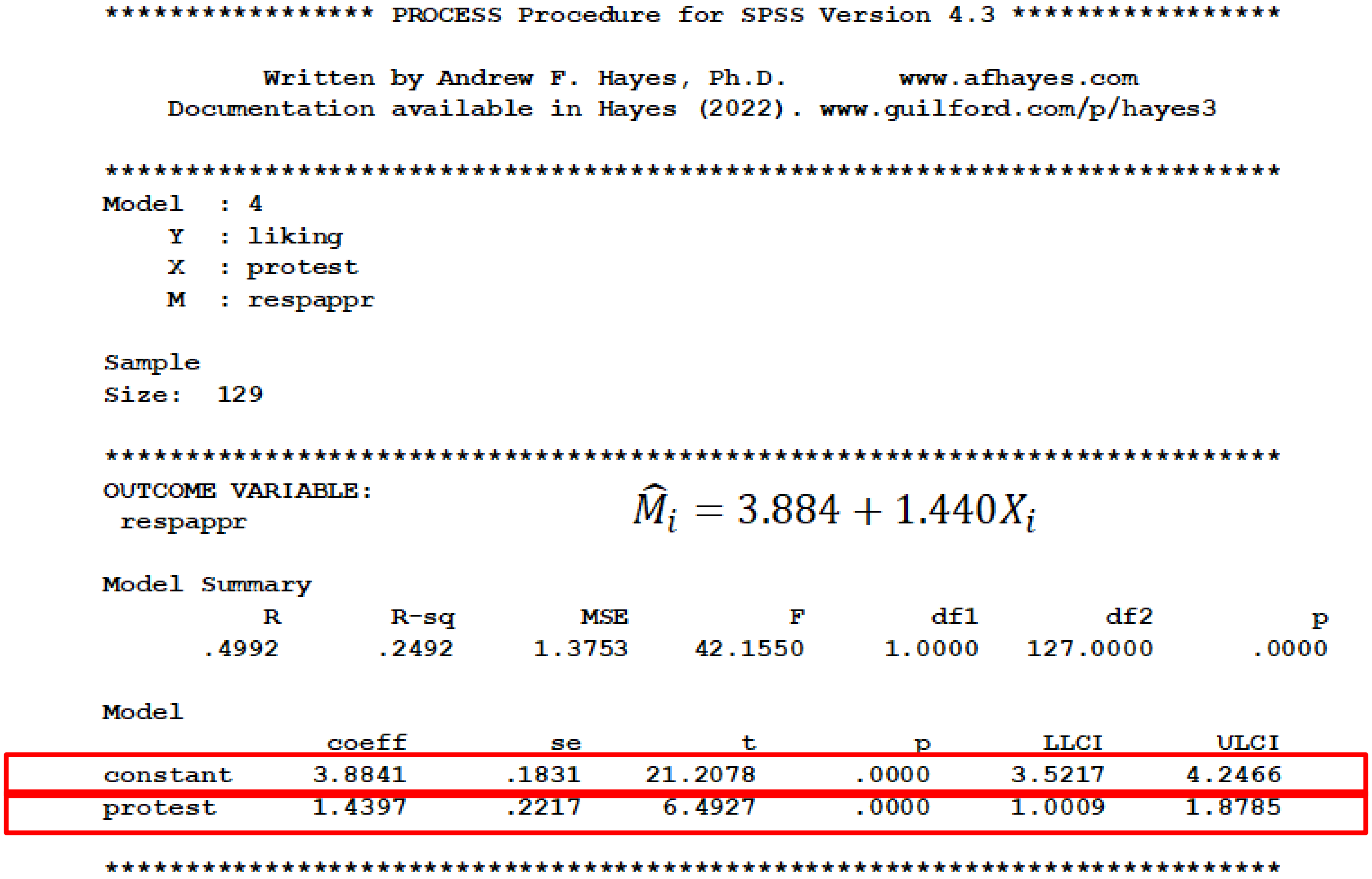
What should the PROCESS command look like?

```
process y=liking/x=protest/m=respappr/model=4/xmint=1/boot=10000/total=1.
```

```
%process (data=protest, y=liking,x=protest,m=respappr,model=4,xmint=1,boot=10000,total=1);
```

```
process(data=protest,y="liking",m="respappr",x="protest",model=4,xmint=1,boot=10000,total=1)
```

PROCESS output



```
***************** PROCESS Procedure for SPSS Version 4.3 *****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2022). www.guilford.com/p/hayes3

**************************************************************************
Model  : 4
    Y  : liking
    X  : protest
    M  : respappr

Sample
Size:  129

**************************************************************************
OUTCOME VARIABLE:
 respappr

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4992      .2492     1.3753    42.1550     1.0000   127.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     3.8841      .1831    21.2078      .0000     3.5217     4.2466
protest      1.4397      .2217     6.4927      .0000     1.0009     1.8785

**************************************************************************
```

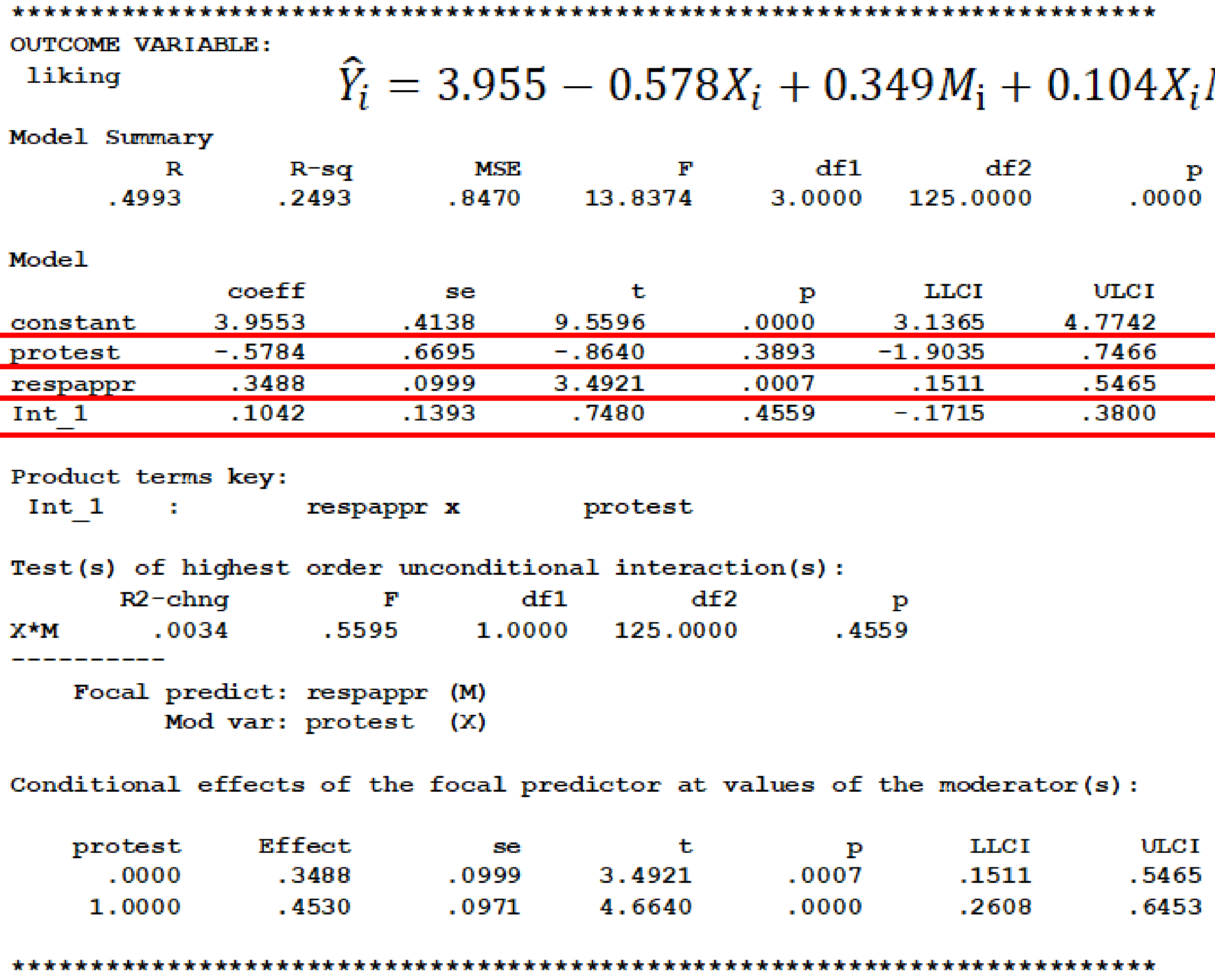
$$\widehat{M}_i = 3.884 + 1.440X_i$$

path a_0

path a

This output is identical to the previous version we ran!

PROCESS output



$$\hat{Y}_i = 3.955 - 0.578X_i + 0.349M_i + 0.104X_iM_i$$

```
**************************************************************************
OUTCOME VARIABLE:
 liking

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4993      .2493      .8470    13.8374     3.0000   125.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     3.9553      .4138     9.5596      .0000     3.1365     4.7742
protest      -.5784      .6695     -.8640      .3893    -1.9035      .7466
respappr      .3488      .0999     3.4921      .0007      .1511      .5465
Int_1         .1042      .1393      .7480      .4559     -.1715      .3800

Product terms key:
 Int_1    :        respappr x        protest

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*M      .0034      .5595     1.0000   125.0000      .4559
----------
    Focal predict: respappr (M)
          Mod var: protest  (X)

Conditional effects of the focal predictor at values of the moderator(s):

    protest     Effect         se          t          p       LLCI       ULCI
      .0000      .3488      .0999     3.4921      .0007      .1511      .5465
     1.0000      .4530      .0971     4.6640      .0000      .2608      .6453

**************************************************************************
```

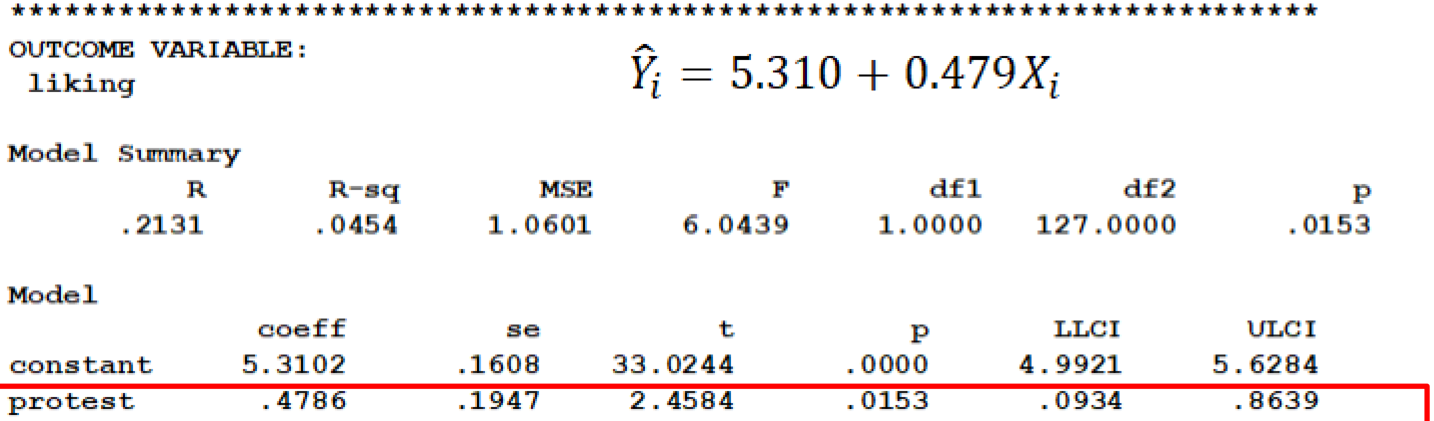
path c_1'

path b

path c_2'

The relationship between the mediator and the outcome in each condition.

PROCESS output



```
***************************************************************************
OUTCOME VARIABLE:
 liking

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .2131      .0454     1.0601     6.0439     1.0000   127.0000      .0153

Model
              coeff         se          t          p       LLCI       ULCI
constant     5.3102      .1608    33.0244      .0000     4.9921     5.6284
protest       .4786      .1947     2.4584      .0153      .0934      .8639
```

$$\hat{Y}_i = 5.310 + 0.479X_i$$

path *c*

Total effect is unchanged by conducting the analysis as “causal” vs. “statistical”.

Individual Coefficient Interpretations

$c = 0.479$ Because the two groups differ by one unit on X, this is the difference between the group means. The attorney was liked more when she protested than when she did not.

$a_0 = 3.884$ This is the group mean of the no protest group. The average appropriateness rating when the attorney did not protest was 3.884 units.

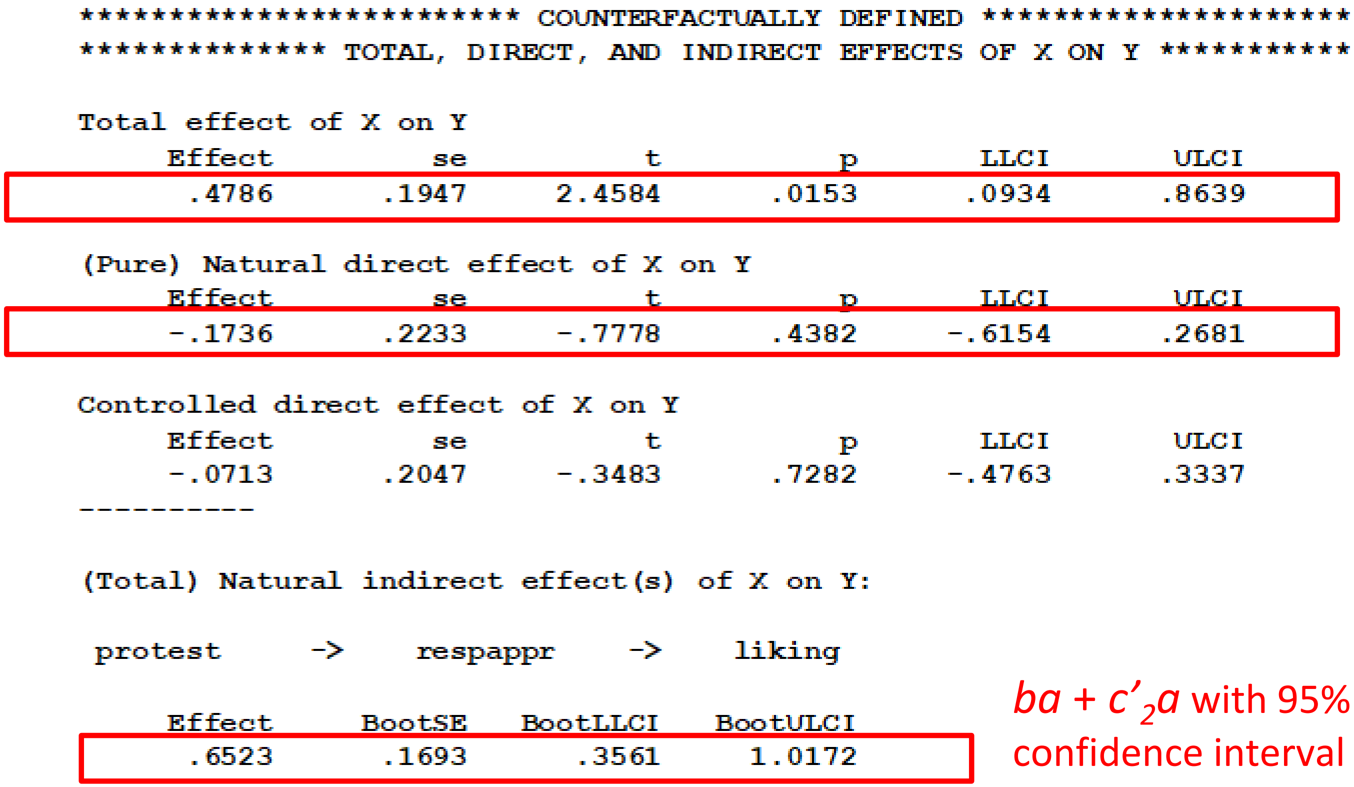
$a = 1.440$ Because the two groups differ by one unit on X, this is the difference between the group means. Protesting was seen as 1.44 units more appropriate for the situation than doing nothing.

$b = 0.349$ For those in the no protest condition, the attorney was liked more by those who saw her behavior as more appropriate for the situation.

$c'_1 = -0.578$ Because the two groups differ by one unit on X, this is the difference between the group means conditional on the appropriateness of her behavior being rated as 0.

$c'_2 = -0.104$ Because the two groups differ by one unit on X, this is the difference in the effect of response appropriateness on liking between the two experimental groups.

PROCESS output



```
************************** COUNTERFACTUALLY DEFINED *************************
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .4786      .1947     2.4584      .0153      .0934      .8639

(Pure) Natural direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
     -.1736      .2233     -.7778      .4382     -.6154      .2681

Controlled direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
     -.0713      .2047     -.3483      .7282     -.4763      .3337
----------

(Total) Natural indirect effect(s) of X on Y:

 protest      ->     respappr     ->     liking

     Effect     BootSE   BootLLCI   BootULCI
      .6523      .1693      .3561     1.0172
```

path *c*

$c'_1 + c'_2(a_0)$

$ba + c'_2a$ with 95% bootstrap confidence interval

Her behavior was perceived as more appropriate if she protested relative to when she did not (*a* = 1.440). When she did not protest, the more appropriate her behavior, the more positively she was perceived (*b* = 0.349). The effect of response appropriateness on liking was 0.104 units smaller if the attorney protested than if she did not (c'_2 = -0.104). If we could change each participant's response appropriateness value to the value they would have given under the no protest condition, but everything else about them is consistent with the protest condition, we would expect a positive effect on liking (point estimate: 0.652, 95% CI = 0.356 to 1.017). After accounting for this mechanism, there was no significant effect of her choice to protest on how she was evaluated (Pure Natural direct effect (PNDE) = -0.174, *p* = 0.27).

Estimating the PNIE and TDE in PROCESS

Note that in the previous output, only the TNIE and CDE are given.

To estimate the PNIE the value of *X* that is assigned to be the reference group can be switched (so instead of the no protest condition (*X* = 0) being the reference group, the protest condition (*X* = 1) must be).

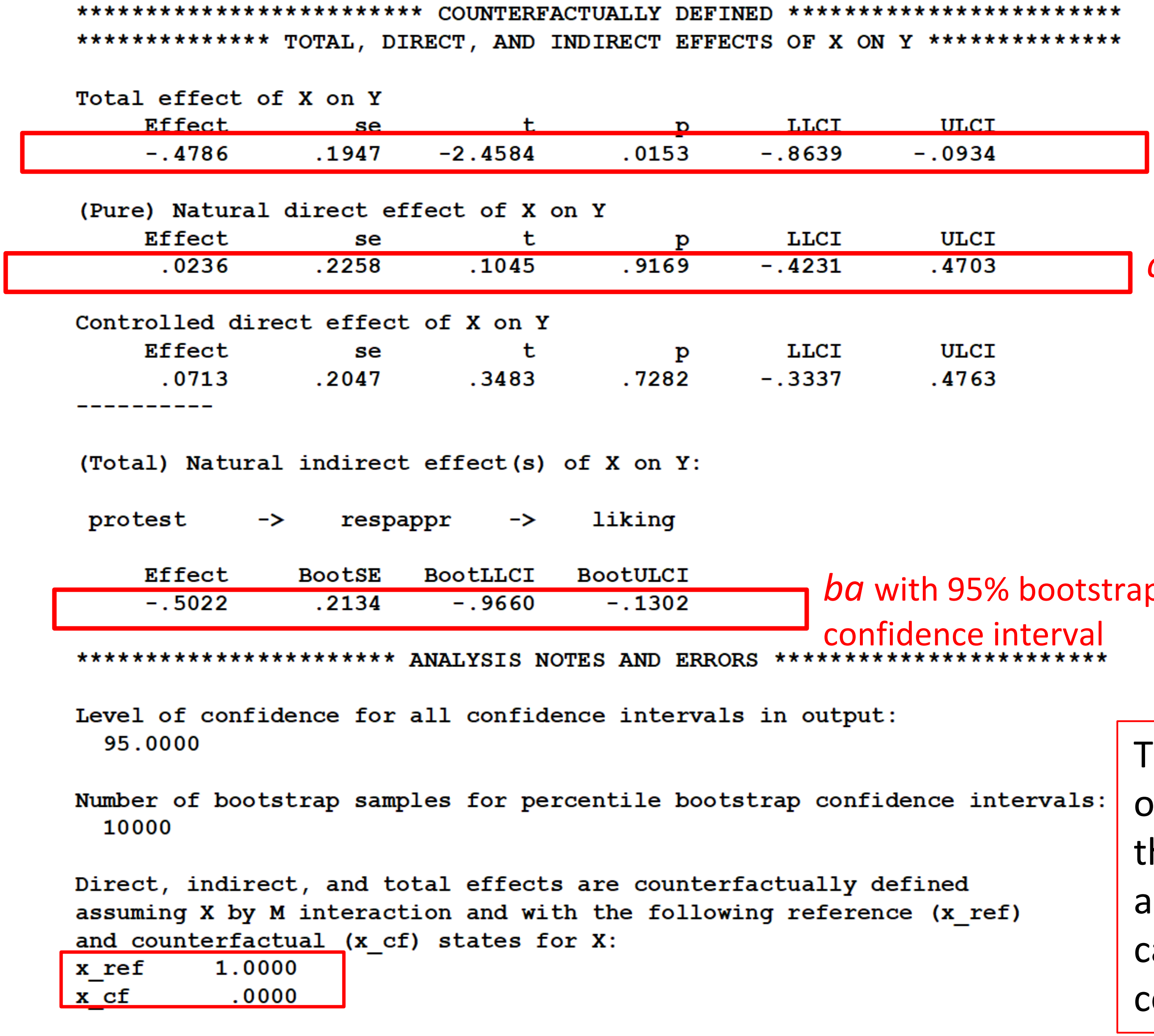
Within PROCESS, this can be done using the **xrefval** option:

```
process y=liking/x=protest/m=respappr/model=4/xmint=1/xrefval=1/boot=10000/total=1.
```

```
%process (data=protest,y=liking,x=protest,m=respappr,model=4,xmint=1,xrefval=1,boot=10000,total=1);
```

```
process(data=protest,y="liking",m="respappr",x="protest",model=4,xmint=1,xrefval=1,boot=10000,total=1)
```

PROCESS output



```
************************* COUNTERFACTUALLY DEFINED ************************
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI
     -.4786      .1947    -2.4584      .0153     -.8639     -.0934

(Pure) Natural direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0236      .2258      .1045      .9169     -.4231      .4703

Controlled direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0713      .2047      .3483      .7282     -.3337      .4763
----------

(Total) Natural indirect effect(s) of X on Y:

 protest     ->    respappr    ->    liking

     Effect     BootSE   BootLLCI   BootULCI
     -.5022      .2134     -.9660     -.1302

*********************** ANALYSIS NOTES AND ERRORS ************************

Level of confidence for all confidence intervals in output:
  95.0000

Number of bootstrap samples for percentile bootstrap confidence intervals:
  10000

Direct, indirect, and total effects are counterfactually defined
assuming X by M interaction and with the following reference (x_ref)
and counterfactual (x_cf) states for X:
x_ref      1.0000
x_cf        .0000
```

path *c*

$c'_1 + c'_2(a + a_0)$

ba with 95% bootstrap confidence interval

The PNDE and TNIE displayed in this output are the TNDE and PNIE of our original model with the signs flipped.

The end of the PROCESS output lets us know what the reference group is assigned to be (in this case, the protest condition, *X* = 1).

Interpretation of total, direct, and indirect effects

Generic

Total: Two people who differ by one unit on *X* are estimated to differ by *c* units on *Y* on average.

TNIE: We would expect a $ba + c'_2a$ unit difference in *Y* if we could change each participant's *M* value to the value under the treatment, but everything else about them is consistent with the treatment condition.

PNIE: We would expect a ba unit difference in the outcome if we could change each participant's *M* value to the value under the treatment, but everything else about them is consistent with the control condition.

TDE: We would expect a $c'_1 + c'_2(a + a_0)$ unit difference in the outcome if we could change each participant to the treatment, but their mediator value is unchanged and aligned with the treatment group.

PNDE: We would expect a $c'_1 + c'_2(a_0)$ unit difference in the outcome if we could change each participant to the treatment, but their mediator value is unchanged and aligned with the control group.

PNDE = -0.578 + 0.104(3.884) = -0.174

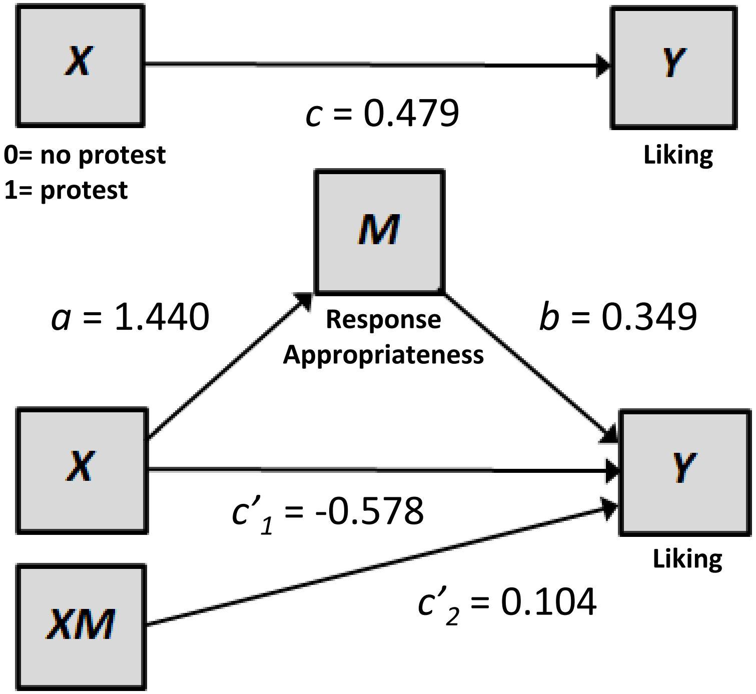
TNIE = 0.349(1.440) + 0.104(1.440) = 0.652

Total effect = -0.174 + 0.652 = 0.478

TDE = -0.578 + 0.104(3.884 + 1.440) = -0.024

PNIE = 0.349(1.440) = 0.503

Total effect = -0.024 + 0.503 = 0.479



Assumptions for Identification

Pearl (2001) outlined four assumptions under which the PNIE and TNIE are **identified** (meaning they provide an unbiased causal estimate).

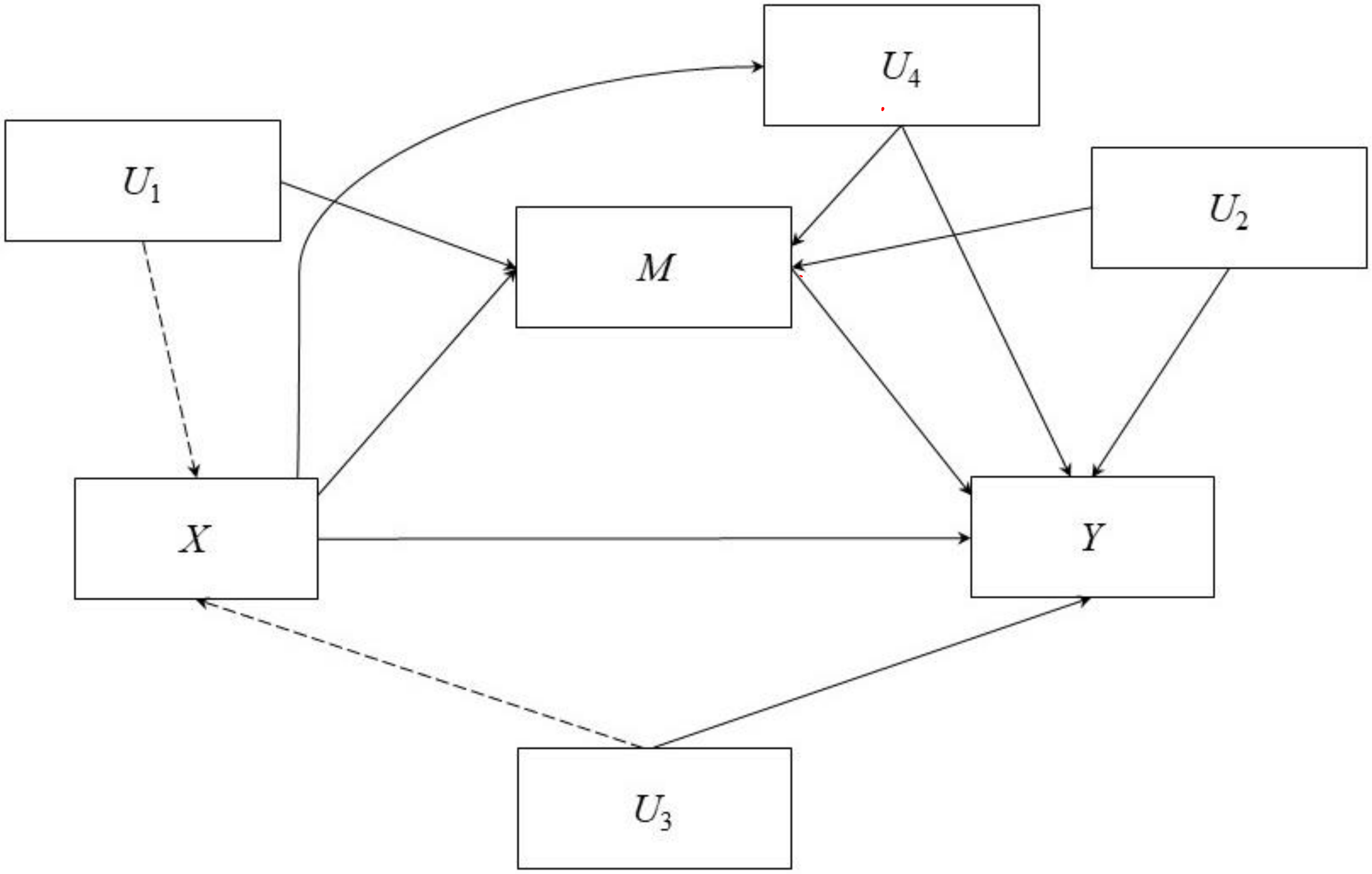
1. No unmeasured confounders of the effect of *X* on *Y* conditioned on covariates
2. No unmeasured confounders of the effect of *M* on *Y* conditioned on covariates and *X*
3. No unmeasured confounders of the effect of *X* on *M* conditioned on covariates
4. No measured or unmeasured confounders of the effect of *M* on *Y* which are affected by *X* conditioned on covariates.

Under these assumptions, even if you have observational/correlational data, causal claims are warranted.

The trouble is these assumptions are untestable, and we can never know when they are met.

Assumptions for Identification

1. No unmeasured confounders of the effect of *X* on *Y* conditioned on covariates (U1)
2. No unmeasured confounders of the effect of *M* on *Y* conditioned on covariates and *X* (U2)
3. No unmeasured confounders of the effect of *X* on *M* conditioned on covariates (U3)
4. No measured or unmeasured confounders of the effect of *M* on *Y* which are affected by *X* conditioned on covariates. (U4)



Assumptions under Randomized *X*

If *X* is randomized, some of these assumptions are met by design, but others remain.

1. ~~No unmeasured confounders of the effect of *X* on *Y* conditioned on covariates~~
2. No unmeasured confounders of the effect of *M* on *Y* conditioned on covariates and *X*
3. ~~No unmeasured confounders of the effect of *X* on *M* conditioned on covariates~~
4. No measured or unmeasured confounders of the effect of *M* on *Y* which are affected by *X* conditioned on covariates.

Sensitivity Analysis

It can be helpful to examine how sensitive your results are to violations of assumptions through sensitivity analyses. Unfortunately, this can not be done within PROCESS.

However, the **medsens** function in the **mediation** package can be used to test the robustness of your results to violations of the sequential ignorability assumption.

This function calculates the average causal mediation effect for varying values of the correlation between the residuals from the *M* and *Y* models (called rho).

```
model.m = lm(respappr ~ protest, data = protest)
model.y = lm(liking ~ protest + respappr, data = protest)
med.model = mediate(model.m=model.m,
model.y=model.y,sims=10000,boot=TRUE,boot.ci.type="perc",
                 treat="protest",mediator="respappr")
sens.analysis = medsens(x=med.model, rho.by = 0.05, effect.type =
"indirect")
```

Sensitivity Analysis

```
summary(med.model)
```

```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

                         Estimate 95% CI Lower 95% CI Upper p-value    
ACME (control)             0.5022       0.1232         0.96  0.0094 ** 
ACME (treated)             0.6523       0.3532         1.01  <2e-16 ***
ADE (control)             -0.1898      -0.6630         0.33  0.4810    
ADE (treated)             -0.0398      -0.4493         0.42  0.8926    
Total Effect               0.4624       0.0521         0.93  0.0262 *  
Prop. Mediated (control)   1.0860       0.2164         3.91  0.0344 *  
Prop. Mediated (treated)   1.4104       0.5228         6.45  0.0262 *  
ACME (average)             0.5772       0.3021         0.90  <2e-16 ***
ADE (average)             -0.1148      -0.4873         0.31  0.6138    
Prop. Mediated (average)   1.2482       0.5366         5.27  0.0262 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 129


Simulations: 10000
```

PNIE

TNIE

Sensitivity Analysis

```
summary(sens.analysis)
```

```
Mediation Sensitivity Analysis: Average Mediation Effect

Sensitivity Region: ACME for Control Group

      Rho ACME(control) 95% CI Lower 95% CI Upper R^2_M*R^2_Y* R^2_M~R^2_Y~
[1,] 0.20        0.2734      -0.0170       0.5638       0.0400       0.0225
[2,] 0.25        0.2128      -0.0730       0.4986       0.0625       0.0352
[3,] 0.30        0.1497      -0.1325       0.4319       0.0900       0.0507
[4,] 0.35        0.0834      -0.1963       0.3631       0.1225       0.0690
[5,] 0.40        0.0130      -0.2656       0.2916       0.1600       0.0902
[6,] 0.45       -0.0626      -0.3418       0.2166       0.2025       0.1141
[7,] 0.50       -0.1450      -0.4269       0.1370       0.2500       0.1409
[8,] 0.55       -0.2360      -0.5234       0.0515       0.3025       0.1705

Rho at which ACME for Control Group = 0: 0.4
R^2_M*R^2_Y* at which ACME for Control Group = 0: 0.16
R^2_M~R^2_Y~ at which ACME for Control Group = 0: 0.0902
```

More information available about the output in Tingley et al., 2014 (included in course materials)

Sensitivity Analysis

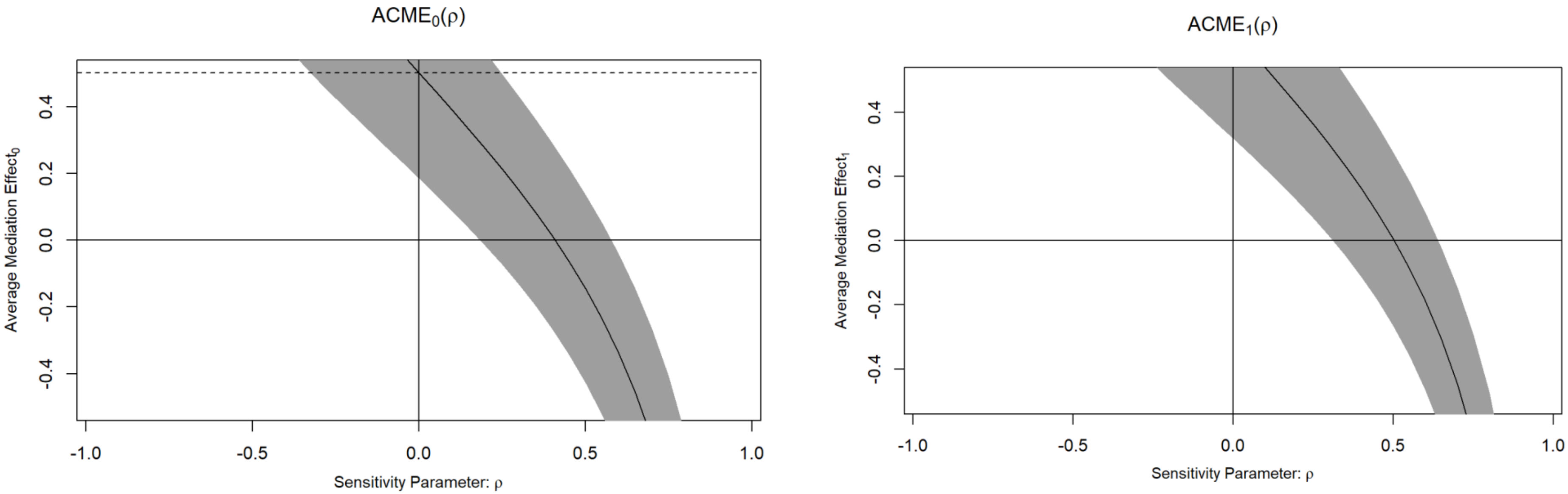
```
summary(sens.analysis)
```

```
Sensitivity Region: ACME for Treatment Group

      Rho ACME(treated) 95% CI Lower 95% CI Upper R^2_M*R^2_Y* R^2_M~R^2_Y~
[1,] 0.35        0.2334      -0.0464       0.5133       0.1225       0.0690
[2,] 0.40        0.1630      -0.1122       0.4383       0.1600       0.0902
[3,] 0.45        0.0874      -0.1847       0.3596       0.2025       0.1141
[4,] 0.50        0.0051      -0.2658       0.2760       0.2500       0.1409
[5,] 0.55       -0.0859      -0.3580       0.1862       0.3025       0.1705
[6,] 0.60       -0.1884      -0.4652       0.0883       0.3600       0.2029

Rho at which ACME for Treatment Group = 0: 0.5
R^2_M*R^2_Y* at which ACME for Treatment Group = 0: 0.25
R^2_M~R^2_Y~ at which ACME for Treatment Group = 0: 0.1409
```

More information available about the output in Tingley et al., 2014 (included in course materials)

Sensitivity Analysis

```
plot(sens.analysis, sens.par = "rho", ylim = c(-0.5, 0.5))
```



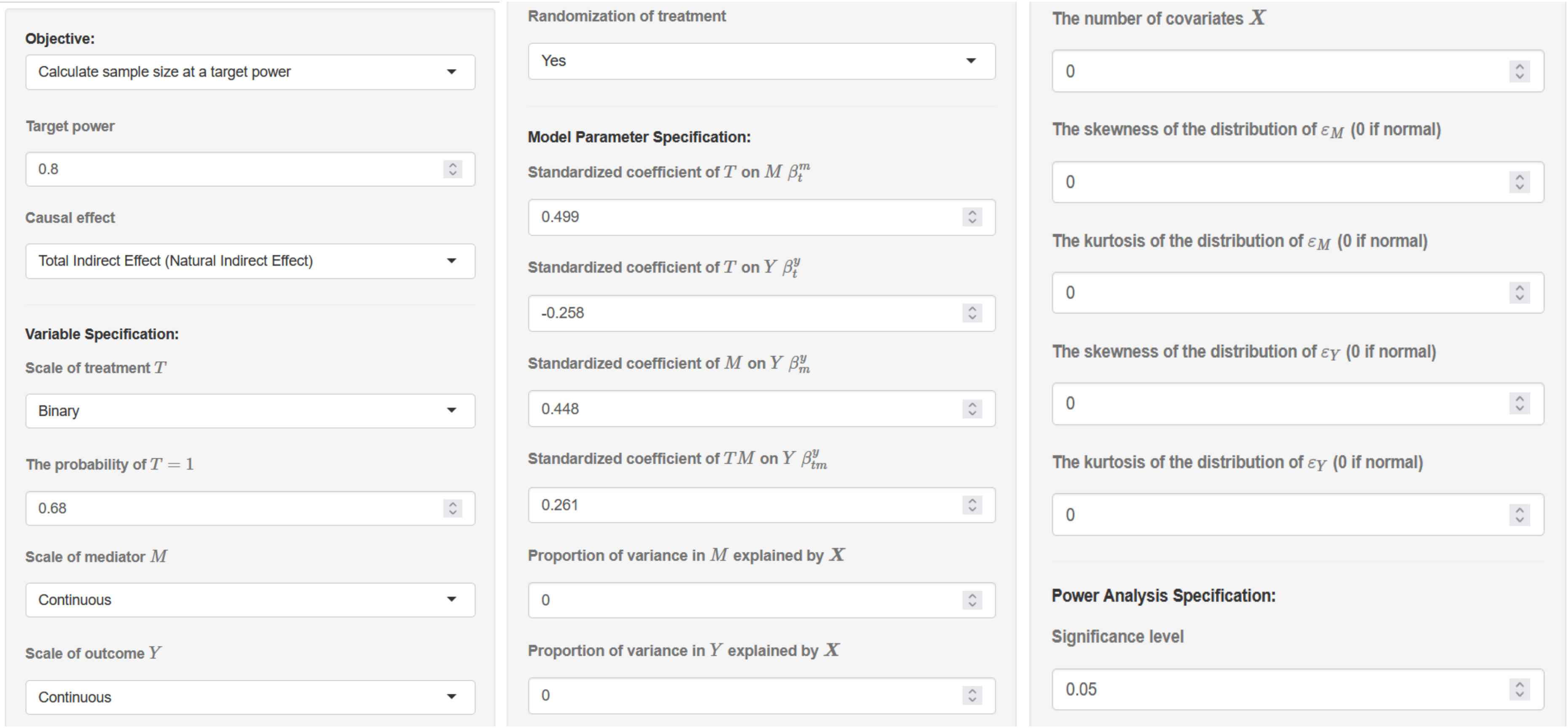
More information available about the output in Tingley et al., 2014 (included in course materials)

Power Analysis Examples

Power Analysis for Causal Mediation Analysis Shiny app

https://xuqin.shinyapps.io/CausalMediationPowerAnalysis/

Assume we want to replicate the protest study while estimating the *XM* interaction, what sample size should we use to have 80% to detect an effect?



Objective:

Calculate sample size at a target power

Target power

0.8

Causal effect

Total Indirect Effect (Natural Indirect Effect)

Variable Specification:

Scale of treatment T

Binary

The probability of $T = 1$

0.68

Scale of mediator M

Continuous

Scale of outcome Y

Continuous

Randomization of treatment

Yes

Model Parameter Specification:

Standardized coefficient of T on M β_t^m

0.499

Standardized coefficient of T on Y β_t^y

-0.258

Standardized coefficient of M on Y β_m^y

0.448

Standardized coefficient of TM on Y β_{tm}^y

0.261

Proportion of variance in M explained by $\boldsymbol{X}$

0

Proportion of variance in Y explained by $\boldsymbol{X}$

0

The number of covariates $\boldsymbol{X}$

0

The skewness of the distribution of ε_M (0 if normal)

0

The kurtosis of the distribution of ε_M (0 if normal)

0

The skewness of the distribution of ε_Y (0 if normal)

0

The kurtosis of the distribution of ε_Y (0 if normal)

0

Power Analysis Specification:

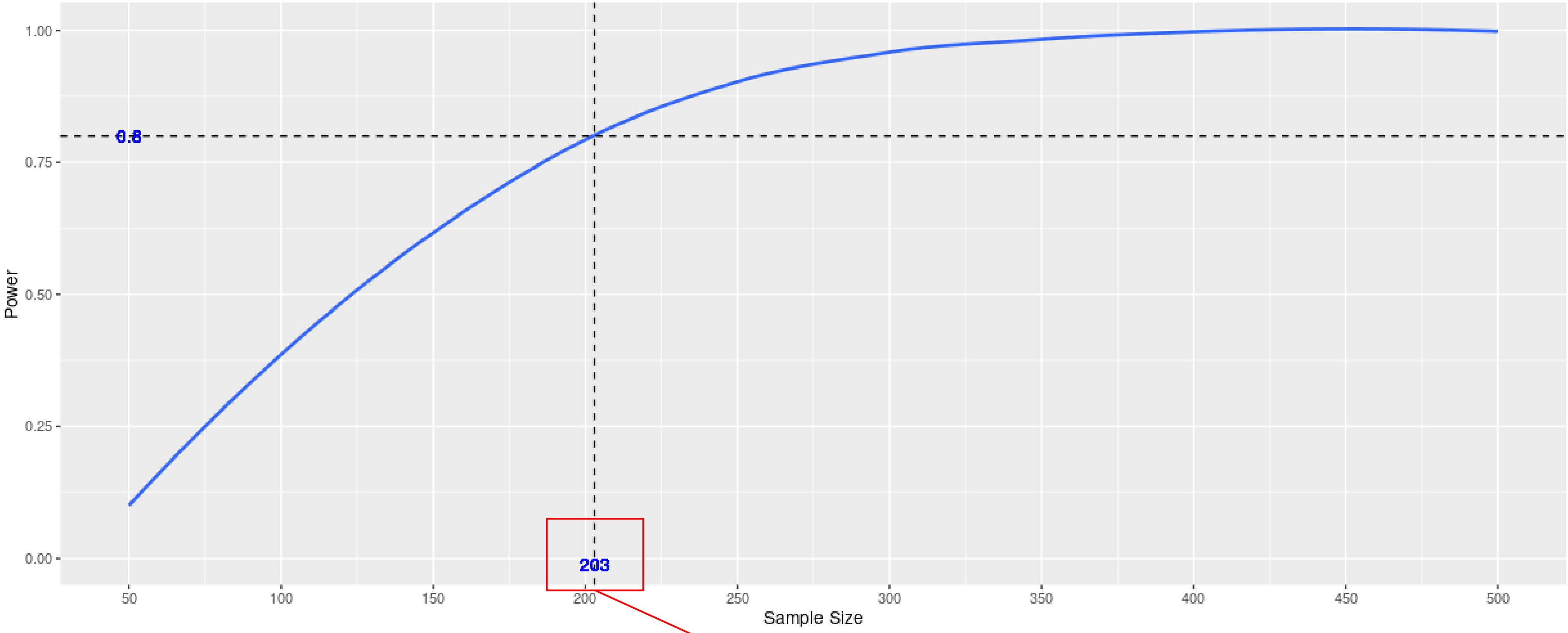
Significance level

0.05

Power Analysis Examples

Power Analysis for Causal Mediation Analysis Shiny app

Output:



It appears we will need a sample size of about 203 to have 80% power.

Testing the *XM* interaction

There are three potential approaches to testing the *XM* interaction that you might take:

1. **Include the *XM* interaction by default**
   - This is what is recommended in causal mediation analysis (Vo et al., 2020)
   - Use the xmint option in PROCESS always
   - Need to differentiate hypotheses about PNIE and TNIE
2. **Test/Evaluation *XM* interaction as a pre-step**
   - Estimate coefficient and use a hypothesis test or effect size measure to evaluate whether it should be included
   - Include your threshold in your preregistration
   - Final model will depend on whether *XM* is included or not
   - May need to differentiate hypotheses about PNIE and TNIE
3. **Robustness Check**
   - Fit the model as hypothesized (*XM* in or out)
   - Afterwards fit the other model and evaluate whether the results are sensitive to the choice (change in effect size or significance level)

Covariates and Sensitivity Analysis

When planning your study you should consider potential confounders, either adjust the design to account for these or measure them.

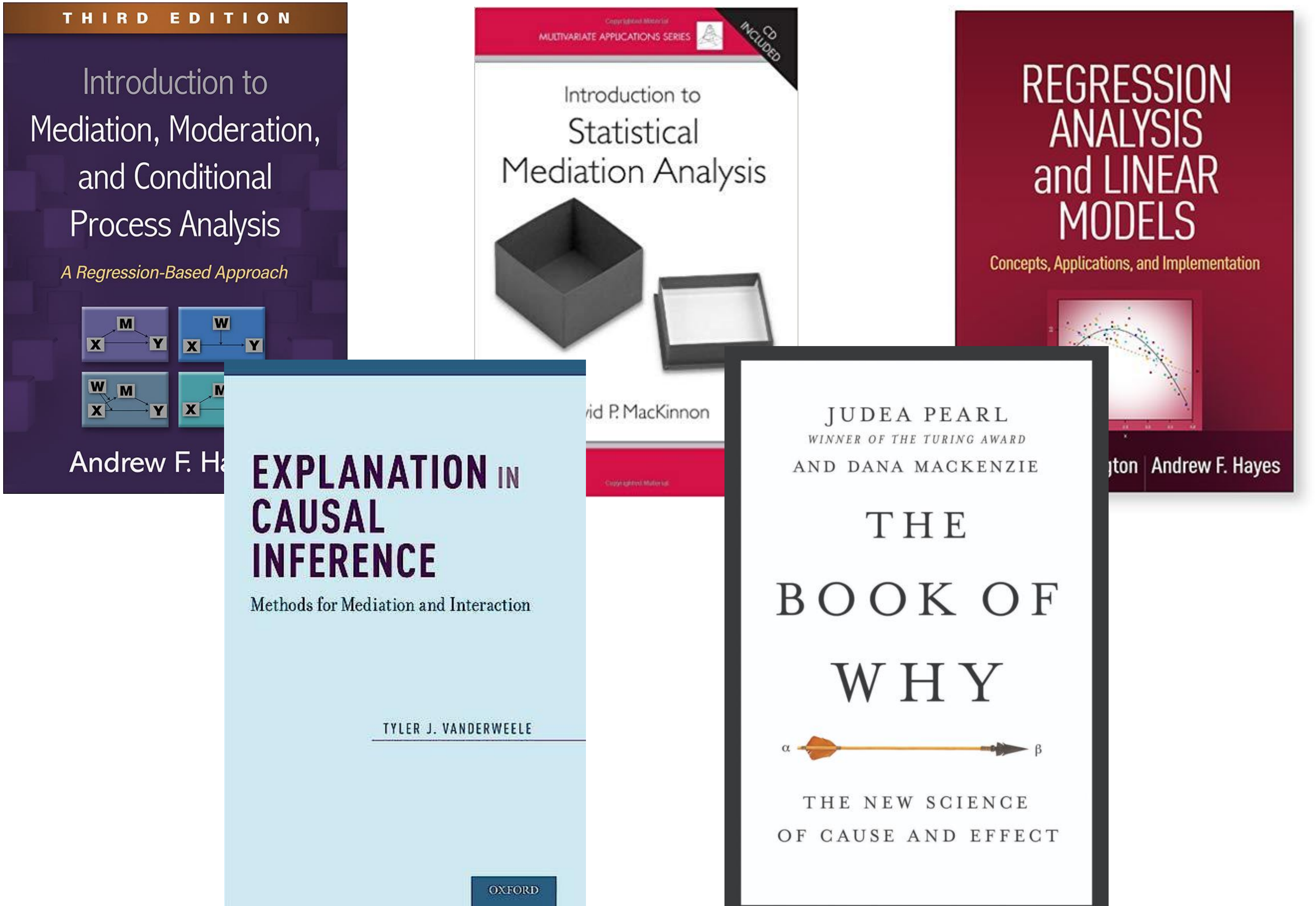
In your analysis include appropriate confounders/covariates.

Conduct sensitivity analysis (even if you included potential confounders)

Preregister the degree of confounding you would consider acceptable

Where to learn more

Chapters 13, 14, and 15


THIRD EDITION
Introduction to Mediation, Moderation, and Conditional Process Analysis
A Regression-Based Approach
M
X
Y
W
Andrew F. Ha
MULTIVARIATE APPLICATIONS SERIES
CD INCLUDED
Introduction to Statistical Mediation Analysis
vid P. MacKinnon
REGRESSION ANALYSIS and LINEAR MODELS
Concepts, Applications, and Implementation
gton
Andrew F. Hayes
EXPLANATION IN CAUSAL INFERENCE
Methods for Mediation and Interaction
TYLER J. VANDERWEELE
OXFORD
JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE
THE BOOK OF WHY
α
β
THE NEW SCIENCE OF CAUSE AND EFFECT