Structural Equation Modeling & Mediation

Introduction to SEM with Lavaan

Kyle M. Lang

Department of Methodology & Statistics
Utrecht University

Outline

Structural Equation Modeling

Mediation

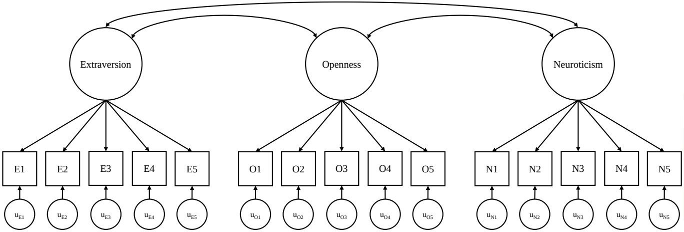
- Simple Mediation
- Bootstrapping

Full SEM

Structural equation modeling (SEM) simply combines path analysis and CFA.

- SEM allows us to model complicated structural relations among latent variables.

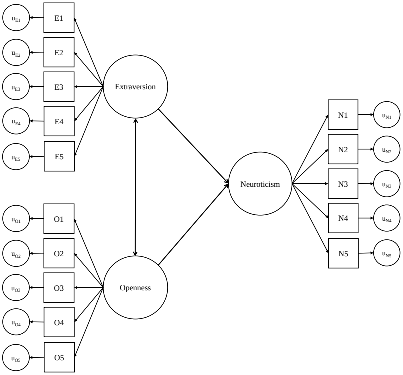
Let's consider a simple, three-factor CFA model.



CFA → SEM

We first evaluate the validity of the measurement model via CFA.

- We then convert the CFA to an SEM by converting some covariances to latent regression paths.



Why SEM?

The beauty of SEM is that we get to model the types of complex relations we can specify via path models while leveraging all the strengths of latent variables.

- When we fit a multiple-group SEM, we are modeling moderation by group.
  - The latent variables give us the ability to evaluate measurement invariance across groups.
  - We'll see more of these ideas in the next lecture.
- Path analysis and SEM lend themselves especially well to mediation analysis and conditional process anlaysis.

Mediation

Mediation vs. Moderation

What do we mean by *mediation* and *moderation*?

Mediation vs. Moderation

What do we mean by *mediation* and *moderation*?

Mediation and moderation are types of hypotheses, not statistical methods or models.

- Mediation tells us *how* one variable influences another.
- Moderation tells us *when* one variable influences another.

Contextualizing Example

Say we wish to explore the process underlying exercise habits.

Our first task is to operationalize “exercise habits”

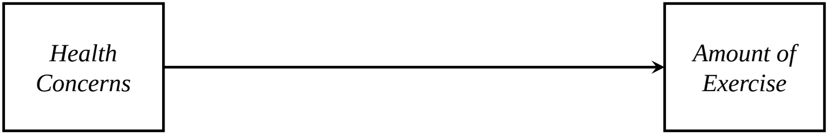
- DV: Hours per week spent in vigorous exercise (*exerciseAmount*).

We may initial ask: what predicts devoting more time to exercise?

- IV: Concerns about negative health outcomes (*healthConcerns*).

Focal Effect Only

The *healthConcerns* → *exerciseAmount* relation is our *focal effect*

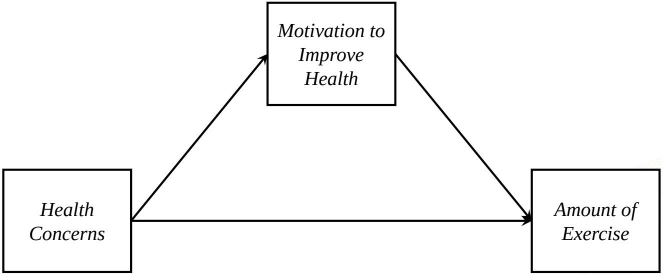


- Mediation, moderation, and conditional process analysis all attempt to describe the focal effect in more detail.
- We always begin by hypothesizing a focal effect.

The Mediation Hypothesis

A mediation analysis will attempt to describe how health concerns affect amount of exercise.

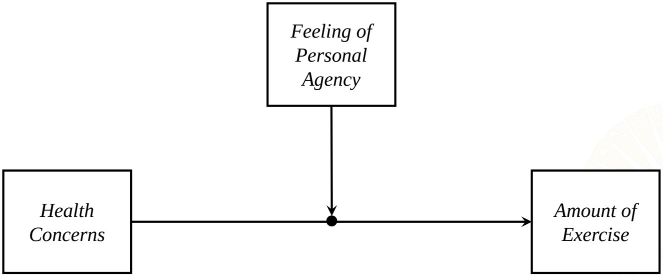
- The *how* is operationalized in terms of intermediary variables.
- Mediator: Motivation to improve health (*motivation*).



Moderation Hypothesis

A moderation hypothesis will attempt to describe when health concerns affect amount of exercise.

- The *when* is operationalized in terms of interactions between the focal predictor and contextualizing variables
- Moderator: Sense of personal agency relating to physical health (*agency*).



Conditional Process Analysis

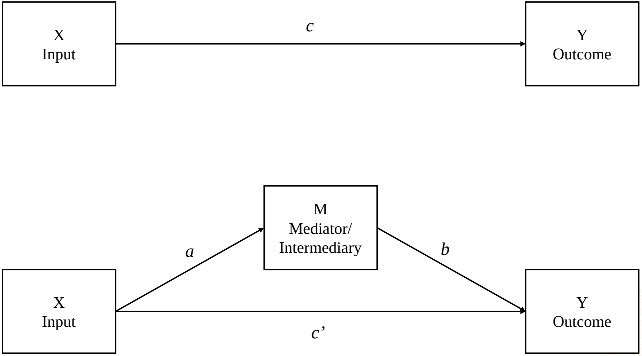
Conditional process analysis combines the mediation and moderation hypotheses into models of moderated mediation.

- Given a mediation model describing *how* health concerns affect exercise amount, what other variables may modulate the indirect effect.

Excellent Resource

Plug Hayes’ book

Path Diagrams



Necessary Equations

To get all the pieces of the preceding diagram using OLS regression, we'll need to fit three seperate models.

$$Y = i_1 + cX + e_1 \tag{1}$$
$$Y = i_2 + c'X + bM + e_2 \tag{2}$$
$$M = i_3 + aX + e_3 \tag{3}$$

- Equation 1 gives us the total effect (c).
- Equation 2 gives us the direct effect (c') and the partialled effect of the mediator on the outcome (b).
- Equation 3 gives us the effect of the input on the outcome (a).

Two Measures of Indirect Effect

Indirect effects can be quantified in two different ways:

$$IE_{diff} = c - c' \tag{4}$$

$$IE_{prod} = a \cdot b \tag{5}$$

IE_{diff} and IE_{prod} are equivalent in simple mediation.

- Both give us information about the proportion of the total effect that is transmitted through the intermediary variable.
- IE_{prod} provides a more direct representation of the actual pathway we're interested in testing.
- IE_{diff} gets at our desired hypothesis indirectly.

The Causal Steps Approach

Baron and Kenny (1986, p. 1176) describe three/four conditions as being sufficient to demonstrate statistical "mediation."

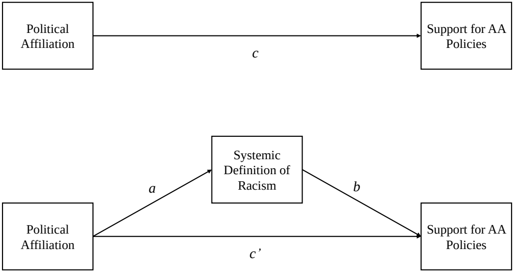
1. Variations in levels of the independent variable significantly account for variations in the presumed mediator (i.e., Path *a*).
   - Need a significant a path.

2. Variations in the mediator significantly account for variations in the dependent variable (i.e., Path *b*).
   - Need a significant b path.

3. When Paths *a* and *b* are controlled, a previously significant relation between the independent and dependent variables is no longer significant.
   - Need a significant total effect
   - The direct effect must be "less" than the total effect

Example Process Model

Consider the following process.



Causal Steps Example

```r
## Load some data:
dat1 <- readRDS("../data/adamsKlpsScaleScore.rds")

## Check pre-conditions:
mod1 <- lm(policy ~ polAffil, data = dat1)
mod2 <- lm(policy ~ sysRac, data = dat1)
mod3 <- lm(sysRac ~ polAffil, data = dat1)

## Partial out the mediator's effect:
mod4 <- lm(policy ~ sysRac + polAffil, data = dat1)
```

Causal Steps Example

```
summary(mod1)


Call:
lm(formula = policy ~ polAffil, data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7357 -0.8254  0.0643  0.6827  3.2481

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.71516    0.35648   7.617 3.32e-11 ***
polAffil     0.23675    0.07775   3.045   0.0031 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.134 on 85 degrees of freedom
Multiple R-squared:  0.09836,Adjusted R-squared:  0.08775
F-statistic: 9.273 on 1 and 85 DF,  p-value: 0.003096
```

Causal Steps Example

```
summary(mod2)


Call:
lm(formula = policy ~ sysRac, data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7700 -0.5593  0.0255  0.6277  3.6835

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9295     0.3896   2.386   0.0193 *
sysRac        0.7557     0.1014   7.450 7.14e-11 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9286 on 85 degrees of freedom
Multiple R-squared:  0.395,Adjusted R-squared:  0.3879
F-statistic:  55.5 on 1 and 85 DF,  p-value: 7.145e-11
```

Causal Steps Example

```
summary(mod3)


Call:
lm(formula = sysRac ~ polAffil, data = dat1)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.44714 -0.50502  0.05286  0.54498  2.25286 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.60605    0.28489   9.147 2.72e-14 ***
polAffil     0.25685    0.06213   4.134 8.34e-05 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.906 on 85 degrees of freedom
Multiple R-squared:  0.1674,Adjusted R-squared:  0.1576 
F-statistic: 17.09 on 1 and 85 DF,  p-value: 8.336e-05
```

Causal Steps Example

```
summary(mod4)


Call:
lm(formula = policy ~ sysRac + polAffil, data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7156 -0.6043  0.0262  0.6474  3.7992

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.83266    0.41246   2.019   0.0467 *
sysRac       0.72236    0.11148   6.480 5.93e-09 ***
polAffil     0.05121    0.06998   0.732   0.4663
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9312 on 84 degrees of freedom
Multiple R-squared:  0.3989,Adjusted R-squared:  0.3845
F-statistic: 27.87 on 2 and 84 DF,  p-value: 5.211e-10
```

Causal Steps Example

```r
## Extract important parameter estimates:
a      <- coef(mod3)["polAffil"]
b      <- coef(mod4)["sysRac"]
c      <- coef(mod1)["polAffil"]
cPrime <- coef(mod4)["polAffil"]

## Compute indirect effects:
ieDiff <- unname(c - cPrime)
ieProd <- unname(a * b)

ieDiff

[1] 0.1855374

ieProd

[1] 0.1855374
```

Sobel's Z

In the previous example, do we have a *significant* indirect effect?

- The direct effect is "substantially" smaller than the total effect, but is the difference statistically significant?
- Sobel (1982) developed an asymptotic standard error for IE_{prod} that we can use to assess this hypothesis.

$$SE_{sobel} = \sqrt{a^2 \cdot SE_b^2 + b^2 \cdot SE_a^2} \tag{6}$$

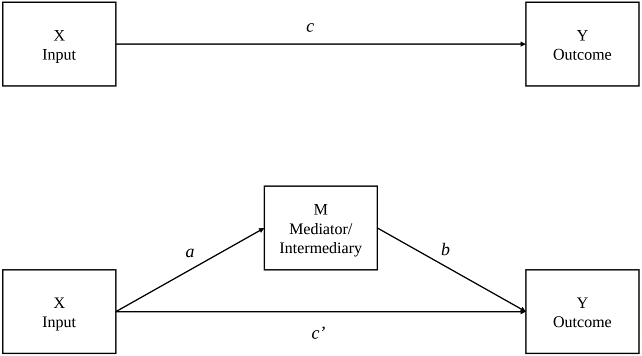
$$Z_{sobel} = \frac{ab}{SE_{sobel}} \tag{7}$$

$$95\%CI_{sobel} = ab \pm 1.96 \cdot SE_{sobel} \tag{8}$$

Sobel Example

```
## SE:
seA <- (mod3 %>% vcov() %>% diag() %>% sqrt())["polAffil"]
seB <- (mod4 %>% vcov() %>% diag() %>% sqrt())["sysRac"]

se <- sqrt(b^2 * seA^2 + a^2 * seB^2) %>% unname()

## z-score:
(z <- ieProd / se)

[1] 3.48501

## p-value:
(p <- 2 * pnorm(z, lower = FALSE))

[1] 0.0004921178

## 95% CI:
c(ieProd - 1.96 * se, ieProd + 1.96 * se)

[1] 0.08118957 0.28988525
```

Recall our Basic Path Diagram



Two Measures of Indirect Effect

Recall the two definitions of an indirect effect:

$$IE_{diff} = c - c' \tag{9}$$

$$IE_{prod} = a \cdot b \tag{10}$$

It pays to remember a few key points:

- IE_{diff} and IE_{prod} are equivalent in simple mediation.
- IE_{diff} is only an indirect indication of IE_{prod}.
- A significant indirect effect can exist without a significant total effect.
- If we only care about the indirect effect, then we don't need to worry about the total effect.

Two Measures of Indirect Effect

Recall the two definitions of an indirect effect:

$$IE_{diff} = c - c' \tag{9}$$

$$IE_{prod} = a \cdot b \tag{10}$$

It pays to remember a few key points:

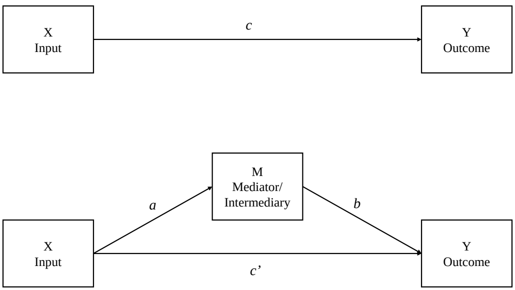
- IE_{diff} and IE_{prod} are equivalent in simple mediation.
- IE_{diff} is only an indirect indication of IE_{prod}.
- A significant indirect effect can exist without a significant total effect.
- If we only care about the indirect effect, then we don't need to worry about the total effect.

These points imply something interesting:

- We don't need to estimate c!

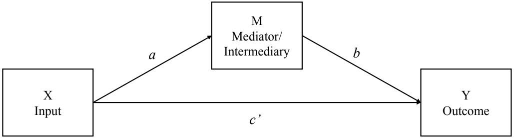
Simplifying our Path Diagram

Question: If we don't care about directly estimating c, how can we simplify this diagram?

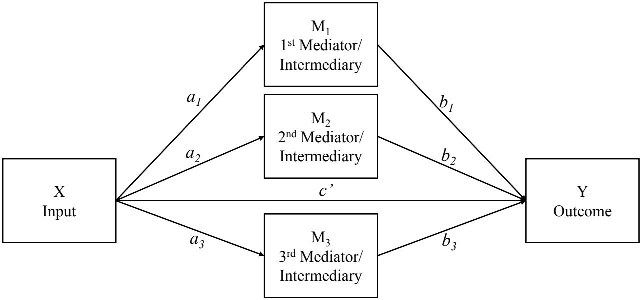


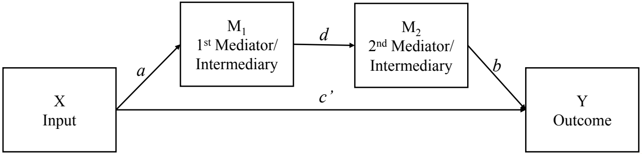
Simplifying our Path Diagram

Answer: We don't fit the upper model.



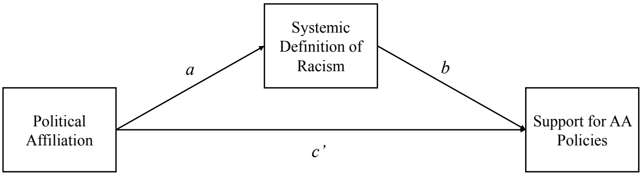
Why Path Analysis?





Example

Let's revisit the above example using path analysis in **lavaan**.



Example

```
## Load the lavaan package:
library(lavaan)

## Specify the basic path model:
mod1 <- "
policy ~ sysRac + polAffil
sysRac ~ polAffil
"

## Estimate the model:
out1 <- sem(mod1, data = dat1)
```

Example

```
## Look at the results:
partSummary(out1, c(5, 6))

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  policy ~
    sysRac            0.722    0.110    6.595    0.000
    polAffil          0.051    0.069    0.745    0.456
  sysRac ~
    polAffil          0.257    0.061    4.182    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.837    0.127    6.595    0.000
   .sysRac            0.802    0.122    6.595    0.000
```

Example

```
## Include the indirect effect:
mod2 <- "
policy ~ b*sysRac + polAffil
sysRac ~ a*polAffil

ab := a*b # Define a parameter for the indirect effect
"

## Estimate the model:
out2 <- sem(mod2, data = dat1)
```

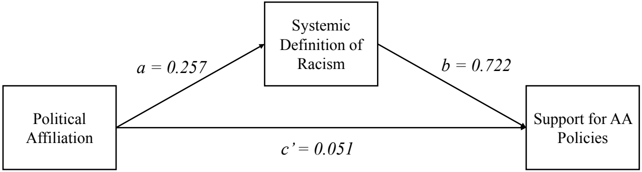
Example

```
## Look at the results:
partSummary(out2, 5:7)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  policy ~
    sysRac     (b)    0.722    0.110    6.595    0.000
    polAffil          0.051    0.069    0.745    0.456
  sysRac ~
    polAffil   (a)    0.257    0.061    4.182    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.837    0.127    6.595    0.000
   .sysRac            0.802    0.122    6.595    0.000

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ab                0.186    0.053    3.532    0.000
```

Example

```r
## We can also get CIs:
parameterEstimates(out2, zstat = FALSE, pvalue = FALSE, ci = TRUE)

       lhs op      rhs label   est    se ci.lower ci.upper
1   policy  ~   sysRac     b 0.722 0.110    0.508    0.937
2   policy  ~ polAffil       0.051 0.069   -0.084    0.186
3   sysRac  ~ polAffil     a 0.257 0.061    0.136    0.377
4   policy ~~   policy       0.837 0.127    0.588    1.086
5   sysRac ~~   sysRac       0.802 0.122    0.564    1.040
6 polAffil ~~ polAffil       2.444 0.000    2.444    2.444
7       ab :=      a*b    ab 0.186 0.053    0.083    0.289
```

Results



Systemic Definition of Racism
a = 0.257
b = 0.722
Political Affiliation
Support for AA Policies
c' = 0.051

We're not there yet...

Path analysis allows us to directly model complex (and simple) relations, but the preceding example still suffers from a considerable limitation.

- The significance test for the indirect effect is still conducted with the Sobel Z approach.

Path analysis (or full SEM) doesn't magically get around distributional problems associated with Sobel's Z test.

- To get a robust significance test of the indirect effect, we need to use *bootstrapping*.

Bootstrapping

Bootstrapping was introduced by Efron (1979) as a tool for non-parametric inference.

- Traditional inference requires that we assume a parametric sampling distribution for our focal parameter.
- We need to make such an assumption to compute the standard errors we require for inferences.
- If we cannot safely make these assumptions, we can use bootstrapping.

Bootstrapping

Assume our observed data $Data_0$ represent the population and:

1. Sample rows of $Data_0$, with replacement, to create B new samples $\{Data_b\}$.
2. Calculate our focal statistic on each of the B bootstrap samples.
3. Make inferences based on the empirical distribution of the B estimates calculated in Step 2

Bootstrapping

$\text{Data}_0 =$

ID	X	Y
1	$x_{i=1}$	$y_{i=1}$
2	$x_{i=2}$	$y_{i=2}$
3	$x_{i=3}$	$y_{i=3}$
4	$x_{i=4}$	$y_{i=4}$
5	$x_{i=5}$	$y_{i=5}$

$\text{Data}_{b=1} =$

ID	X	Y
3	$x_{i=3}$	$y_{i=3}$
2	$x_{i=2}$	$y_{i=2}$
4	$x_{i=4}$	$y_{i=4}$
2	$x_{i=2}$	$y_{i=2}$
3	$x_{i=3}$	$y_{i=3}$

, $\text{Data}_{b=2} =$

ID	X	Y
1	$x_{i=1}$	$y_{i=1}$
3	$x_{i=3}$	$y_{i=3}$
3	$x_{i=3}$	$y_{i=3}$
4	$x_{i=4}$	$y_{i=4}$
5	$x_{i=5}$	$y_{i=5}$

, $\text{Data}_{b=3} =$

ID	X	Y
3	$x_{i=3}$	$y_{i=3}$
5	$x_{i=5}$	$y_{i=5}$
1	$x_{i=1}$	$y_{i=1}$
4	$x_{i=4}$	$y_{i=4}$
3	$x_{i=3}$	$y_{i=3}$

, $\cdots$ $\text{Data}_{b=B} =$

ID	X	Y
2	$x_{i=2}$	$y_{i=2}$
5	$x_{i=5}$	$y_{i=5}$
1	$x_{i=1}$	$y_{i=1}$
1	$x_{i=1}$	$y_{i=1}$
4	$x_{i=4}$	$y_{i=4}$

Example

Suppose I'm on the lookout for a retirement location. Since I want to relax in my old-age, I'm concerned with ensuring a low probability of dragon attacks, so I have a few salient considerations:

- Shooting for a location with no dragons, whatsoever, is a fools errand (since dragons are, obviously, ubiquitous).
- I merely require a location that has at least two times as many dragon-free days as other kinds.

Example

I've been watching several candidate locales over the course of my (long and illustrious) career, and I'm particularly hopeful about one quiet hamlet in the Patagonian highlands.

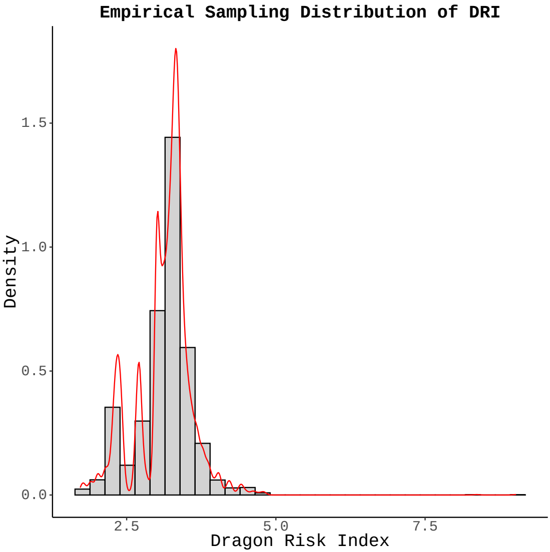
- To ensure that my required degree of dragon-freeness is met, I'll use the *Dragon Risk Index* (DRI):

$$DRI = Median\left(\frac{\text{Dragon-Free Days}}{\text{Dragonned Days}}\right)$$

Example

```r
## Read in the observed data:
rawData <- readRDS("../data/daysData.rds")

## Compute the observed test statistic:
obsDRI <- median(rawData$goodDays / rawData$badDays)
obsDRI

[1] 3.24476

## Draw the bootstrap samples:
set.seed(235711)
nSams <- 5000
bootDRI <- rep(NA, nSams)
for(b in 1:nSams) {
    bootSam    <- rawData[sample(1:nrow(rawData), replace = TRUE), ]
    bootDRI[b] <- median(bootSam$goodDays / bootSam$badDays)
}
```

Example



Empirical Sampling Distribution of DRI
Density
1.5
1.0
0.5
0.0
2.5
5.0
7.5
Dragon Risk Index

Example

To see if I can be confident in the dragon-freeness of my potential home, I'll summarize the preceding distribution with a (one-tailed) percentile confidence interval:

```
bootLB <- sort(bootDRI)[0.05 * nSams]
bootUB <- Inf

## The bootstrapped Percentile CI:
c(bootLB, bootUB)

[1] 2.288555      Inf
```

Bootstrapped Inference for Indirect Effects

We can apply the same procedure to testing the indirect effect.

- The problem with Sobel's Z is exactly the type of issue for which bootstrapping was designed
  - We don't know a reasonable finite-sample sampling distribution for the ab parameter.
- Bootstrapping will allow us to construct an empirical sampling distribution for ab and construct confidence intervals for inference.

Bootstrapped Inference for Indirect Effects

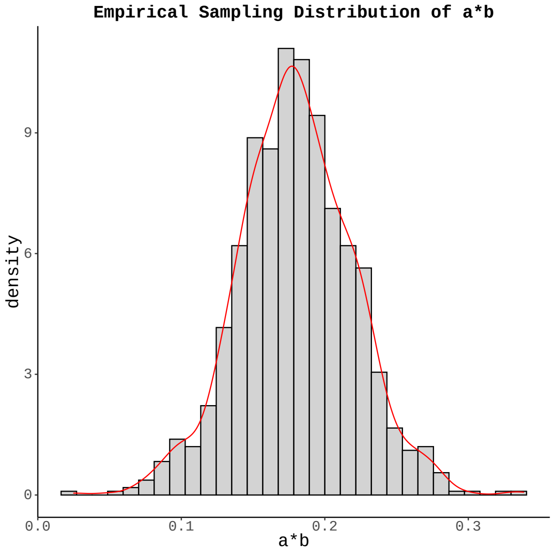
PROCEDURE:

1. Resample our observed data with replacement
2. Fit our hypothesized path model to each bootstrap sample
3. Store the value of ab that we get each time
4. Summarize the empirical distribution of ab to make inferences

Example

```r
nSams <- 1000
abVec <- rep(NA, nSams)
for(i in 1:nSams) {
    ## Resample the data:
    bootSam <- dat1[sample(1:nrow(dat1), replace = TRUE), ]

    ## Fit the path model:
    bootOut <- sem(mod2, data = bootSam)

    ## Store the estimated indirect effect:
    abVec[i] <- coef(bootOut)[c("a", "b")] %>% prod()
}
```

Example



Empirical Sampling Distribution of a*b
density
9
6
3
0
0.0
0.1
0.2
0.3
a*b

Example

```
## Calculate the percentile CI:
lb <- sort(abVec)[0.025 * nSams]
ub <- sort(abVec)[0.975 * nSams]
c(lb, ub)

[1] 0.09804422 0.26334737
```

Example

```
## Much more parsimoniously:
bootOut2 <- sem(mod2, data = dat1, se = "boot", bootstrap = 1000)

parameterEstimates(bootOut2, zstat = FALSE, pvalue = FALSE)

       lhs op      rhs label   est    se ci.lower ci.upper
1   policy  ~   sysRac     b 0.722 0.105    0.519    0.926
2   policy  ~ polAffil       0.051 0.078   -0.095    0.209
3   sysRac  ~ polAffil     a 0.257 0.060    0.129    0.375
4   policy ~~   policy       0.837 0.163    0.530    1.158
5   sysRac ~~   sysRac       0.802 0.128    0.557    1.063
6 polAffil ~~ polAffil       2.444 0.000    2.444    2.444
7       ab :=      a*b    ab 0.186 0.041    0.102    0.265
```

Monte Carlo Method/Parametric Bootstrap

We can also use a *parametric bootstrap* of the individual parameters a and b to get a somewhat robust test of the indirect effect.

- Assuming normal sampling distributions for a and b is not, generally, problematic
- We can save ourselves a lot of computational effort by assuming normality for a and b, then:

  1. Fit the hypothesized path model to the raw data
  2. Extract a, b, and $ACOV(\{a, b\})$ from the fitted path model
  3. Parameterize a bivariate normal distribution $\mathrm{N}(a, b|\mu, \Sigma)$ with $\mu = \{a, b\}$ and $\Sigma = ACOV(\{a, b\})$
  4. Draw simulated values $\{\tilde{a}, \tilde{b}\}$ from $\mathrm{N}(a, b|\mu, \Sigma)$
  5. Compute the simulated indirect effect $\widetilde{ab} = \tilde{a} \cdot \tilde{b}$ and store it
  6. Summarize the empirical distribution of $\widetilde{ab}$ for inference.

Example

```r
## Load package to draw the Monte Carlo samples
library(mvtnorm)

## Specify the model (note the parameter labels):
mod3 <- "
policy ~ polAffil + b*sysRac
sysRac ~ a*polAffil
"

## Fit the model:
out4 <- sem(mod3, data = dat1)

## Extract the important estimates:
a <- coef(out4)["a"]
b <- coef(out4)["b"]
acov <- vcov(out4)[c("a", "b"), c("a", "b")]
```

Aside Regarding Asymptotic Covariances

The *asymptotic covariance matrix* (ACOV) is (-1 times) the inverse of the Fisher information matrix of the model parameters.

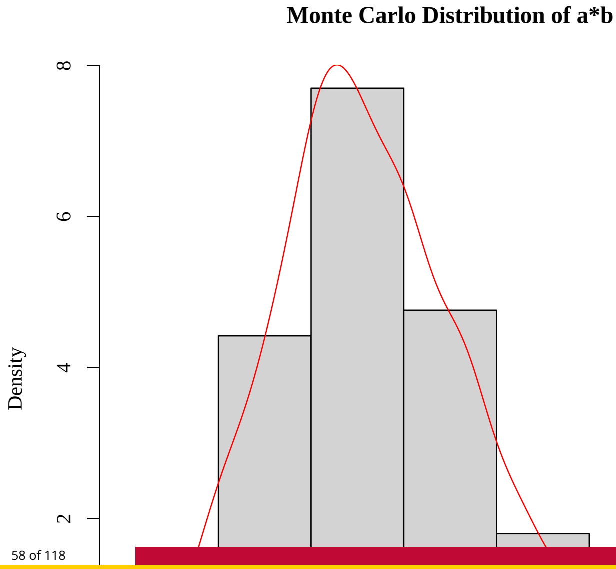
- The $ACOV$ contains the expected covariance among the ML estimates of the model parameters.
- The diagonal elements of the matrix (i.e., the asymptotic variances) are the square of the usual ML SE estimates.

```
round(vcov(out4), 4)

                plcy~A b      a      plcy~~ syR~~R
policy~polAffil  0.005
b               -0.003  0.012
a                0.000  0.000  0.004
policy~~policy   0.000  0.000  0.000  0.016
sysRac~~sysRac   0.000  0.000  0.000  0.000  0.015
```

Back to the Example

```
## Look at ACOV(a,b):
round(acov, 4)

       a     b
a 0.0038 0.000
b 0.0000 0.012

## Draw the Monte Carlo samples:
samMat <- rmvnorm(nSams, c(a, b), acov)
abVec <- samMat[ , "a"] * samMat[ , "b"]
```

Back to the Example

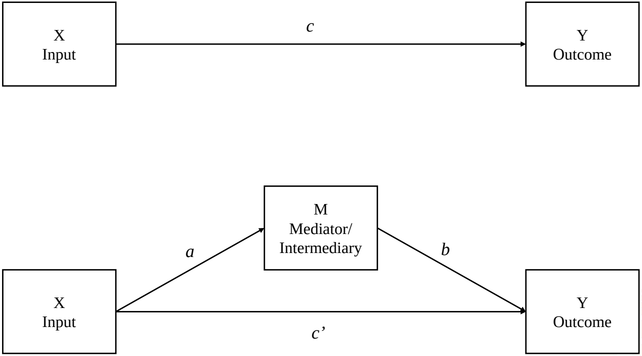


Kris Preacher's Website

If you don't want to program the Monte Carlo approach yourself (although each of you easily can), you should consider the very handy Web App on Professor Kris Preacher's website `http://www.quantpsy.org`.

- Kris' website has a vast array of hugely helpful resources for anyone doing mediation or moderation analysis.
- You should definitely check it out!

Simple Mediation

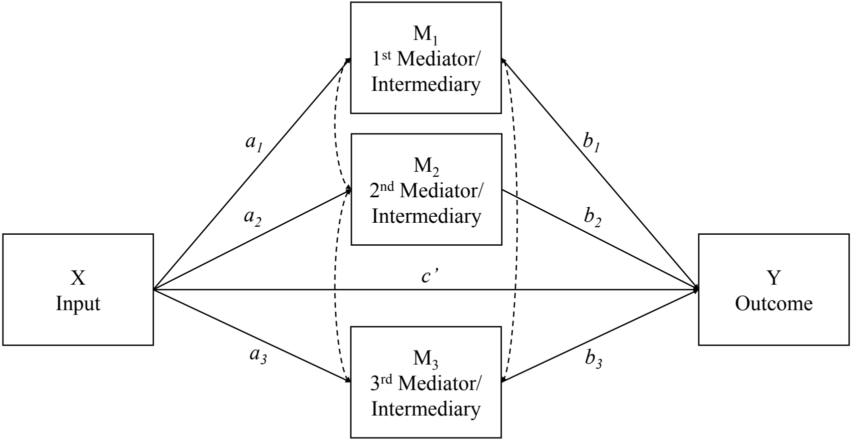


Simple Mediation is Too Simple

We can justify multiple mediator models by asking: “What mediates the effects in a simple mediation model?”

- Mediation of the direct effect leads to *parallel multiple mediator models*.
- Mediation of the a or b paths produces *serial multiple mediator models*.

Parallel Multiple Mediation



Parallel Multiple Mediation

To get all of the information in the preceding diagram, we need to fit four equations:

$$
\begin{aligned}
Y &= i_Y + b_1 M_1 + b_2 M_2 + b_3 M_3 + c' X + e_Y \\
M_1 &= i_{M1} + a_1 X + e_{M1} \\
M_2 &= i_{M2} + a_2 X + e_{M2} \\
M_3 &= i_{M3} + a_3 X + e_{M3}
\end{aligned}
$$

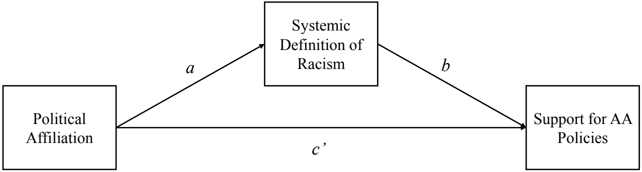
In general, a parallel mediator model with K mediator variables will required $K + 1$ separate equations.

Path modeling can make this task much simpler.

- Also allows us to explicitly estimate the correlations between parallel mediators.

Parallel Multiple Mediation

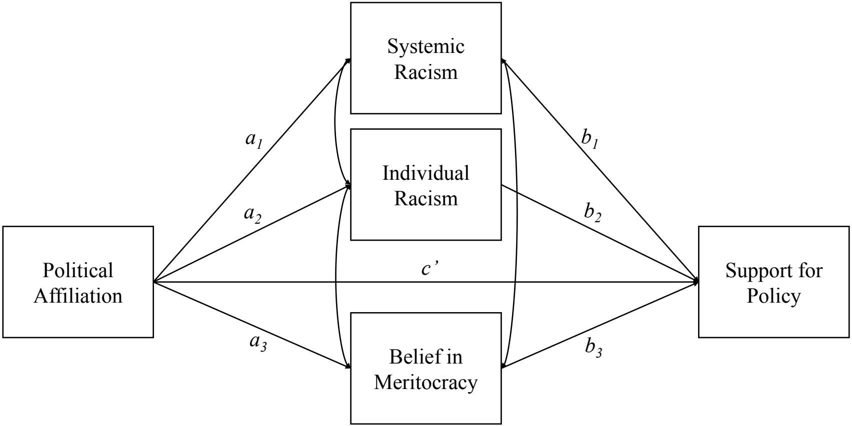
Let's reconsider the example from last week:



Question: What might be mediating the residual direct effect?

Parallel Multiple Mediation

Potential Answer:



A Quick Note on Inference

In simple mediation:

- We have one indirect effect: ab.
- The total effect is equal to the direct effect plus the indirect effect $c = c' + ab$

In parallel multiple mediation:

- We have K *specific indirect effects*, where K is the number of mediators: $a_1b_1, a_2b_2, \ldots, a_Kb_K$.
- The *Total Indirect Effect* is equal to the sum of all the specific indirect effects: $IE_{tot} = \sum_{k=1}^{K} a_kb_k$.
- The *Total Effect* is equal to the direct effect plus the total indirect effect: $c = c' + IE_{tot}$

Inference for the specific indirect effects is basically the same as it is for the sole indirect effect in simple mediation.

- Caveat: Each specific indirect effect must be interpreted as conditional on all other mediators in the model.

Example

Example

```
library(lavaan)

## Read in the data
dataDir <- "../data/"
fileName <- "adamsKlpsScaleScore.rds"
dat1 <- readRDS(paste0(dataDir, fileName))

nBoot <- 2500 # Number of bootstrap samples
bootType <- "bca.simple" # Type of CI

## Parallel Multiple Mediator Model:
mod1.1 <- "
policy ~ b1*sysRac + b2*indRac + b3*merit + cp*polAffil
sysRac ~ a1*polAffil
indRac ~ a2*polAffil
merit ~ a3*polAffil

sysRac ~~ indRac + merit
indRac ~~ merit

ab1 := a1*b1
ab2 := a2*b2
ab3 := a3*b3
```

Example

```
parameterEstimates(out1.1, boot.ci.type = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out1.1, boot.ci.type = bootType):  object 'out1.1'
not found
```

Comparing Specific Indirect Effects

When we have multiple specific indirect effects in a single model, we can test if they are statistically different from one another.

Question: How might we go about doing such a test (assuming we're using path modeling)?

Comparing Specific Indirect Effects

When we have multiple specific indirect effects in a single model, we can test if they are statistically different from one another.

Question: How might we go about doing such a test (assuming we're using path modeling)?

Answer: There are, at least, two reasonable methods:

1. Use nested model $\Delta\chi^2$ tests
2. Define a new parameter corresponding to the null hypothesis and use bootstrapping

Example

Example

```r
## Test differences in specific indirect effects:
mod1.2 <- "
policy ~ b1*sysRac + b2*indRac + b3*merit + cp*polAffil
sysRac ~ a1*polAffil
indRac ~ a2*polAffil
merit ~ a3*polAffil

sysRac ~~ indRac + merit
indRac ~~ merit

ab1 := a1*b1
ab2 := a2*b2
ab3 := a3*b3
totalIE := ab1 + ab2 + ab3

ab1 == ab2
"

out1.2 <-
    sem(mod1.2, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod1.2, data = dat1, se = "boot", bootstrap
= nBoot, :  lavaan ERROR: missing observed variables in dataset: merit
```

Example

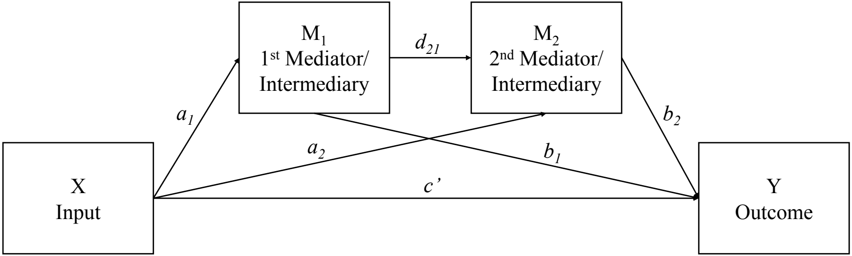
Example

```
## Test differences in specific indirect effects:
mod1.3 <- "
policy ~ b1*sysRac + b2*indRac + b3*merit + cp*polAffil
sysRac ~ a1*polAffil
indRac ~ a2*polAffil
merit ~ a3*polAffil

sysRac ~~ indRac + merit
indRac ~~ merit

ab1 := a1*b1
ab2 := a2*b2
ab3 := a3*b3
totalIE := ab1 + ab2 + ab3

test1 := ab2 - ab1
"

out1.3 <-
    sem(mod1.3, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod1.3, data = dat1, se = "boot", bootstrap
= nBoot, :  lavaan ERROR: missing observed variables in dataset: merit
```

Example

```
parameterEstimates(out1.3, boot.ci.type = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out1.3, boot.ci.type = bootType):  object 'out1.3'
not found
```

Serial Multiple Mediation



Serial Multiple Mediation

To get all of the information in the preceding diagram, we need to fit three equations:

$$
\begin{aligned}
Y &= i_Y + b_1 M_1 + b_2 M_2 + c' X + e_Y \\
M_2 &= i_{M2} + d_{21} M_1 + a_2 X + e_{M2} \\
M_1 &= i_{M1} + a_1 X + e_{M1}
\end{aligned}
$$

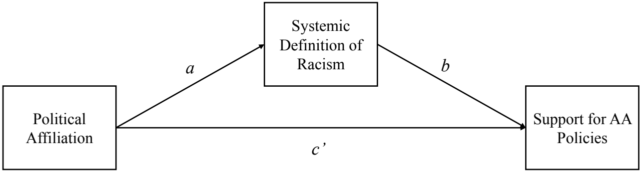
As with parallel mediator models, a serial mediator model with K mediator variables will required $K + 1$ separate equations.

Again, path modeling can make this task much simpler.

- Also allows us to fit more parsimonious, restricted models.

Serial Multiple Mediation

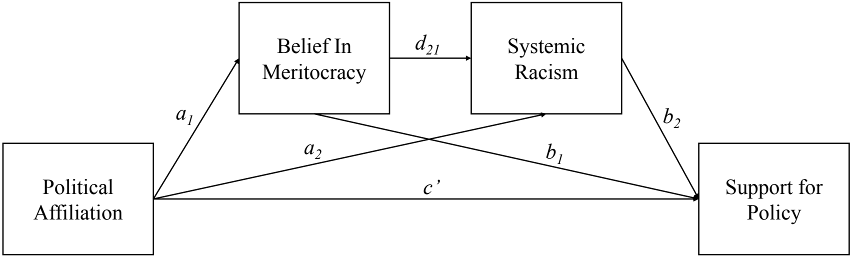
Okay, back to our simple mediation example:



Question: What could be mediating the a path?

Serial Multiple Mediation

Potential Answer:



A Quick Note on Inference

Parallel multiple mediation operates much like a number of combined simple mediation models, serial multiple mediation is not so straight-forward.

In serial multiple mediation:

- Every possible path from X to Y that passes through, at least, one mediator is a specific indirect effect.
  - With the saturated two-mediator model shown above, we have: $IE_{spec} = \{a_1b_1, a_2b_2, a_1d_{21}b_2\}$
- The *Total Indirect Effect* is, again, equal to the sum of all the specific indirect effects: $IE_{tot} = \sum_{k=1}^{|\{IE_{spec}\}|} IE_{spec,k}$.
- The *Total Effect* is equal to the direct effect plus the total indirect effect: $c = c' + IE_{tot}$

A Quick Note on Inference

Inference for the specific indirect effects is basically the same as it is for the sole indirect effect in simple mediation, when using path modeling or bootstrapping.

- Caveat: Normal-theory, Sobel-Type, standard errors for the specific indirect effects that involve more than two constituent paths can be very complex.
    - This isn't really a problem since you should always use bootstrapping, anyway!

Example

Example

```
## Serial Multiple Mediator Model:
mod2.1 <- "
policy ~ b1*merit + b2*sysRac + cp*polAffil
sysRac ~ d21*merit + a2*polAffil
merit ~ a1*polAffil

ab1 := a1*b1
ab2 := a2*b2
fullIE := a1*d21*b2
totalIE := ab1 + ab2 + fullIE
"

out2.1 <-
    sem(mod2.1, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod2.1, data = dat1, se = "boot", bootstrap
= nBoot, :  lavaan ERROR: missing observed variables in dataset:  merit

summary(out2.1)

Error in h(simpleError(msg, call)):  error in evaluating the argument
'object' in selecting a method for function 'summary':  object 'out2.1' not
found
```

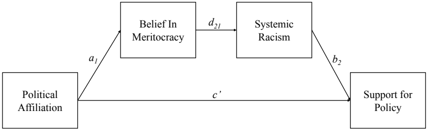
Example

```
parameterEstimates(out2.1, boot.ci.type = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out2.1, boot.ci.type = bootType):  object 'out2.1'
not found
```

Restricted Models

In this example, the a_2 and b_1 paths are non-significant as are the simple specific indirect effects a_1b_1 and a_2b_2.

- There is a school of thinking that would prescribe constraining the a_2 and b_1 paths to zero as in:



- This model will ascribe a larger effect size to $a_1d_{21}b_2$ since it must convey all of the indirect influence of X on Y.
  - We should first fit a saturated model, but subsequently culling non-significant paths can, sometimes, be appropriate.

Example

```
mod2.2 <- "
policy ~ cp*polAffil + b2*sysRac
merit ~ a1*polAffil
sysRac ~ d21*merit

fullIE := a1*d21*b2
"

out2.2 <-
    sem(mod2.2, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod2.2, data = dat1, se = "boot", bootstrap
= nBoot, :  lavaan ERROR: missing observed variables in dataset:  merit

summary(out2.2)

Error in h(simpleError(msg, call)):  error in evaluating the argument
'object' in selecting a method for function 'summary':  object 'out2.2' not
found
```

Example

```
parameterEstimates(out2.2, bootstrap = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out2.2, bootstrap = bootType):  unused argument
(bootstrap = bootType)
```

Example

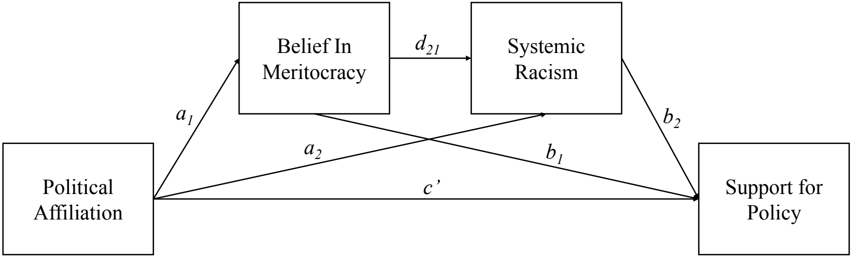
As in parallel multiple mediation, we can test for differences in the specific indirect effects of a serial multiple mediator model:

Example

```
## Test Differences between Indirect Effects
## in Serial Multiple Mediator Model (Method 1):
mod2.3 <- "
policy ~ cp*polAffil + b1*merit + b2*sysRac
merit ~ a1*polAffil
sysRac ~ a2*polAffil + d21*merit

ab1 := a1*b1
ab2 := a2*b2
fullIE := a1*d21*b2
totalIE := ab1 + ab2 + fullIE

fullIE == ab1
fullIE == ab2
"

out2.3 <-
    sem(mod2.3, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod2.3, data = dat1, se = "boot", bootstrap
= nBoot, :  lavaan ERROR: missing observed variables in dataset:  merit

summary(out2.3)
```

Serial Multiple Mediation

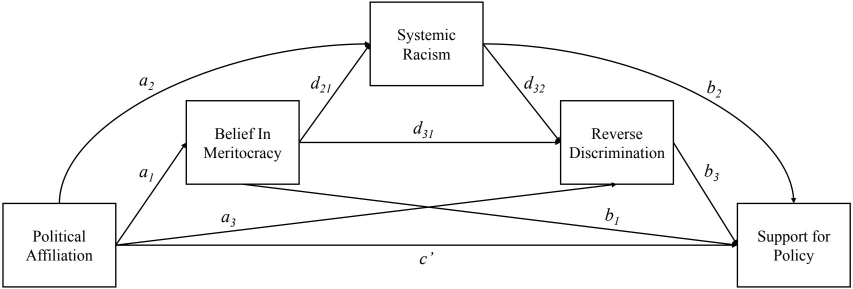
Okay, so we've supported an interesting hypothesis with the following model, but why stop there?



Question: What might mediated the b_2 path?

Serial Multiple Mediation

Potential Answer:



Serial Multiple Mediation

Question: How many equations do we need to get the information in the preceding diagram?

Serial Multiple Mediation

Question: How many equations do we need to get the information in the preceding diagram?

$$\begin{aligned}
Policy &= i_Y + b_1 Merit + b_2 SysRac + b_3 RevDisc + c' PolAff + e_Y \\
RevDisc &= i_{M3} + d_{31} Merit + d_{32} SysRac + a_3 PolAff + e_{M3} \\
SysRac &= i_{M2} + d_{21} Merit + a_2 PolAff + e_{M2} \\
Merit &= i_{M1} + a_1 PolAff + e_{M1}
\end{aligned}$$

Which produces the following set of specific indirect effects:

- $a_1 b_1$
- $a_2 b_2$
- $a_3 b_3$
- $a_1 d_{31} b_3$
- $a_1 d_{21} b_2$
- $a_2 d_{32} b_3$
- $a_1 d_{21} d_{32} b_3$

Example

Example

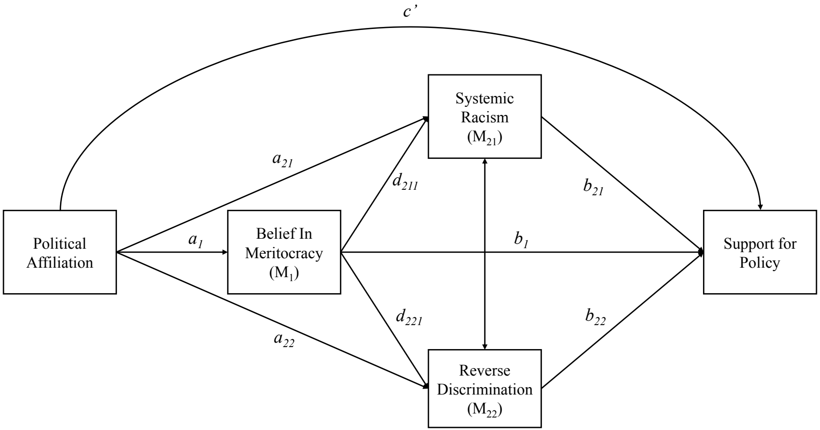
```
## Serial Multiple Mediator Model with 3 Mediators:
mod3.1 <- "
policy ~ b1*merit + b2*sysRac + b3*revDisc + cp*polAffil
revDisc ~ d31*merit + d32*sysRac + a3*polAffil
sysRac ~ d21*merit + a2*polAffil
merit ~ a1*polAffil

ab1 := a1*b1
ab2 := a2*b2
ab3 := a3*b3

partIE1 := a1*d31*b3
partIE2 := a1*d21*b2
partIE3 := a2*d32*b3

fullIE := a1*d21*d32*b3

totalIE := ab1 + ab2 + ab3 + partIE1 + partIE2 + partIE3 + fullIE
"

out3.1 <-
    sem(mod3.1, data = dat1, se = "boot", bootstrap = nBoot)
```

Example

```
parameterEstimates(out3.1, boot.ci.type = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out3.1, boot.ci.type = bootType):  object 'out3.1'
not found
```

Hybrid Multiple Mediation

We can also combine parallel and serial mediation models:



Example

Example

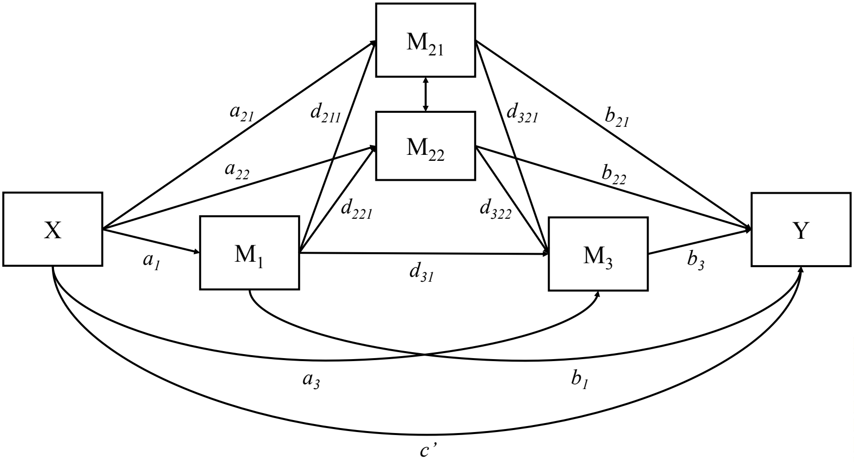
```
## Hybrid Multiple Mediator Model:
mod4.1 <- "
policy ~ b1*merit + b21*sysRac + b22*revDisc + cp*polAffil
sysRac ~ d211*merit + a21*polAffil
revDisc ~ d221*merit + a22*polAffil
merit ~ a1*polAffil

sysRac ~~ revDisc

ab1 := a1*b1
ab21 := a21*b21
ab22 := a22*b22

fullIE21 := a1*d211*b21
fullIE22 := a1*d221*b22

totalIE := ab1 + ab21 + ab22 + fullIE21 + fullIE22
"

out4.1 <-
    sem(mod4.1, data = dat1, se = "boot", bootstrap = nBoot)

Error in lavaan::lavaan(model = mod4.1, data = dat1, se = "boot", bootstrap
```

Example

```
parameterEstimates(out4.1, boot.ci.type = bootType)[ , -c(1 : 3)]

Error in parameterEstimates(out4.1, boot.ci.type = bootType):  object 'out4.1'
not found
```

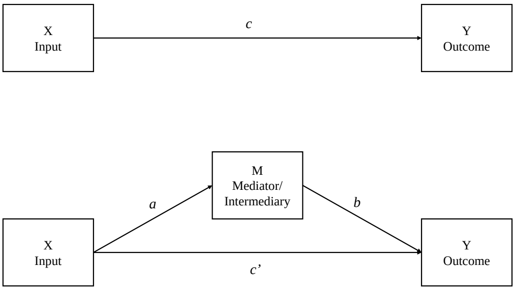
Practice

List all of the specific indirect effects present in this model:



Boring Model

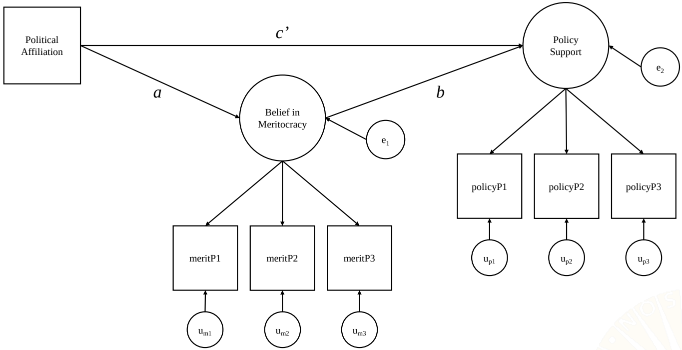
So far, all of our models have been similar to:



But there is no reason that we need to restrict ourselves to mucking about with observed variables.

Better Model

We can (and should) test for indirect effects using *latent variable models* such as:



Measurement error can be a big problem for mediation analysis, so latent variable modeling is highly recommended.

Example

Example

```
library(lavaan)
dataDir <- "../data/"
dat1 <- readRDS(paste0(dataDir, "adamsKlpsData.rds"))

## Specify the CFA model:
mod1.1 <- "
merit =~ meritP1 + meritP2 + meritP3
policy =~ policyP1 + policyP2 + policyP3
"

## Fit the CFA and check model:
out1.1 <- cfa(mod1.1, data = dat1, std.lv = TRUE)

Error in lavaan::lavaan(model = mod1.1, data = dat1, std.lv = TRUE,
model.type = "cfa", :  lavaan ERROR: some latent variable names collide
with observed
variable names:  merit policy

## Check model fit:
round(fitMeasures(out1.1)[c("chisq", "df", "pvalue", "cfi",
                            "tli", "rmsea", "srmr")], 4)

Error in h(simpleError(msg, call)):  error in evaluating the argument
'object' in
```

Interpretation of Indirect Effects

Although indirect effects are composed parameters, they have direct interpretations, independent of the interpretations of their constituent paths:

- The $X \rightarrow M \rightarrow Y$ indirect effect ab is interpreted as:
  - The expected change in Y for a unit change in X that is transmitted indirectly through M, or...
  - For a unit change in X, Y is expected to change by ab units, indirectly through M, or...
  - Participants who differ by one unit on X are expect to differ by ab units on Y as a results of the effect of X on M which, in turn, affects Y.

- The interpretation/scaling of the indirect effect is entirely defined by the input X and outcome Y
  - The scaling of the intermediary variable M does not affect the interpretation of the indirect effect.

Partially Standardized Indirect Effect

$$ab_{ps} = \frac{ab}{SD_Y}$$
$$c'_{ps} = \frac{c'}{SD_Y}$$
$$c_{ps} = \frac{c}{SD_Y} = ab_{ps} + c'_{ps}$$

- Simple
- Removes binding to the scale of Y
- Still scale-bound by X
- Not clear what constitutes a “large” effect

Completely Standardized Indirect Effect

$$ab_{cs} = \frac{SD_X ab}{SD_Y}$$

$$c'_{cs} = \frac{SD_X c'}{SD_Y}$$

$$c_{cs} = \frac{SD_X c}{SD_Y} = ab_{cs} + c'_{cs}$$

- Simple
- Removes all scale binding
- Not clear what constitutes a “large” effect

Ratio of the Indirect Effect to the Total Effect

$$P_M = \frac{ab}{c} = \frac{ab}{c' + ab}$$

- Very simple
- Not bounded by 0 and 1
- Explodes toward $\pm\infty$ as $c \to 0$
- Very unstable
  - High between-sample variability
  - Requires $N \geq 500$

Ratio of the Indirect Effect to the Direct Effect

$$R_M = \frac{ab}{c'} = \frac{P_M}{1 - P_M}$$

- Very simple
- Not bounded by 0 and 1
- Explodes toward $\pm\infty$ as $c' \to 0$
- Very unstable
  - High between-sample variability
  - Requires $N \geq 2000$

Proportion of Variance in Y Explained by the Indirect Effect

Developed by Fairchild, MacKinnon, Taborga, and Taylor (2009).

- Given a non-zero total effect, represents the proportion of variance in Y accounted for by the indirect effect.

$$R^2_{med} = r^2_{MY} - (R^2_{Y.MX} - r^2_{XY})$$

- Mostly sensible interpretation
- Predicated on the assumption that $\beta_{YX} \neq 0$
- $|ab| > |c| \Rightarrow R^2_{med} < 0$
  - Not a strict proportion

Kappa Squared

Developed by Preacher and Kelley (2011).

- Gives the proportion of the *maximum possible* indirect effect represented by ab.

$$\kappa^2 = \frac{ab}{\max(ab)}$$

- Bounded by 0 and 1
- Values closer to 1.0 indicate a bigger effect
- A bit of a pain to calculate.

Computing $\max(ab)$

$$a \in \left\{ \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_X^2\sigma_Y^2} \right\} = [a_{low}, a_{high}],$$

$$b \in \left\{ \pm \frac{\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2\sigma_M^2 - \sigma_{MX}^2}} \right\} = [b_{low}, b_{high}],$$

$$\max(a) = \begin{cases} a_{high}, & \text{if} \quad \hat{a} > 0 \\ a_{low}, & \text{if} \quad \hat{a} < 0 \end{cases}, \quad \max(b) = \begin{cases} b_{high}, & \text{if} \quad \hat{b} > 0 \\ b_{low}, & \text{if} \quad \hat{b} < 0 \end{cases},$$

$$\max(ab) = \max(a)\max(b)$$

Example

Example

```
## Specify the model:
mod2 <- "
policy ~ b*sysRac + cp*polAffil
sysRac ~ a*polAffil

ab := a*b
"

## Estimate the model:
out2 <- sem(mod2, data = dat1)
##
## Extract/compute the necessary quantities:
ab <- prod(coef(out2)[c("a", "b")])
ab

[1] 0.1015958

cPrime <- coef(out2)["cp"]
##
sdY <- sd(dat1$policy)
sdX <- sd(dat1$polAffil)
##
r2MY <- with(dat1, cor(policy, sysRac))^2
```

Compute κ^2

```r
## Subset the data:
tmpData <- dat1[ , c("polAffil", "sysRac", "policy")]
colnames(tmpData) <- c("x", "m", "y")
##
## Extract pertinent variance/covariance elements:
cov1 <- cov(tmpData)

sYM <- cov1["x", "m"]
sYX <- cov1["y", "x"]
sMX <- cov1["m", "x"]
s2X <- cov1["x", "x"]
s2M <- cov1["m", "m"]
s2Y <- cov1["y", "y"]
```

Compute κ^2

```
## Possible range of a:
aMarg <- sqrt(s2M * s2Y - sYM^2) * sqrt(s2X * s2Y - sYX^2)
aInt <- c(
    (sYM * sYX - aMarg) / (s2X * s2Y),
    (sYM * sYX + aMarg) / (s2X * s2Y)
)
aInt

[1] -0.4378558  0.5793099

##
## Possible range of b:
bMarg <- sqrt(s2X * s2Y - sYX^2) / sqrt(s2X * s2M - sMX^2)
bInt <- c(-1 * bMarg, bMarg)
bInt

[1] -1.289996  1.289996

##
## max(a):
aMax <- ifelse(coef(out2)["a"] < 0,
               aInt[1],
               aInt[2])
aMax
```

Practice

Suppose:

1. Σ is given by:

	x	m	y
x	1.5		
m	0.3	1.4	
y	0.6	0.45	1.55

2. The estimated paths are:
   - $a = 0.2$
   - $b = 0.246$
   - $ab = 0.049$

Compute κ^2 for the estimated ab.

References

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.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26. doi: 10.1214/aos/1176344552

Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R-squared effect-size measures for mediation analysis. *Behavior Research Methods*, *41*(2), 486–498.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*(2), 93.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*(1982), 290–312.