

1 What are Mediation and Moderation?

1.1 Getting Started

2 Mediation Analyses

2.1 Example Mediation Data

2.2 Method 1: Baron & Kenny

2.3 Interpreting Barron & Kenny Results

2.4 Method 2: The Mediation Pacakge Method

2.5 Interpreting Mediation Results

3 Moderation Analyses

3.1 Example Moderation Data

3.2 Moderation Analysis

3.3 Interpreting Moderation Results

4 References and Further Reading

Chapter 14: Mediation and Moderation

[Code ▼](#)

Alyssa Blair

1 What are Mediation and Moderation?

Mediation analysis tests a hypothetical causal chain where one variable X affects a second variable M and, in turn, that variable affects a third variable Y . Mediators describe the how or why of a (typically well-established) relationship between two other variables and are sometimes called intermediary variables since they often describe the process through which an effect occurs. This is also sometimes called an indirect effect. For instance, people with higher incomes tend to live longer but this effect is explained by the mediating influence of having access to better health care.

In R, this kind of analysis may be conducted in two ways: Baron & Kenny's (1986) 4-step indirect effect method and the more recent *mediation* package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). The Baron & Kelly method is among the original methods for testing for mediation but tends to have low statistical power. It is covered in this chapter because it provides a very clear approach to establishing relationships between variables and is still occassionally requested by reviewers. However, the *mediation* package method is highly recommended as a more flexible and statistically powerful approach.

Moderation analysis also allows you to test for the influence of a third variable, Z , on the relationship between variables X and Y . Rather than testing a causal link between these other variables, moderation tests for when or under what conditions an effect occurs. Moderators can strength, weaken, or reverse the nature of a relationship. For example, academic self-efficacy (confidence in own's ability to do well in school) moderates the relationship between task importance and the amount of test anxiety a student feels (Nie, Lau, & Liau, 2011). Specifically, students with high self-efficacy experience less anxiety on important tests than students with low self-efficacy while all students feel relatively low anxiety for less important tests. Self-efficacy is considered a moderator in this case because it interacts with task importance, creating a different effect on test anxiety at different levels of task importance.

In general (and thus in R), moderation can be tested by interacting variables of interest (moderator with IV) and plotting the simple slopes of the interaction, if present. A variety of packages also include functions for testing moderation but as the underlying statistical approaches are the same, only the “by hand” approach is covered in detail in here.

Finally, this chapter will cover these basic mediation and moderation techniques only. For more complicated techniques, such as multiple mediation, moderated mediation, or mediated moderation please see the *mediation* package's full documentation.

1.1 Getting Started

If necessary, review the Chapter on regression. Regression test assumptions may be tested with *gvlma*. You may load all the libraries below or load them as you go along. Review the help section of any packages you may be unfamiliar with `?(packagename)`.

```
library(mediation) #Mediation package
library(rockchalk) #Graphing simple slopes; moderation
library(multilevel) #Sobel Test
library(bda) #Another Sobel Test option
library(gvlma) #Testing Model Assumptions
library(stargazer) #Handy regression tables

#Useful Help
?lm
?mediation
```

```
## No documentation for 'mediation' in specified packages and libraries:
## you could try '??mediation'
```

```
?rockchalk
?stargazer

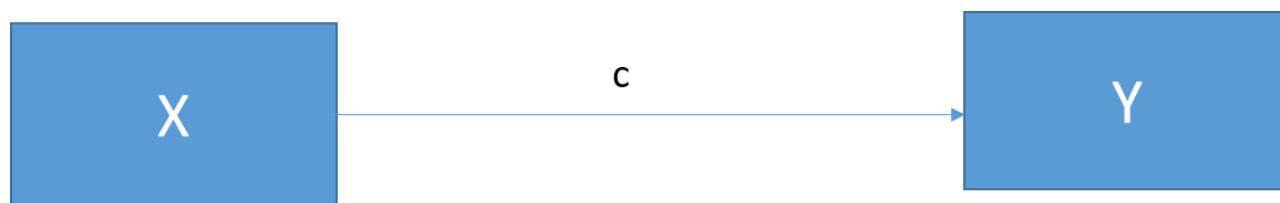
#Optional packages
library(QuantPsyc)
library(pequod)
?moderate.lm
?pequod
```

```
## No documentation for 'pequod' in specified packages and libraries:
## you could try '??pequod'
```

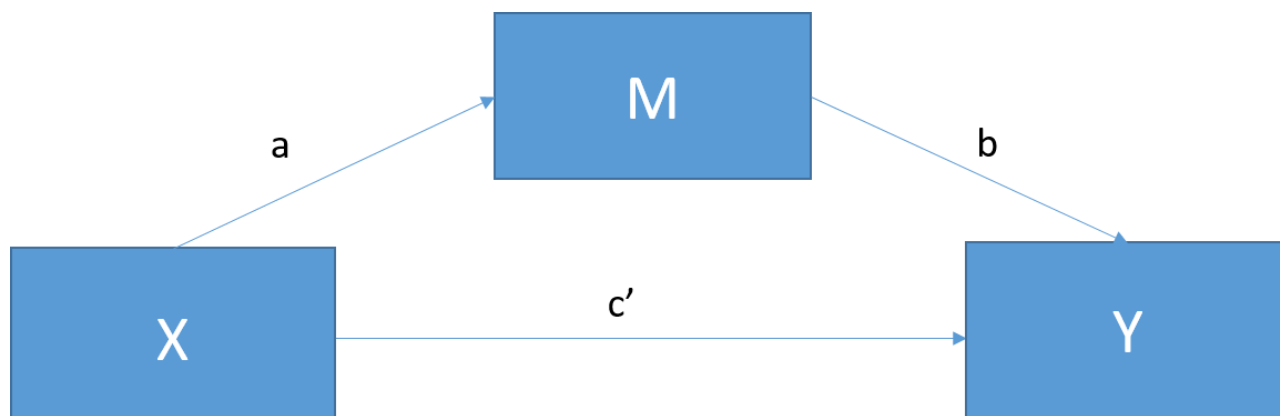
2 Mediation Analyses

Mediation tests whether the effects of X (the independent variable) on Y (the dependent variable) operate through a third variable, M (the mediator). In this way, mediators explain the causal relationship between two variables or “how” the relationship works, making it a very popular method in psychological research.

Both mediation and moderation assume that there is little to no measurement error in the mediator/moderator variable and that the DV **did not CAUSE** the mediator/moderator. If mediator error is likely to be high, researchers should collect multiple indicators of the construct and use SEM to estimate latent variables. The safest ways to make sure your mediator is not caused by your DV are to experimentally manipulate the variable or collect the measurement of your mediator before you introduce your IV.



Total Effect Model.



Basic Mediation Model.

c = the total effect of X on Y $c = c' + ab$ c' = the direct effect of X on Y after controlling for M; $c' = c - ab$
 ab = indirect effect of X on Y

The above shows the standard mediation model. Perfect mediation occurs when the effect of X on Y decreases to 0 with M in the model. Partial mediation occurs when the effect of X on Y decreases by a nontrivial amount (the actual amount is up for debate) with M in the model.

2.1 Example Mediation Data

Set an appropriate working directory and generate the following data set.

In this example we'll say we are interested in whether the number of hours since dawn (X) affect the subjective ratings of wakefulness (Y) 100 graduate students through the consumption of coffee (M).

Note that we are intentionally creating a mediation effect here (because statistics is always more fun if we have something to find) and we do so below by creating M so that it is related to X and Y so that it is related to M. This creates the causal chain for our analysis to parse.

```
#setwd("user location") #Working directory
set.seed(123) #Standardizes the numbers generated by rnorm; see Chapter 5
N <- 100 #Number of participants; graduate students
X <- rnorm(N, 175, 7) #IV; hours since dawn
M <- 0.7*X + rnorm(N, 0, 5) #Suspected mediator; coffee consumption
Y <- 0.4*M + rnorm(N, 0, 5) #DV; wakefulness
Meddata <- data.frame(X, M, Y)
```

2.2 Method 1: Baron & Kenny

This is the original 4-step method used to describe a mediation effect. Steps 1 and 2 use basic linear regression while steps 3 and 4 use multiple regression. For help with regression, see Chapter 10.

The Steps: 1. Estimate the relationship between X on Y (hours since dawn on degree of wakefulness) - Path “c” must be significantly different from 0; must have a total effect between the IV & DV

2. Estimate the relationship between X on M (hours since dawn on coffee consumption) -Path “a” must be significantly different from 0; IV and mediator must be related.
3. Estimate the relationship between M on Y controlling for X (coffee consumption on wakefulness, controlling for hours since dawn) -Path “b” must be significantly different from 0; mediator and DV must be related. -The effect of X on Y decreases with the inclusion of M in the model
4. Estimate the relationship between Y on X controlling for M (wakefulness on hours since dawn, controlling for coffee consumption) -Should be non-significant and nearly 0.

```
#1. Total Effect
fit <- lm(Y ~ X, data=Meddata)
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ X, data = Meddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.917   -3.738   -0.259    2.910   12.540
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.88368   14.26371   1.394   0.1665
## X           0.16899    0.08116   2.082   0.0399 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.16 on 98 degrees of freedom
## Multiple R-squared:  0.04237,    Adjusted R-squared:  0.0326
## F-statistic: 4.336 on 1 and 98 DF,  p-value: 0.03993
```

```
#2. Path A (X on M)
fita <- lm(M ~ X, data=Meddata)
summary(fita)
```

```
##
## Call:
## lm(formula = M ~ X, data = Meddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.5367 -3.4175 -0.4375  2.9032 16.4520
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.04494   13.41692   0.451   0.653
## X            0.66252    0.07634   8.678 8.87e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.854 on 98 degrees of freedom
## Multiple R-squared:  0.4346, Adjusted R-squared:  0.4288
## F-statistic: 75.31 on 1 and 98 DF, p-value: 8.872e-14
```

```
#3. Path B (M on Y, controlling for X)
fitb <- lm(Y ~ M + X, data=Meddata)
summary(fitb)
```

```
##
## Call:
## lm(formula = Y ~ M + X, data = Meddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3651 -3.3037 -0.6222  3.1068 10.3991
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.32177   13.16216   1.316   0.191
## M            0.42381    0.09899   4.281 4.37e-05 ***
## X           -0.11179    0.09949  -1.124   0.264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.756 on 97 degrees of freedom
## Multiple R-squared:  0.1946, Adjusted R-squared:  0.1779
## F-statistic: 11.72 on 2 and 97 DF, p-value: 2.771e-05
```

```
#4. Reversed Path C (Y on X, controlling for M)
fitc <- lm(X ~ Y + M, data=Meddata)
summary(fitc)
```

```
##
## Call:
## lm(formula = X ~ Y + M, data = Meddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.438  -2.573  -0.030   3.010  11.779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  96.11234    9.27663   10.361 < 2e-16 ***
## Y            -0.11493    0.10229   -1.124    0.264
## M             0.69619    0.08356    8.332 5.27e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.823 on 97 degrees of freedom
## Multiple R-squared:  0.4418, Adjusted R-squared:  0.4303
## F-statistic: 38.39 on 2 and 97 DF,  p-value: 5.233e-13
```

```
#Summary Table
stargazer(fit, fita, fitb, fitc, type = "text", title = "Baron and Kenny Method")
```

```

##
## Baron and Kenny Method
## =====
##
##                                     Dependent variable:
## -----
##
##                                     Y           M           Y
X                                     (1)         (2)         (3)
##                                     (4)
## -----
## Y
-0.115
##
(0.102)
##
## M                                     0.424***
0.696***
##                                     (0.099)
(0.084)
##
## X                                     0.169**      0.663***      -0.112
##                                     (0.081)      (0.076)      (0.099)
##
## Constant      19.884      6.045      17.322
96.112***
##                                     (14.264)      (13.417)      (13.162)
(9.277)
##
## -----
## Observations      100      100      100
100
## R2      0.042      0.435      0.195
0.442
## Adjusted R2      0.033      0.429      0.178
0.430
## Residual Std. Error  5.160 (df = 98)      4.854 (df = 98)      4.756 (df = 97)
4.823 (df = 97)
## F Statistic      4.336** (df = 1; 98) 75.313*** (df = 1; 98) 11.715*** (df = 2;
97) 38.389*** (df = 2; 97)
## =====
## Note:
p<0.1; **p<0.05; ***p<0.01

```

2.3 Interpreting Barron & Kenny Results

Here we find that our total effect model shows a significant positive relationship between hours since dawn (X) and wakefulness (Y). Our Path A model shows that hours since dawn (X) is also positively related to coffee consumption (M). Our Path B model then shows that coffee consumption (M) positively predicts wakefulness (Y) when controlling for hours since dawn (X). Finally, wakefulness (Y) does not predict hours since dawn (X) when controlling for coffee consumption (M).

Since the relationship between hours since dawn and wakefulness is no longer significant when controlling for coffee consumption, this suggests that coffee consumption does in fact mediate this relationship. However, this method alone does not allow for a formal test of the indirect effect so we don't know if the change in this relationship is truly meaningful.

There are two primary methods for formally testing the significance of the indirect test: the Sobel test & bootstrapping (covered under the *mediation* method).

```
#Sobel Test
library(multilevel)
?sobel
sobel(Meddata$X, Meddata$M, Meddata$Y)
```

```
## $`Mod1: Y~X`
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 19.8836805 14.2637142  1.394004 0.16646905
## pred         0.1689931  0.0811601  2.082220 0.03992761
##
## $`Mod2: Y~X+M`
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 17.3217682 13.16215851  1.316028 1.912663e-01
## pred        -0.1117904  0.09949262 -1.123605 2.639537e-01
## med          0.4238113  0.09899469  4.281152 4.371472e-05
##
## $`Mod3: M~X`
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  6.0449365 13.41692114  0.4505457 6.533122e-01
## pred          0.6625203  0.07634187  8.6783345 8.871741e-14
##
## $Indirect.Effect
## [1] 0.2807836
##
## $SE
## [1] 0.07313234
##
## $z.value
## [1] 3.83939
##
## $N
## [1] 100
```

```
#or
library(bda)
mediation.test(M,X,Y)
```



```
##              Sobel      Aroian      Goodman
## z.value 3.8393902040 3.8190525305 3.8600562907
## p.value 0.0001233403 0.0001339652 0.0001133609
```

The Sobel Test uses a specialized t-test to determine if there is a significant reduction in the effect of X on Y when M is present. Using the `sobel` function of the *multilevel* package will show provide you with three of the basic models we ran before (Mod1 = Total Effect; Mod2 = Path B; and Mod3 = Path A) as well as an estimate of the indirect effect, the standard error of that effect, and the z-value for that effect. You can either use this value to calculate your p-value or run the `mediation.test` function from the *bda* package to receive a p-value for this estimate.

In this case, we can now confirm that the relationship between hours since dawn and feelings of wakefulness are significantly mediated by the consumption of coffee ($z' = 3.84$, $p < .001$).

However, the Sobel Test is largely considered an outdated method since it assumes that the indirect effect (ab) is normally distributed and tends to only have adequate power with large sample sizes. Thus, again, it is highly recommended to use the *mediation* bootstrapping method instead.

2.4 Method 2: The *Mediation* Pacakge Method

This package uses the more recent bootstrapping method of Preacher & Hayes (2004) to address the power limitations of the Sobel Test. This method computes the point estimate of the indirect effect (ab) over a large number of random sample (typically 1000) so it does not assume that the data are normally distributed and is especially more suitable for small sample sizes than the Barron & Kenny method.

To run the *mediate* function, we will again need a model of our IV (hours since dawn), predicting our mediator (coffee consumption) like our Path A model above. We will also need a model of the direct effect of our IV (hours since dawn) on our DV (wakefulness), when controlling for our mediator (coffee consumption). When can then use *mediate* to repeatedly simulate a comparsion between these models and to test the signifcance of the indirect effect of coffee consumption.

```
#Mediate package
library(mediation)
?mediate
fitM <- lm(M ~ X,      data=Meddata) #IV on M; Hours since dawn predicting coffee consu
mption
fitY <- lm(Y ~ X + M, data=Meddata) #IV and M on DV; Hours since dawn and coffee predi
cting wakefulness
gvlma(fitM) #data is positively skewed; could log transform (see Chap. 10 on assumptio
ns)
```

```
##
## Call:
## lm(formula = M ~ X, data = Meddata)
##
## Coefficients:
## (Intercept)          X
##      6.0449      0.6625
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
##
## Call:
## gvlma(x = fitM)
##
##              Value p-value              Decision
## Global Stat      8.833 0.06542  Assumptions acceptable.
## Skewness         6.314 0.01198 Assumptions NOT satisfied!
## Kurtosis         1.219 0.26949  Assumptions acceptable.
## Link Function    1.076 0.29959  Assumptions acceptable.
## Heteroscedasticity 0.223 0.63674  Assumptions acceptable.
```

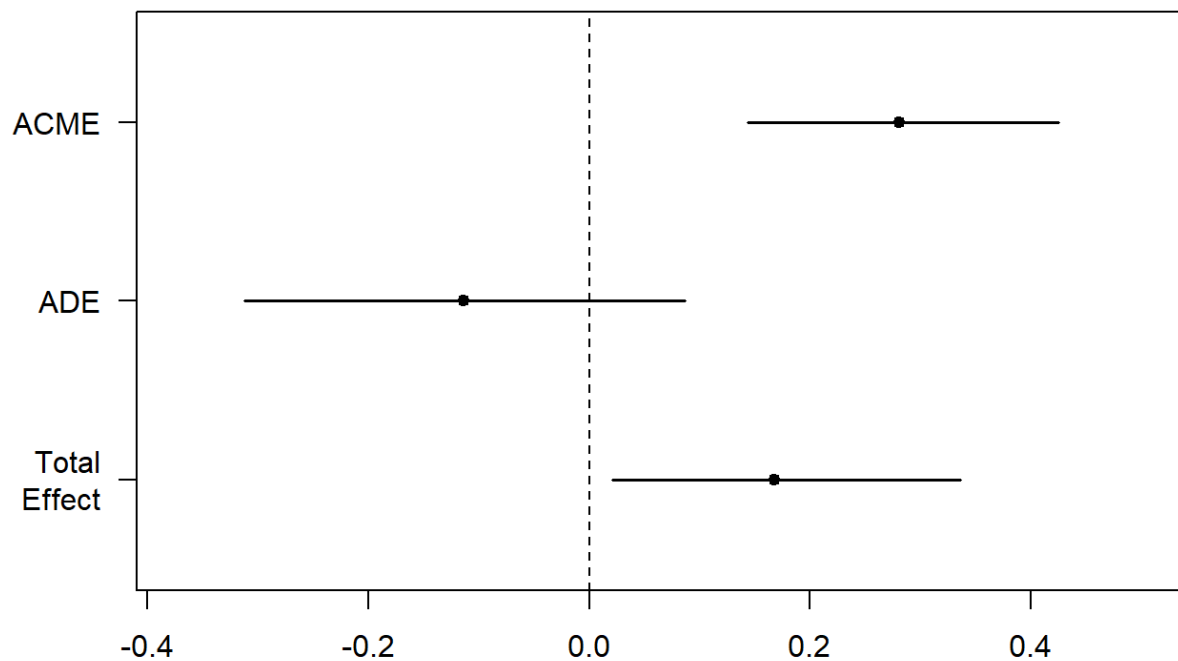
```
gvlma(fitY)
```

```
##
## Call:
## lm(formula = Y ~ X + M, data = Meddata)
##
## Coefficients:
## (Intercept)          X          M
##      17.3218      -0.1118      0.4238
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
##
## Call:
## gvlma(x = fitY)
##
##              Value p-value              Decision
## Global Stat      3.41844 0.4904 Assumptions acceptable.
## Skewness         1.85648 0.1730 Assumptions acceptable.
## Kurtosis         0.77788 0.3778 Assumptions acceptable.
## Link Function    0.71512 0.3977 Assumptions acceptable.
## Heteroscedasticity 0.06896 0.7929 Assumptions acceptable.
```

```
fitMed <- mediate(fitM, fitY, treat="X", mediator="M")
summary(fitMed)
```

```
##
## Causal Mediation Analysis
##
## Quasi-Bayesian Confidence Intervals
##
##           Estimate 95% CI Lower 95% CI Upper p-value
## ACME           0.2808    0.1437    0.42 <2e-16 ***
## ADE            -0.1133   -0.3116    0.09  0.258
## Total Effect    0.1674    0.0208    0.34  0.028 *
## Prop. Mediated  1.6428    0.5631    8.44  0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 100
##
##
## Simulations: 1000
```

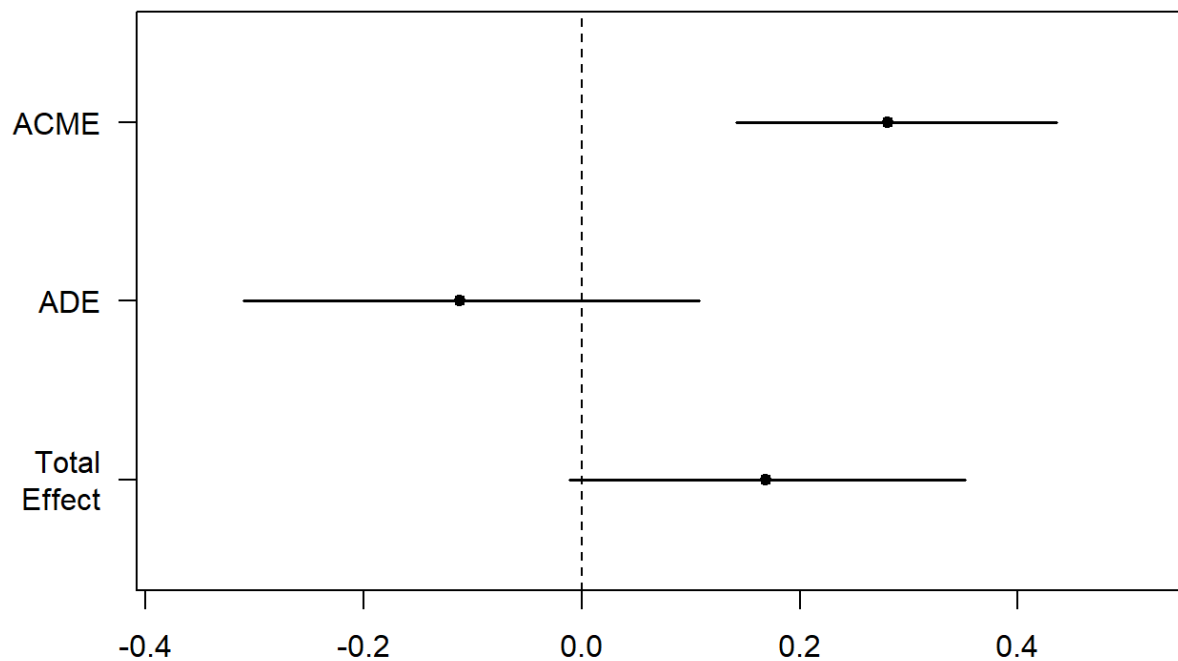
```
plot(fitMed)
```



```
#Bootstrap
fitMedBoot <- mediate(fitM, fitY, boot=TRUE, sims=999, treat="X", mediator="M")
summary(fitMedBoot)
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##           Estimate 95% CI Lower 95% CI Upper p-value
## ACME           0.2808    0.1420    0.44 <2e-16 ***
## ADE            -0.1118   -0.3099    0.11  0.280
## Total Effect    0.1690   -0.0112    0.35  0.066 .
## Prop. Mediated  1.6615   -5.4019   11.54  0.066 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 100
##
##
## Simulations: 999
```

```
plot(fitMedBoot)
```



2.5 Interpreting *Mediation* Results

The *mediate* function gives us our Average Causal Mediation Effects (ACME), our Average Direct Effects (ADE), our combined indirect and direct effects (Total Effect), and the ratio of these estimates (Prop. Mediated). The ACME here is the indirect effect of M (total effect - direct effect) and thus this value tells us if our mediation effect is significant.

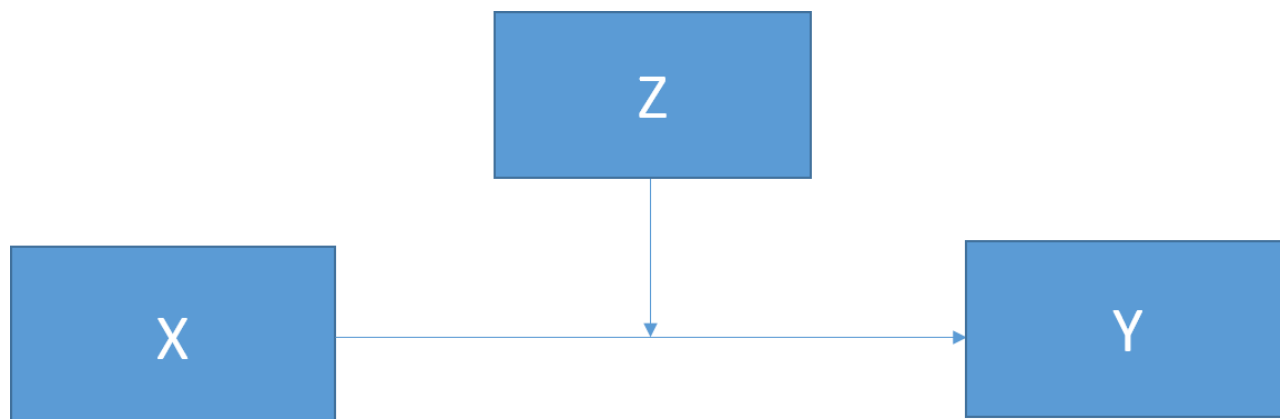
In this case, our **fitMed** model again shows a significant affect of coffee consumption on the relationship between hours since dawn and feelings of wakefulness, (ACME = .28, $p < .001$) with no direct effect of hours since dawn (ADE = -0.11, $p = .27$) and significant total effect ($p < .05$).

We can then bootstrap this comparison to verify this result in **fitMedBoot** and again find a significant mediation effect (ACME = .28, $p < .001$) and no direct effect of hours since dawn (ADE = -0.11, $p = .27$). However, with increased power, this analysis no longer shows a significant total effect ($p = .08$).

3 Moderation Analyses

Moderation tests whether a variable (Z) affects the direction and/or strength of the relation between an IV (X) and a DV (Y). In other words, moderation tests for interactions that affect WHEN relationships between variables occur. Moderators are conceptually different from mediators (when versus how/why) but some variables may be a moderator or a mediator depending on your question. See the *mediation* package documentation for ways of testing more complicated mediated moderation/moderated mediation relationships.

Like mediation, moderation assumes that there is little to no measurement error in the moderator variable and that the DV did not CAUSE the moderator. If moderator error is likely to be high, researchers should collect multiple indicators of the construct and use SEM to estimate latent variables. The safest ways to make sure your moderator is not caused by your DV are to experimentally manipulate the variable or collect the measurement of your moderator before you introduce your IV.



Basic Moderation Model.

3.1 Example Moderation Data

Set an appropriate working directory and generate the following data set.

In this example we'll say we are interested in whether the relationship between the number of hours of sleep (X) a graduate student receives and the attention that they pay to this tutorial (Y) is influenced by their consumption of coffee (Z). Here we create the moderation effect by making our DV (Y) the product of levels of the IV (X) and our moderator (Z).

```
#setwd("location") #Working directory
set.seed(123)#Standardizes the numbers generated by rnorm; see Chapter 5
N <- 100 #Number of participants; graduate students
X <- abs(rnorm(N, 6, 4)) #IV; Hours of sleep
X1 <- abs(rnorm(N, 60, 30)) #Adding some systematic variance for our DV
Z <- rnorm(N, 30, 8) #Moderator; Ounces of coffee consumed
Y <- abs((-0.8*X) * (0.2*Z) - 0.5*X - 0.4*X1 + 10 + rnorm(N, 0, 3)) #DV; Attention Pa
id
Moddata <- data.frame(X, X1, Z, Y)

summary(Moddata)
```

##	X	X1	Z	Y
## Min.	: 0.195	Min. : 1.597	Min. :15.95	Min. : 2.386
## 1st Qu.:	4.025	1st Qu.: 35.967	1st Qu.:25.75	1st Qu.: 30.155
## Median :	6.247	Median : 53.225	Median :30.29	Median : 47.761
## Mean :	6.483	Mean : 56.806	Mean :30.96	Mean : 47.763
## 3rd Qu.:	8.767	3rd Qu.: 74.035	3rd Qu.:36.11	3rd Qu.: 61.727
## Max.	:14.749	Max. :157.231	Max. :48.34	Max. :136.947

3.2 Moderation Analysis

Moderation can be tested by looking for significant interactions between the moderating variable (Z) and the IV (X). Notably, it is important to mean center both your moderator and your IV to reduce multicollinearity and make interpretation easier. Centering can be done using the *scale* function, which subtracts the mean of a variable from each value in that variable. For more information on the use of centering, see ?scale and any number of statistical textbooks that cover regression (we recommend Cohen, 2008).

A number of packages in R can also be used to conduct and plot moderation analyses, including the *moderate.lm* function of the *QuantPsyc* package and the *pequod* package. However, it is simple to do this “by hand” using traditional multiple regression, as shown here, and the underlying analysis (interacting the moderator and the IV) in these packages is identical to this approach. The *rockchalk* package used here is one of many graphing and plotting packages available in R and was chosen because it was especially designed for use with regression analyses (unlike the more general graphing options described in Chapters 8 & 9).

```
#Centering Data
Xc <- c(scale(X, center=TRUE, scale=FALSE)) #Centering IV; hours of sleep
Zc <- c(scale(Z, center=TRUE, scale=FALSE)) #Centering moderator; coffee consumpti
on

#Moderation "By Hand"
library(gvlma)
fitMod <- lm(Y ~ Xc + Zc + Xc*Zc) #Model interacts IV & moderator
summary(fitMod)
```

```
##
## Call:
## lm(formula = Y ~ Xc + Zc + Xc * Zc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.466  -8.972  -0.233   6.180  38.051
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.54443    1.17286  41.390 < 2e-16 ***
## Xc           5.20812    0.34870  14.936 < 2e-16 ***
## Zc           1.10443    0.15537   7.108 2.08e-10 ***
## Xc:Zc        0.23384    0.04134   5.656 1.59e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.65 on 96 degrees of freedom
## Multiple R-squared:  0.7661, Adjusted R-squared:  0.7587
## F-statistic: 104.8 on 3 and 96 DF,  p-value: < 2.2e-16
```

```
coef(summary(fitMod))
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 48.5444271 1.17285613 41.389925 5.149708e-63
## Xc           5.2081205 0.34870152 14.935755 8.862490e-27
## Zc           1.1044337 0.15537153  7.108340 2.077645e-10
## Xc:Zc        0.2338362 0.04134056  5.656338 1.592946e-07
```

```
gvlma(fitMod) #data is positively skewed; could log transform (see Chap. 10)
```

```
##
## Call:
## lm(formula = Y ~ Xc + Zc + Xc * Zc)
##
## Coefficients:
## (Intercept)          Xc          Zc          Xc:Zc
##    48.5444      5.2081      1.1044      0.2338
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
##
## Call:
## gvlma(x = fitMod)
##
##
```

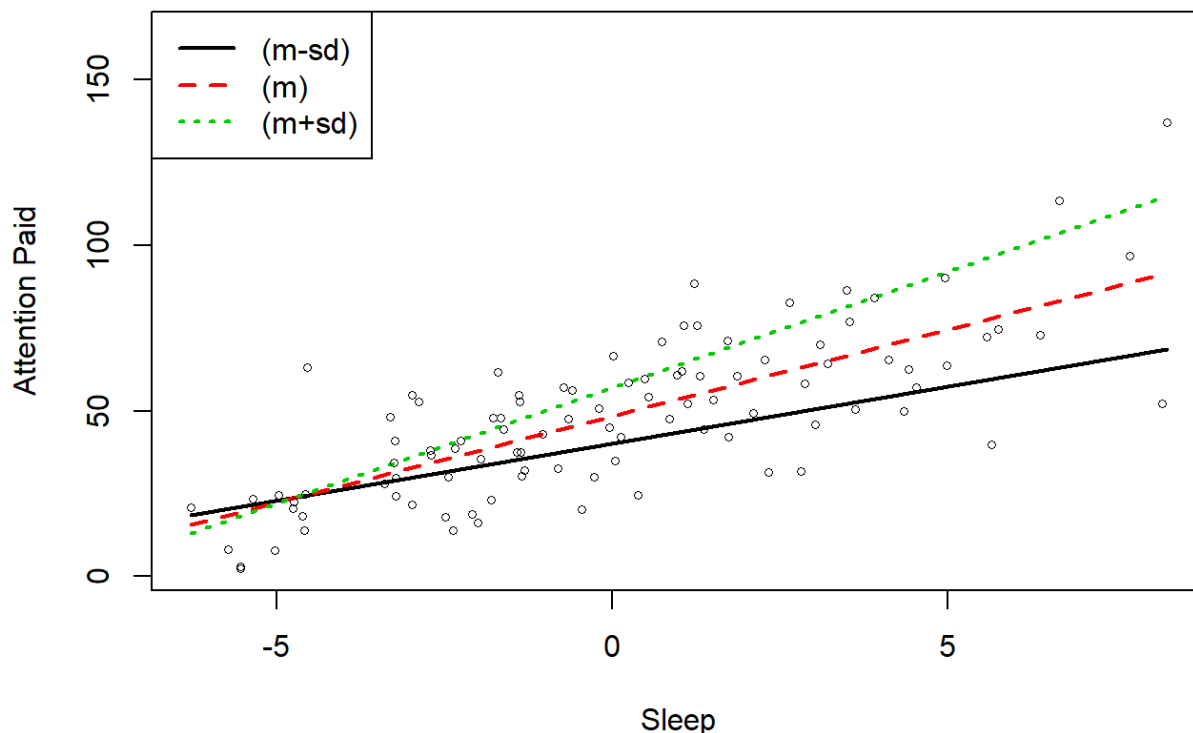
	Value	p-value	Decision
## Global Stat	7.68778	0.10371	Assumptions acceptable.
## Skewness	5.97432	0.01452	Assumptions NOT satisfied!
## Kurtosis	0.94082	0.33207	Assumptions acceptable.
## Link Function	0.73540	0.39114	Assumptions acceptable.
## Heteroscedasticity	0.03724	0.84698	Assumptions acceptable.

```
#Data Summary
library(stargazer)
stargazer(fitMod,type="text", title = "Sleep and Coffee on Attention")
```



```
##
## Sleep and Coffee on Attention
## =====
##                               Dependent variable:
##                               -----
##                               Y
## -----
## Xc                           5.208***
##                               (0.349)
##
## Zc                           1.104***
##                               (0.155)
##
## Xc:Zc                        0.234***
##                               (0.041)
##
## Constant                     48.544***
##                               (1.173)
##
## -----
## Observations                 100
## R2                           0.766
## Adjusted R2                  0.759
## Residual Std. Error         11.647 (df = 96)
## F Statistic                  104.784*** (df = 3; 96)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
#Plotting
library(rockchalk)
ps <- plotSlopes(fitMod, plotx="Xc", modx="Zc", xlab = "Sleep", ylab = "Attention Paid", modxVals = "std.dev")
```



3.3 Interpreting Moderation Results

Results are presented similar to regular multiple regression results (see Chapter 10). Since we have significant interactions in this model, there is no need to interpret the separate main effects of either our IV or our moderator.

Our by hand model shows a significant interaction between hours slept and coffee consumption on attention paid to this tutorial ($b = .23$, $SE = .04$, $p < .001$). However, we'll need to unpack this interaction visually to get a better idea of what this means.

The *rockchalk* function will automatically plot the simple slopes (1 SD above and 1 SD below the mean) of the moderating effect. This figure shows that those who drank less coffee (the black line) paid more attention with the more sleep that they got last night but paid less attention overall than average (the red line). Those who drank more coffee (the green line) paid more when they slept more as well and paid more attention than average. The difference in the slopes for those who drank more or less coffee shows that coffee consumption moderates the relationship between hours of sleep and attention paid.

4 References and Further Reading

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Cohen, B. H. (2008). *Explaining psychological statistics*. John Wiley & Sons.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4), 309.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, 7(1), 83.

Nie, Y., Lau, S., & Liao, A. K. (2011). Role of academic self-efficacy in moderating the relation between task importance and test anxiety. *Learning and Individual Differences*, 21(6), 736-741.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.