Mediation & Moderation

Theory Construction and Statistical Modeling

Kyle M. Lang

Department of Methodology & Statistics
Utrecht University

Outline

Mediation

- Indirect Effects
- Causal Steps Approach
- Sobel's Test
- Bootstrapping

Moderation

- Testing Moderation
- Post Hoc Analysis

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”

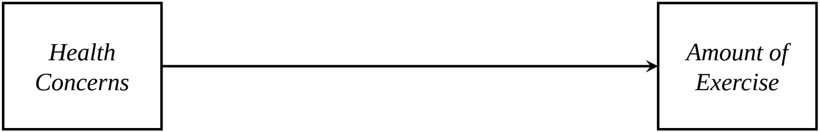
- DV: Hours per week spent in vigorous exercise (*exerciseAmount*).

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*

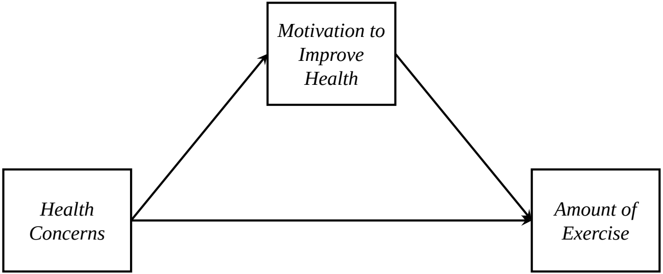


- Mediation and moderation both attempt to describe the focal effect in more detail.
- We always begin by hypothesizing a focal effect.

The Mediation Hypothesis

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

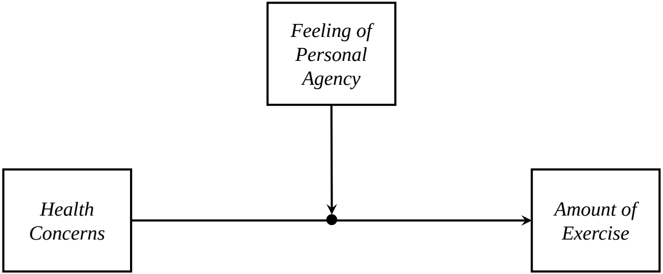
- The *how* is operationalized in terms of intermediary variables.
- Mediator: Motivation to improve health (*motivation*).



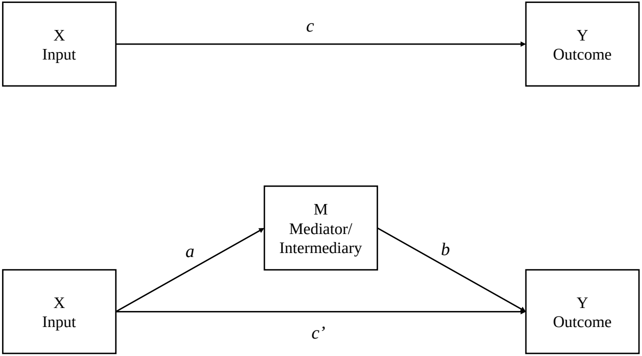
Moderation Hypothesis

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*).



Path Diagrams



Necessary Equations

To get all the pieces of the preceding diagram using OLS regression, we'll need to fit three separate 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 partial 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:

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

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

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.

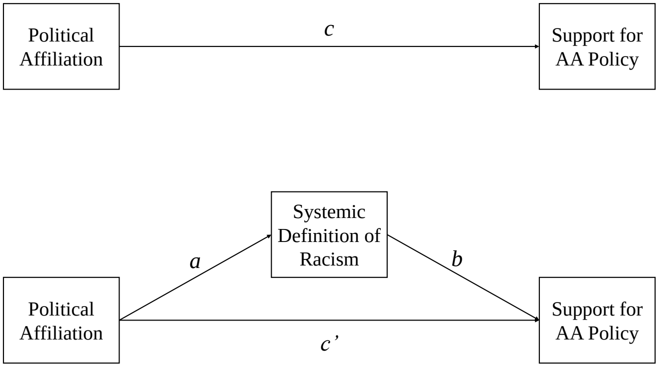
The Causal Steps Approach

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 
-2.28970 -0.53821  0.08866  0.64015  3.08343 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.1218     0.4883   2.297   0.0241 *  
sysRac        0.6649     0.1210   5.494 4.03e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.026 on 85 degrees of freedom
Multiple R-squared:  0.262,Adjusted R-squared:  0.2534 
F-statistic: 30.18 on 1 and 85 DF,  p-value: 4.029e-07
```

Causal Steps Example

```
summary(mod3)


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

Residuals:
    Min      1Q  Median      3Q     Max 
-2.2187 -0.5449 -0.2115  0.6182  1.9516 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.19726    0.27634  11.570   <2e-16 ***
polAffil     0.17023    0.06027   2.825   0.0059 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8788 on 85 degrees of freedom
Multiple R-squared:  0.08581,Adjusted R-squared:  0.07505 
F-statistic: 7.978 on 1 and 85 DF,  p-value: 0.005898
```

Causal Steps Example

```
summary(mod4)


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

Residuals:
    Min      1Q  Median      3Q     Max 
-2.1370 -0.6338 -0.0020  0.6658  3.4674 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.80704    0.51013   1.582   0.1174    
sysRac       0.59680    0.12478   4.783  7.3e-06 ***
polAffil     0.13515    0.07252   1.864   0.0658 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.011 on 84 degrees of freedom
Multiple R-squared:  0.2913,Adjusted R-squared:  0.2745 
F-statistic: 17.27 on 2 and 84 DF,  p-value: 5.228e-07
```

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] 0.1015958

ieProd

[1] 0.1015958
```

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.

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

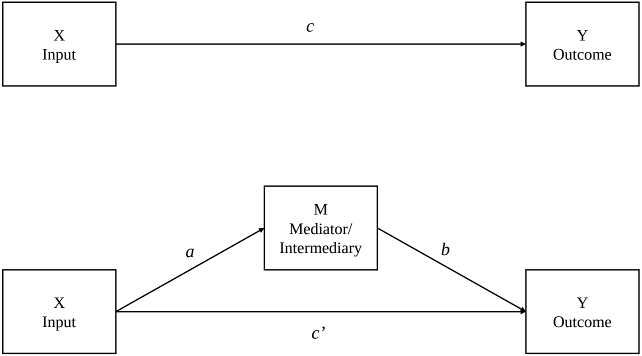
$$Z_{sobel} = \frac{ab}{SE_{sobel}} \tag{7}$$

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

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] 2.432107

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

[1] 0.01501126

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

[1] 0.01972121 0.18347034
```

Recall our Basic Path Diagram



Two Measures of Indirect Effect

Recall the two definitions of an indirect effect:

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

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

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}.
- 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:

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

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

It pays to remember a few key points:

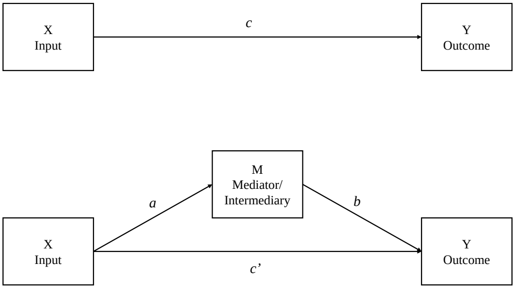
- IE_{diff} and IE_{prod} are equivalent in simple mediation.
- IE_{diff} is only an indirect indication of IE_{prod}.
- If we only care about the indirect effect, then we don't need to worry about the total effect.

These points imply something interesting:

- We don't need to estimate c!

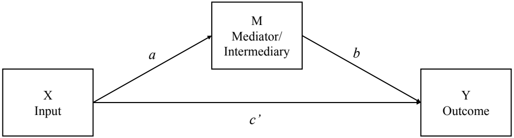
Simplifying our Path Diagram

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

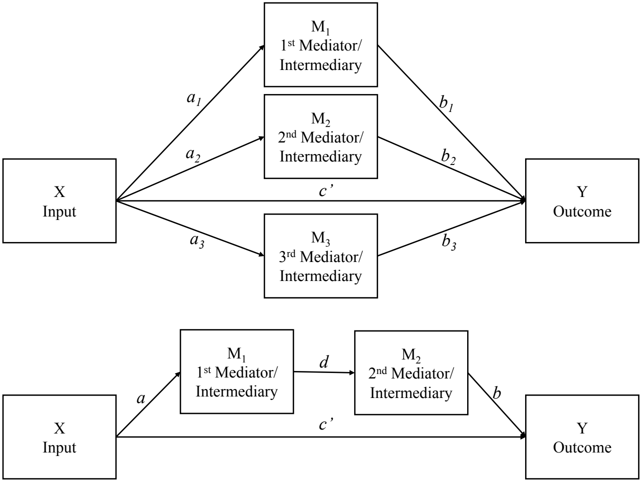


Simplifying our Path Diagram

ANSWER: We don't fit the upper model.

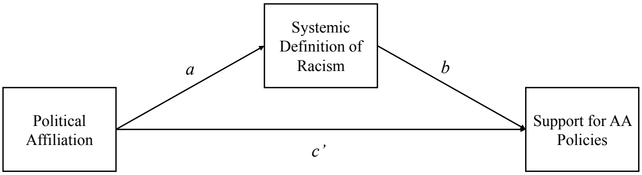


Why Path Analysis?



Example

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



Example

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

## Specify the basic path model:
mod1 <- '
policy ~ 1 + sysRac + polAffil
sysRac ~ 1 + polAffil
'

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

Example

```
## Look at the results:
partSummary(out1, 7:9)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  policy ~
    sysRac            0.597    0.123    4.867    0.000
    polAffil          0.135    0.071    1.897    0.058
  sysRac ~
    polAffil          0.170    0.060    2.858    0.004

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.807    0.501    1.610    0.107
   .sysRac            3.197    0.273   11.705    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.987    0.150    6.595    0.000
   .sysRac            0.755    0.114    6.595    0.000
```

Example

```
## Include the indirect effect:
mod2 <- '
policy ~ 1 + b*sysRac + polAffil
sysRac ~ 1 + 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, 7:8)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  policy ~
    sysRac     (b)    0.597    0.123    4.867    0.000
    polAffil          0.135    0.071    1.897    0.058
  sysRac ~
    polAffil   (a)    0.170    0.060    2.858    0.004

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.807    0.501    1.610    0.107
   .sysRac            3.197    0.273   11.705    0.000
```

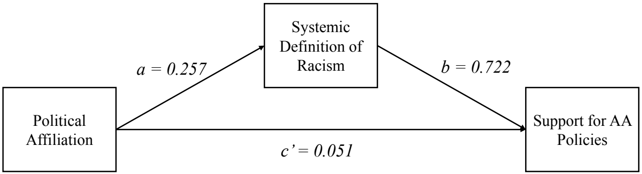
Example

```
partSummary(out2, 9:10)

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .policy            0.987    0.150    6.595    0.000
   .sysRac            0.755    0.114    6.595    0.000

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ab                0.102    0.041    2.464    0.014
```

Example

```
## 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 ~1                 0.807 0.501   -0.175    1.789
2    policy  ~    sysRac     b 0.597 0.123    0.356    0.837
3    policy  ~ polAffil        0.135 0.071   -0.005    0.275
4    sysRac ~1                 3.197 0.273    2.662    3.733
5    sysRac  ~ polAffil     a 0.170 0.060    0.053    0.287
6    policy ~~    policy        0.987 0.150    0.694    1.280
7    sysRac ~~    sysRac        0.755 0.114    0.530    0.979
8  polAffil ~~ polAffil        2.444 0.000    2.444    2.444
9  polAffil ~1                 4.310 0.000    4.310    4.310
10       ab :=       a*b    ab 0.102 0.041    0.021    0.182
```

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):

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

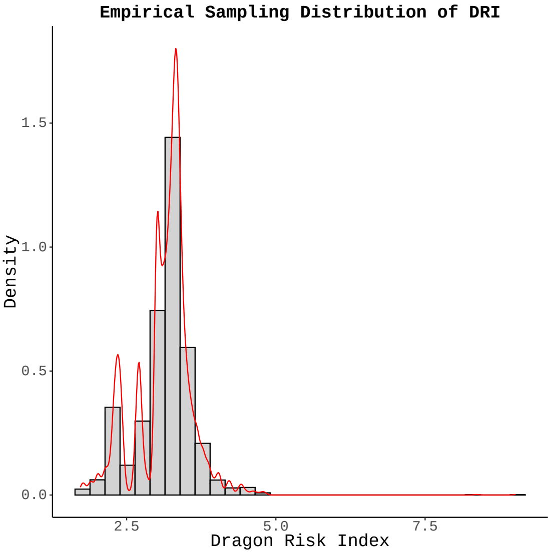
Example

```r
## Load some useful packages:
library(dplyr)
library(magrittr)

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

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

[1] 3.24476
```

Example

```r
## Define the number of bootstrap samples:
nSams <- 5000

## Set a seed for the RNG:
set.seed(235711)

## Bootstrap the DRI statistic:
bootDRI <- rep(NA, nSams)
for(b in 1:nSams)
  bootDRI[b] <- rawData %>%
    slice_sample(prop = 1, replace = TRUE) %$% # Resample the data
    median(goodDays / badDays)                 # Calculate the DRI
```

Example


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

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:

```
## Compute the 95% bootstrapped percentile CI:
quantile(bootDRI, c(0.05, 1.0))

      5%     100%
2.288555 9.016917
```

Since we have a directional hypothesis, the upper bound of this interval is a bit misleading.

```
max(bootDRI)

[1] 9.016917

qnorm(1.0, mean(bootDRI), sd(bootDRI))

[1] 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

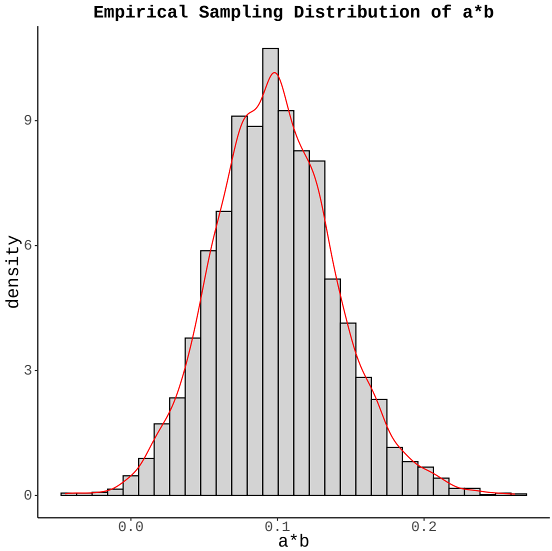
PROCEDURE:

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

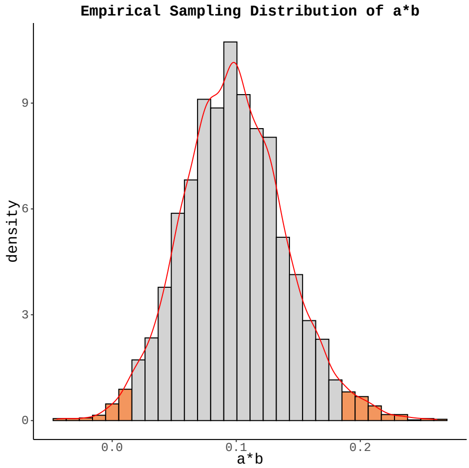
```r
abVec <- rep(NA, nSams)
for(i in 1:nSams)
  abVec[i] <- dat1 %>%
    slice_sample(prop = 1, replace = TRUE) %>% # Resample the data
    sem(mod2, data = .) %>%                   # Fit the model
    coef() %>%                                # Extract estimates
    extract(c("a", "b")) %>%                  # Isolate a and b
    prod()                                    # Calculate IE
```

Example


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

Example

```
## Calculate the 95% CI:
quantile(abVec, c(0.025, 0.975))

      2.5%      97.5%
0.01983687 0.18521091
```



Example

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

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

        lhs op      rhs label   est    se ci.lower ci.upper
1    policy ~1                0.807 0.568   -0.273    1.938
2    policy  ~   sysRac     b 0.597 0.137    0.313    0.848
3    policy  ~ polAffil       0.135 0.084   -0.029    0.300
4    sysRac ~1                3.197 0.277    2.689    3.779
5    sysRac  ~ polAffil     a 0.170 0.064    0.035    0.291
6    policy ~~   policy       0.987 0.164    0.659    1.302
7    sysRac ~~   sysRac       0.755 0.108    0.535    0.956
8  polAffil ~~ polAffil       2.444 0.000    2.444    2.444
9  polAffil ~1                4.310 0.000    4.310    4.310
10       ab :=      a*b    ab 0.102 0.041    0.020    0.186
```

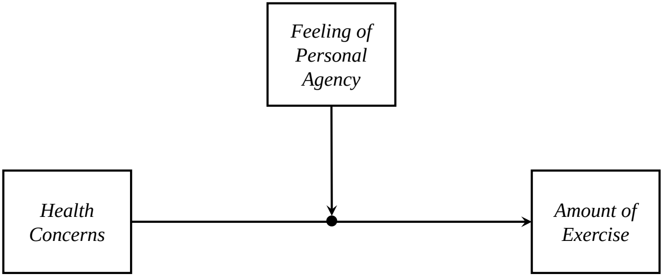
MODERATION

Refresher: 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*).



Equations

In additive MLR, we might have the following equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$

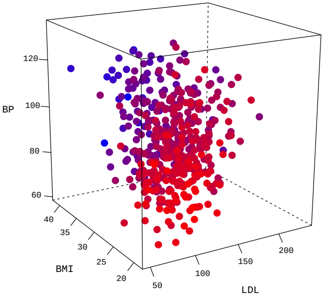
This additive equation assumes that X and Z are independent predictors of Y.

When X and Z are independent predictors, the following are true:

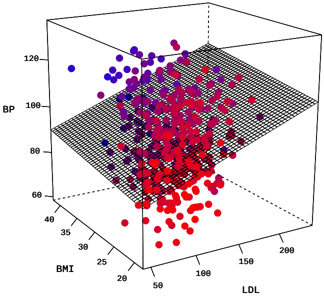
- X and Z *can* be correlated.
- β_1 and β_2 are *partial* regression coefficients.
- The effect of X on Y is the same at **all levels** of Z, and the effect of Z on Y is the same at **all levels** of X.

Additive Regression

The effect of X on Y is the same at **all levels** of Z.

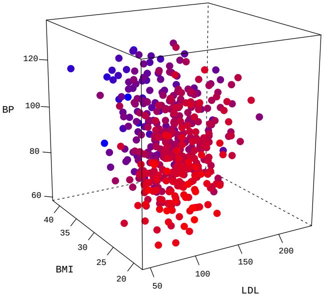


→

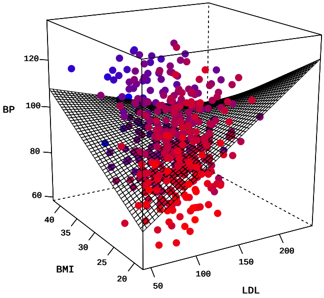


Moderated Regression

The effect of X on Y varies **as a function** of Z.



$\longrightarrow$



Equations

The following derivation is adapted from Hayes (2022).

- When testing moderation, we hypothesize that the effect of X on Y varies as a function of Z.
- We can represent this concept with the following equation:

$$Y = \beta_0 + f(Z)X + \beta_2 Z + \varepsilon \tag{11}$$

Equations

The following derivation is adapted from Hayes (2022).

- When testing moderation, we hypothesize that the effect of X on Y varies as a function of Z.
- We can represent this concept with the following equation:

$$Y = \beta_0 + f(Z)X + \beta_2 Z + \varepsilon \tag{11}$$

- If we assume that Z linearly (and deterministically) affects the relationship between X and Y, then we can take:

$$f(Z) = \beta_1 + \beta_3 Z \tag{12}$$

Equations

- Substituting Equation 12 into Equation 11 leads to:

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

Equations

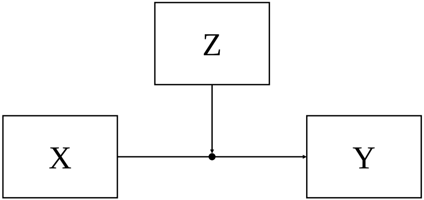
- Substituting Equation 12 into Equation 11 leads to:

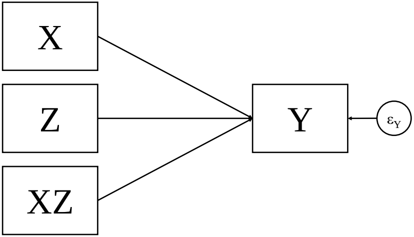
$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

- Which, after distributing X and reordering terms, becomes:

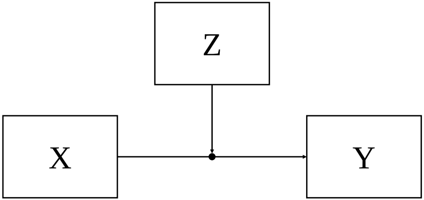
$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

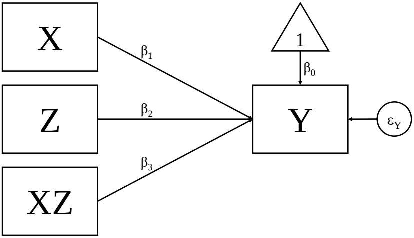
Conceptual vs. Analytic Diagrams





Conceptual vs. Analytic Diagrams





Testing Moderation

Now, we have an estimable regression model that quantifies the linear moderation we hypothesized.

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

- To test for significant moderation, we simply need to test the significance of the interaction term, XZ.
  - Check if $\hat{\beta}_3$ is significantly different from zero.

Interpretation

Given the following equation:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 Z + \hat{\beta}_3 XZ + \hat{\varepsilon}$$

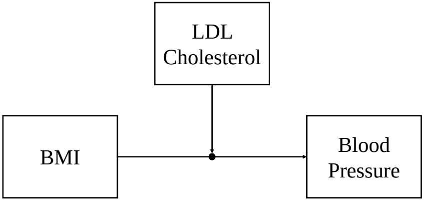
- $\hat{\beta}_3$ quantifies the effect of Z on the focal effect (the $X \rightarrow Y$ effect).
  - For a unit change in Z, $\hat{\beta}_3$ is the expected change in the effect of X on Y.
- $\hat{\beta}_1$ and $\hat{\beta}_2$ are *conditional effects*.
  - Interpreted where the other predictor is zero.
  - For a unit change in X, $\hat{\beta}_1$ is the expected change in Y, when $Z = 0$.
  - For a unit change in Z, $\hat{\beta}_2$ is the expected change in Y, when $X = 0$.

Example

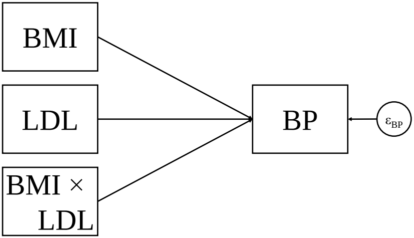
Looking at the *diabetes* dataset.

- We suspect that patients' BMIs are predictive of their average blood pressure.
- We further suspect that this effect may be differentially expressed depending on the patients' LDL levels.

Diagrams



LDL
Cholesterol
BMI
Blood
Pressure



BMI
LDL
BMI ×
LDL
BP
ε_{BP}

Example

```
dDat <- readRDS("../data/diabetes.rds")

## Focal Effect:
out0 <- lm(bp ~ bmi, data = dDat)
partSummary(out0, -c(1, 2))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  61.9973     3.6659   16.91   <2e-16
bmi           1.2379     0.1371    9.03   <2e-16

Residual standard error: 12.72 on 440 degrees of freedom
Multiple R-squared:  0.1563,Adjusted R-squared:  0.1544
F-statistic: 81.54 on 1 and 440 DF,  p-value: < 2.2e-16
```

Example

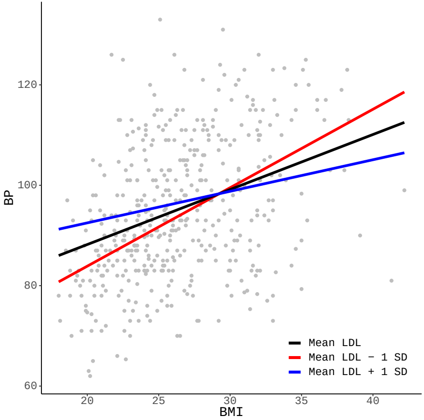
```
## Additive Model:
out1 <- lm(bp ~ bmi + ldl, data = dDat)
partSummary(out1, -c(1, 2))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 59.26577    3.91281  15.147  < 2e-16
bmi          1.16567    0.14156   8.235 2.08e-15
ldl          0.04016    0.02056   1.953   0.0515

Residual standard error: 12.68 on 439 degrees of freedom
Multiple R-squared:  0.1636,Adjusted R-squared:  0.1598
F-statistic: 42.94 on 2 and 439 DF,  p-value: < 2.2e-16
```

Example

```
## Moderated Model:
out2 <- lm(bp ~ bmi * ldl, data = dData)
partSummary(out2, -c(1, 2))

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.480616  14.291677   1.013 0.311514
bmi          2.867825   0.541312   5.298 1.86e-07
ldl          0.448771   0.127160   3.529 0.000461
bmi:ldl     -0.015352   0.004716  -3.255 0.001221

Residual standard error: 12.54 on 438 degrees of freedom
Multiple R-squared:  0.1834,Adjusted R-squared:  0.1778
F-statistic: 32.78 on 3 and 438 DF,  p-value: < 2.2e-16
```

Visualizing the Interaction

We can get a better idea of the patterns of moderation by plotting the focal effect at conditional values of the moderator.



Example

Of course, we can fit the same model in **lavaan**.

```
library(lavaan)

## Specify the model:
mod <- 'bp ~ 1 + bmi + ldl + bmi:ldl'

## Estimate the model:
lavOut <- sem(mod, data = dDat)
```

Example

```
partSummary(lavOut, 7:9)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  bp ~
    bmi               2.868    0.539    5.322    0.000
    ldl               0.449    0.127    3.545    0.000
    bmi:ldl          -0.015    0.005   -3.270    0.001

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
   .bp               14.481   14.227    1.018    0.309

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .bp              155.871   10.485   14.866    0.000
```

Probing the Interaction

A significant estimate of β_3 tells us that the effect of X on Y depends on the level of Z, but not much more.

- The plot above gives a descriptive illustration of the pattern, but does not support statistical inference.
  - The three conditional effects we plotted look different, but we cannot say much about how they differ with only the plot and $\hat{\beta}_3$.
- This is the purpose of *probing* the interaction.
  - Try to isolate areas of Z's distribution in which $X \rightarrow Y$ effect is significant and areas where it is not.

Probing the Interaction

The most popular method of probing interactions is to do a so-called *simple slopes* analysis.

- Pick-a-point approach
- Spotlight analysis

In simple slopes analysis, we test if the slopes of the conditional effects plotted above are significantly different from zero.

- To do so, we test the significance of *simple slopes*.

Simple Slopes

Recall the derivation of our moderated equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

We can reverse the process by factoring out X and reordering terms:

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

Where $f(Z) = \beta_1 + \beta_3 Z$ is the linear function that shows how the relationship between X and Y changes as a function of Z.

$f(Z)$ is the *simple slope*.

- By plugging different values of Z into $f(Z)$, we get the value of the conditional effect of X on Y at the chosen level of Z.

Significance Testing of Simple Slopes

The values of Z used to define the simple slopes are arbitrary.

- The most common choice is: $\{(\bar{Z} - SD_Z), \bar{Z}, (\bar{Z} + SD_Z)\}$
- You could also use interesting percentiles of Z's distribution.

The standard error of a simple slope is given by:

$$SE_{f(Z)} = \sqrt{SE^2_{\beta_1} + 2Z \times cov(\beta_1, \beta_3) + Z^2 SE^2_{\beta_3}}$$

So, you can test the significance of a simple slope by constructing a t-statistic or confidence interval using $\hat{f}(Z)$ and $SE_{f(Z)}$:

$$t = \frac{\hat{f}(Z)}{SE_{f(Z)}}, \quad CI = \hat{f}(Z) \pm t_{crit} \times SE_{f(Z)}$$

Example

We can use **semTools** routines to probe interaction in **lavaan** models.

- `probe2WayMC()`: simple slopes/intercepts analysis
- `plotProbe()`: simple slopes plots

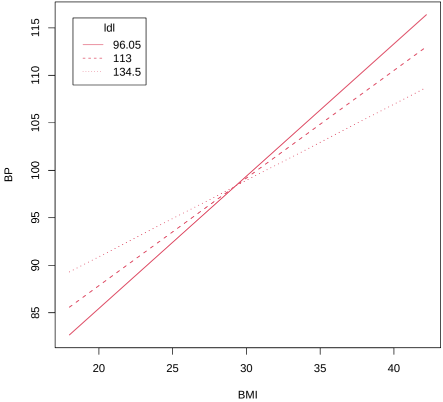
```
library(semTools)

## Estimate and test simple slopes and simple intercepts:
ssOut <- probe2WayMC(lavOut,
                     nameX = c("bmi", "ldl", "bmi:ldl"),
                     nameY = "bp",
                     modVar = "ldl",
                     valProbe = quantile(dDat$ldl, c(0.25, 0.50, 0.75))
)
```

Example

```
## View the results:
ssOut

$SimpleIntcept
        ldl    est    se      z pvalue
25%  96.05 57.585 4.017 14.334      0
50% 113.00 65.192 3.736 17.449      0
75% 134.50 74.840 4.944 15.139      0

$SimpleSlope
        ldl   est    se     z pvalue
25%  96.05 1.393 0.156 8.942      0
50% 113.00 1.133 0.140 8.107      0
75% 134.50 0.803 0.178 4.508      0
```

Example

```
## Plot the simple slopes:
plotProbe(ssOut, xlim = range(dDat$bmi), xlab = "BMI", ylab = "BP")
```



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

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). New York: Guilford Press.

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