2018 ■ Vol. 14 ■ no. 2

A computational description of simple mediation analysis

Pier-Olivier Caron [a, ✉] and Philippe Valois [b]

[a]Télé-Université, Université du Québec, Département des Sciences humaines, Lettres et Communications
[b]Laboratoire des sciences appliquées du comportement et Laboratoire d'étude des troubles de l'ordre de la psychopathologie en enfance, Département de Psychologie, Université du Québec à Montréal and Centre d'études troubles obsessionnels-compulsifs et tics, Centre de recherche de l'Institut universitaire en santé mentale de Montréal

Abstract ■ Simple mediation analysis is an increasingly popular statistical analysis in psychology and in other social sciences. However, there is very few detailed account of the computations within the model. Articles are more often focusing on explaining mediation analysis conceptually rather than mathematically. Thus, the purpose of the current paper is to introduce the computational modelling within simple mediation analysis accompanied with examples with R. Firstly, mediation analysis will be described. Then, the method to simulate data in R (with standardized coefficients) will be presented. Finally, the bootstrap method, the Sobel test and the Baron and Kenny test all used to evaluate mediation (i.e., indirect effect) will be developed. The R code to implement the computation presented is offered as well as a script to carry a power analysis and a complete example.

Keywords ■ mediation analysis, indirect effect, power analysis. **Tools** ■ R.

✉ pocaron19@gmail.com

POC: 0000-0001-6346-5583; *PV*: 0000-0002-9594-2230

10.20982/tqmp.14.2.p147

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ One anonymous reviewer

Introduction

Mediation analysis is an increasingly popular statistical analysis in psychology and in other social sciences. It seeks to explain the (biological, psychological, cognitive, etc.) mechanism that underlies the relationship between an independent variable and a dependent variable by the inclusion of a third variable, i.e., the mediator variable. As mediation analysis becomes more and more popular, there is also an increasing body of scientific literature on the subject. However, very few detail the computation within the model. They are more often focusing on explaining conceptually rather than mathematically (see, for instance, Kane & Ashbaugh, 2017). Herein, this paper will adopt the latter approach to help the reader understand and apply the modelling within mediation analysis.

The purpose of the current article is to introduce the computational modelling within simple mediation analysis. It is worth noting that only simple mediation (a single mediator variable) will be presented, but that other forms of mediation (parallel, serial or moderated), may be understood by extending the presented formulas. The first part consists of the description of mediation analysis. Then, a method to simulate data (with standardized coefficients) will be presented. Finally, the bootstrap method used to evaluate mediation (i.e., indirect effect), the Baron and Kenny test and the Sobel test will be developed. The R code to implement the computation will be presented. For the sake of simplicity and without lack of generality, the presentation will mainly focus on standardized regression coefficients. The computation to unstandardize data will be presented. As a cautionary reminder, strong statistical analyses do not supersede strong theoretical framework and experimental design which are imperative when investigating potential mediating variable. Mediation analysis is useful, but must be used properly.

Simple mediation analysis

Mediation analysis is a subset of path analysis in which the researcher is interested in the relation between the independent variable (x) on the dependent variable (y) through the mediator variable (m). The path diagram corresponding to a simple mediation model is presented in the top panel of Figure 1. When there is no m, the existing relation

between x and y is said to be the total effect, represented by c_{xy}. It corresponds to the regression coefficient between x and y. The total effect can be divided into two other effects : the direct effect (c') and the indirect effect (ab). Deterministically, the indirect effect is the interpretation that x causes variability to m, which then causes variability to y. Mathematically speaking, the indirect effect is the product of the paths between x and m, and m and y (or paths a and b in top panel of Figure 1). The indirect effect is the effect of interest in mediation analysis. The other effect is the direct effect which is the relation remaining between x and y when the effect of m has been partialled out. As such, the more correct mathematical representation of c' is $c_{xy|m}$.

Mediation analysis can be seen as a regression analysis carried in two steps. The first step is to regress m on x to obtain the parameter a. Then, the second step is to regress y on x and m to obtain c' and b respectively. Finally, the product ab is tested to see if it is statistically different from 0 which would support a mediating effect. As it should become apparent, top panel of Figure 1, even though it is widespread, is conceptually ambiguous and can be misleading. For instance, a and c are simple coefficients whereas b and c' are partial coefficients. We will thus more clearly defined each parameter in the mediation model. Bottom panel of Figure 1 depicts the mediation models with the more appropriately labelled parameters. The path a is more appropriately the path a_{xm} which represents the correlation between x and m. As already pointed out for the relations between x and y, there is a total effect, c_{xy}, and the direct effect $c_{xy|m}$. The parameter b usually presented in mediation analysis is $b_{my|x}$, that is, the relation between m and y when controlling for the effect of x. There is also a parameter for the relation between m and y, b_{my}, which exists but is neglected, because it plays no role in the interpretation of mediation analysis. Both are dependent from one another with the partial correlation equation :

$$b_{my} = b_{my|x}\left(1 - a_{xm}^2\right) + a_{xm}c_{xy} \tag{1}$$

or consequently ;

$$b_{my|x} = \frac{b_{my} - a_{xm}c_{xy}}{(1 - a_{xm}^2)} \tag{2}$$

Finally, there is the indirect effect ab, which is the product of a_{xm} and $b_{my|x}$. The indirect effect, ab, and the direct effect, $c_{xy|m}$, sum to the total effect c_{xy}. Hence mathematically, $c_{xy} = c_{xy|m} + a_{xm} \times b_{my|x}$. In order to simulate a mediation model, three parameters must be known and defined because a, b and c are interrelated. To help illustrate, the next section explains how to generate data containing mediation.

Generating data

Modelling of the data is presented using, as it was previously mentioned, standardized coefficients. Parameters could be any value between -1 and 1. In mediation analysis, there are two predictors (x and m) and two dependent variables (m and y). In order to generate data, we must first generate data for X (capital letters represent data). Let X be a normally distributed variable with a mean of 0 and variance of 1, $X \sim \mathcal{N}(0, 1)$, then generate M with is computed by

$$M = a_{xm}X + e_m \tag{3}$$

where e_m is the error in M (i.e., $\text{var}(e_m)$ is the variance of the residual). The structural equation modelling of the mediation analysis (showing the error parameters) is presented in the bottom panel of Figure 1. Residual error has a mean of 0 and, to keep variance to 1, the error variance, e_m, is set to :

$$\text{var}(e_m) = 1 - a_{xm}^2 \tag{4}$$

so that M is normally distributed, $M \sim \mathcal{N}(0, 1)$. Because, the variance is additive, to get a variance equals to 1, the variance of other sources have to be subtracted. Finally, Y is generated in the following manner

$$Y = c_{xy|m}X + b_{my|x}M + \text{sqrt}(e_y) \tag{5}$$

which corresponds to the second regression analysis of mediation analysis. The variance of the error term of Y, e_y, is computed by

$$\text{var}\left(e_y\right) = 1 - (c_{xy|m}^2 + b_{my|x}^2 + 2a_{xm}c_{xy|m}b_{my|x}) \tag{6}$$

so that Y follows a normal distribution, $Y \sim \mathcal{N}(0, 1)$. The first two terms refer to the coefficients in equation 5 and the last one refers to the covariance between x and m (that is, the sum of two correlated variables is the sum of their variance plus twice their covariance; Howell, 2012). Equation 6 comes from the fact that the sum of two normally distributed correlated random variables is

$$\text{var}\left(x + m\right) = \text{var}\left(x\right) + \text{var}\left(m\right) + 2\text{cov}\left(xm\right) \tag{7}$$

Listing 1 shows the code to implement the generation of data in R with $a_{xm} = .50$, $b_{my|x} = .60$, $c_{xy} = .000$. From equation 1, we can compute b_{my} which is .45. The mediation model is presented in Figure 2. The resulting variance-covariance matrix is showed in Table 1. The covariance matrix is approximately the same as a correlation matrix in this case. Since data contain some error, it is only approximately the same. Given that the sample size was 10^6, results are strongly accurate. As such, the above values were true to the population parameters. It is worth noting that the variance of each variable is very close to 1.000 as expected from equations 3 to 6.

Figure 1 ■ Illustration of mediation analysis. Top panel depicts the usual diagram describing mediation. Bottom panel shows the parameters with a more appropriate notation which is used throughout the current paper. It depicts a mediation analysis from a structural equation modelling perspective as it includes error parameters.

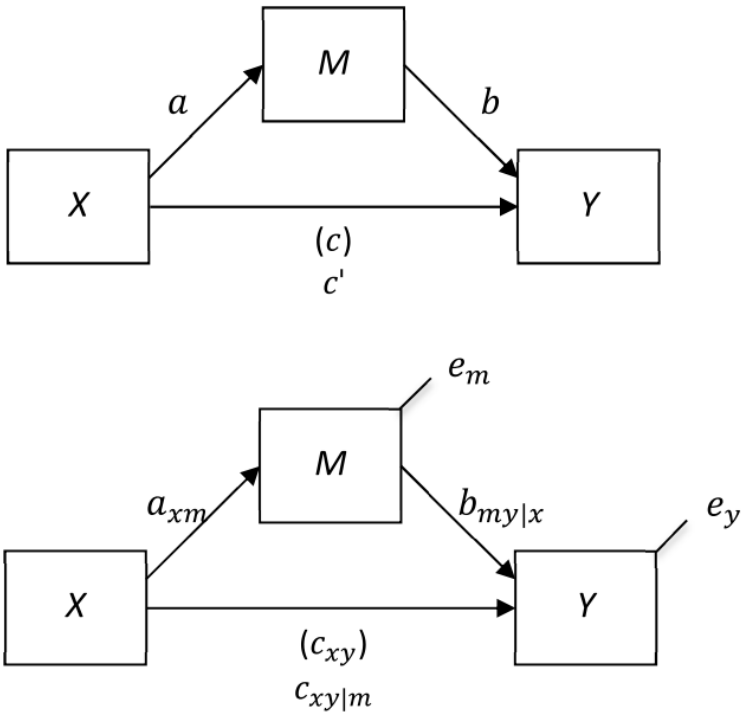


Table 1 ■ Variance-Covariance matrix of simulated data [a]

Variables	X	M	Y
X	1.001		
M	.495	1.002	
Y	-.006	0.451	1.003

Note. [a] obtained with the function `var()`

To unstandardized data, if needed, the data contained in a standardized variable (X, M, or Y) after being computed must be multiplied by the desired standard deviation (square root of the variance, σ^2) and the mean, μ, has to be added, such as, for the variable x :

$$X_{unstd} = \sigma_X X_{std} + \mu_X \tag{8}$$

in which x_{unstd}represents unstandardized data and x_{std}refers to standardized data. One could also use the code in Listing 1 to generate unstandardized data by specifying means and standard deviations.

Hypothesis testing

There are three ways to determine if ab is statistically significant. The first is the Baron and Kenny (1986) method, which is a three-step regression analysis. The first step is to check if the relation between x and y, that is c_{xy}, is significant, meaning there is a relation to be potentially explained by a mediator. The second step is to check if a_{xm} is significant, or testing if there is a relation between the mediator and the predictor. Finally, the last step is to regress y on x and m to obtain $b_{my|x}$ and $c_{xy|m}$. If $b_{my|x}$ is significant then the method suggests that a mediation process occurred. If $c_{xy|m}$ no longer is significant (compared to c_{xy}), the mediation is said to be complete, otherwise it is deemed partially mediated. We offer a R script to carry out the Baron and Kenny method in Listing 2. The Baron and Kenny method has been left out of favor because of its inappropriate assumptions, mostly on whether the hierar-

Listing 1 ■ Code to generate data. The figure shows the R code to generate data according to a simple mediation model with standardized parameters a_{xm}, $b_{my|x}$ and c_{xy} defined.

```
GenerateMediationData <- function(n = 1000, a = .50, b =.60, c =.00, mean.x = 0,
    sd.x = 1, mean.m = 0, sd.m = 1, mean.y = 0, sd.y = 1) {
  # a is a_xm
  # b is b_my|x
  # c is c_xy
  # mean.x, sd.x, mean.m, sd.m, mean.y and sd.y will create unstandardized data according to the specified
  # means and standard deviations
  if(missing(a) | missing(b) | missing(c)){
    stop("One or more arguments are missing")
  }
  ab <- a*b
  cp <- c-ab          # cp = c' = c_xy|m
  ey <- 1-(cp^2 +b^2 + 2*a*cp*b)
  if ((ey < 0) | (ey > 1)){print("WARNING : Sum of square of coefficients is too
   high to generate standardized data")}
  # Generate data
  x <- rnorm(n, mean = 0, sd = 1)
  em <- sqrt(1-a^2)
  m <- a*x + em*rnorm(n, mean = 0, sd = 1)
  ey2 <- sqrt(ey)
  y <- cp*x + b*m + ey2*rnorm(n, mean = 0, sd = 1)
  x <- x * sd.x + mean.x
  m <- m * sd.m + mean.m
  y <- y * sd.y + mean.y
  data <- as.data.frame(cbind(x, m, y))
  return(data)
}
```

chical steps have to be followed, and the rise of newer and more powerful statistical techniques (Hayes, 2013).

The second test to assess mediation is the Sobel test, which is a z-distributed statistic computed from the indirect effect as

$$z = \frac{a_{xm} b_{my|x}}{SE} \tag{9}$$

where SE is the standard error of the indirect effect computed with the following equation :

$$SE = \sqrt{\left(a_{xm}^2 s_{b_{my|x}}^2 + b_{my|x}^2 s_{a_{xm}}^2\right)} \tag{10}$$

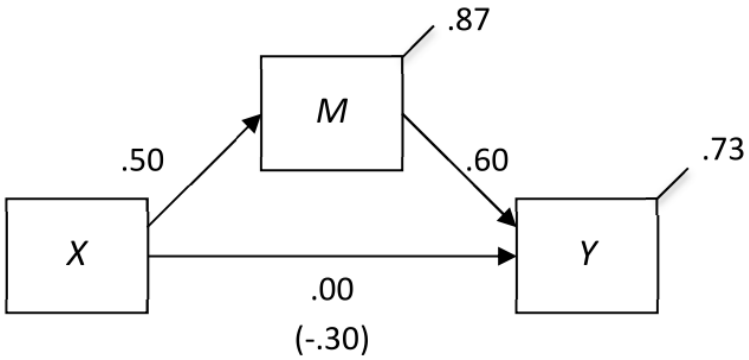
and where s_i^2 represents the variance of the path i, $i = a_{xm}, b_{mx|y}$. Listing 3 shows the R code to implement the Sobel test. This test has the assumption that the product of two correlation coefficients is normally distributed, which is not always true in practice. Consequently, it is less powerful than the last method, which is the bootstrap method, emphasized by Hayes (2013). The bootstraps test resamples data in order to build a 95% confidence interval (or any percentage actually) of the indirect effect and test if it entails the null hypothesis (i.e., the indirect effect is 0). As it is a bootstrap method, it is free from the statistical distribution assumption (more robustness) compared to the Sobel test, because even if data is normally distributed, this is not necessarily true for the indirect effect, and is more powerful (less type II error) than the Baron and Kenny test and the Sobel test (Preacher, Rucker, & Hayes, 2007).

The bootstrap method (Efron & Tibshirani, 1979) is a computer-based method which treats the sample as a pseudo-population (that is, the sample distributions reflect the population distribution). It randomly selects with replacement subjects of the original sample in order to generate another sample and compute the desired statistics. Then, it repeatedly does this last step a tremendous amount of time (for instance, a general recommendation is over 5 000) in order to create an empirical sampling distribution of the desired statistics. Confidence intervals can be computed from the sampling distribution and inference regarding hypothesis testing can be done. Bootstrapping is easily implemented in R. The bias-corrected and accel-

Listing 2 ■ Code for Baron & Kenny Test.

```
BaronKenny <- function(x, m, y, data, alpha = 0.05) {
  # x is the column name of the predictor in data
  # m is the column name of the mediator in data
  # y is the column name of the dependent variable in data
  # data is a data.frame or a matrix that contain columns with the names of x, y and m.
  # If x, m or y are missing, data[,1] will be used for x, data[,2] will be
  # used for m and data[,3] will be used for y.

  if(missing(data)) {
    stop("There's no data")
  }

  if(is.data.frame(data) != TRUE & is.matrix(data)) {
    d <- as.data.frame(data)
  } else if (is.data.frame(data) != TRUE & is.matrix(data) != TRUE) {
    stop('"data" should be a matrix or a data.frame')
  }

  if(missing(x)){x <- data[,1]} else if (is.numeric(x) == TRUE) {x <- data[,x]}
    else {x <- data[,match(x, table = colnames(data))]}
  if(missing(m)){m <- data[,1]} else if (is.numeric(m) == TRUE) {m <- data[,m]}
    else {m <- data[,match(m, table = colnames(data))]}
  if(missing(y)){y <- data[,1]} else if (is.numeric(y) == TRUE) {y <- data[,y]}
    else {y <- data[,match(y, table = colnames(data))]}

  Sig <- FALSE
  out <- 'No Mediation'

  #regression 1
  step1 <- lm(formula = y ~ x)
  pC <- summary(step1)$coefficients[2,4]
  if (pC <= alpha){

    #regression 2
    step2 <- lm(formula = m ~ x)
    pA <- summary(step2)$coefficients[2,4]
    if (pA <= alpha){

      #regression 3
      step3 <- lm(formula = y ~ x + m)
      pB <- summary(step3)$coefficients[3,4]
      Sig <- (pB <= alpha)
      if (Sig){
        if (summary(step3)$coefficients[2,4] <= alpha){
          out <- 'This is a partial mediation'} else {
          out <- 'This is a complete mediation'}
        }
      }
    }
    return(list(sig=Sig, conclusion=out))
  }
```

Figure 2 ■ Illustration of the mediation for the example. The population parameters are also used for the power analysis.



erated (BCa) bootstrap interval is a method introduced to correct bias and skewness in the distribution of bootstrap estimates. Listing 4 shows the code to apply the bootstrap method to mediation analysis. It also uses an additional function to compute the indirect effect for the boot function that needs to be called in the primary function.

Power analysis

It might be also interesting to put the previous tutorial into practice. For instance, let us consider a power analysis to evaluate the type II error rate of `BootTest()`, `SolbelTest()` and `BaronKenny()` functions. Power refers to the probability to find a significant result when the null hypothesis is false (there is an indirect effect). Failure to find a significant result is a type II error. Listings 5 and 6 shows the code to implement a power analysis. The purpose of power analysis is to simulate an experiment with known and non-null population parameters, check whether the result is significant or not, and redo the above a tremendous amount of times. There are two main components in the script: the generation of data (Listing 1) and the indirect effect test (Listings 2 to 4). The outcome of the function is the power of the mediation test given a sample size n.

To conclude this section, three power analyses were carried out following the parameters of the previous example with a sample size of 40. Table 2 shows the results of the power analysis of the three tests. The Baron and Kenny test had a poor performance (power of .029), because of the really low (null) total effect which is a tricky scenario for that test. The Sobel test had a power of 0.606. Finally, the bootstrap method obtained a power of 0.786. To sum up, the results demonstrate the lack of power of the Baron and Kenny test and the Sobel test, and the more powerful estimation of the BootTest.

A complete example

In order to illustrate mediation analysis, a complete example will be carried. Listing 7 shows the complete script to run the example. Four hundred twenty-nine people were asked to complete the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996) and a short survey that included questions about the average weekly alcoholic beverage intake (further referenced as alcohol intake) and number of weekly positive social interaction (further referenced as positive social interaction). The BDI is a short self-report questionnaire used to assess intensity of depression. The main hypothesis is the effect of the alcohol intake (the independent variable x) on depression (the dependent variable y) will be partially mediated by positive social interaction (the mediator m). Table 3 presents the population parameters of the example.

Table 4 shows the descriptive analysis and histogram with density curve (see Figure 3) for the three variables showed a normal distribution of data. These information can be found with the of the psych package (Revelle, 2017). Tables 5 presents the variance-covariance matrix with function `cov()` and correlation matrix with the `cor()` function. It is worth to note that the correlation matrix summarizes approximately the expected relations given by the population parameters. Table 6 depicts the first step of the mediation analysis conducted by testing a regression model of alcohol intake on positive social interaction (using the function `lm()` in R) with a significant model, $F(1, 427) = 67.72, p < .001$, and a significant effect of alcoholic intake over positive social interaction, $\beta = 0.189, p < .001$. Step two tests the regression model of alcohol intake and positive social interaction on depression (see table 6) found a significant model, $F(2, 426) = 38.08, p < .001$, and significant effects of alcoholic intake, $\beta = 0.824, p < .001$, and positive social

Table 2 ■ Summary of power analyses

Indirect test	Power
Baron & Kenny test	0.029
Sobel test	0.606
Bootstrap test	0.786

Table 3 ■ Population parameters of the complete example

Parameter	Value	Equation
a_{xm}	0.400	Fixed
$b_{my\|x}$	-0.350	Fixed [a]
b_{my}	0.247	$b_{my} = b_{my\|x}\left(1 - a_{xm}^2\right) + a_{xm}c_{xy}$
c_{xy}	0.250	Fixed [a]
$c_{xy\|m}$	0.390	$c_{xy} - ab$
ab	-0.140	$a_{xm} \times b_{my\|x}$
n	429	-

Note. [a] because their counterpart (b_{my} and $c_{xy|m}$) were fixed first.

interaction, $\beta = -1.629, p < .001$, over BDI score.

To test for the significant mediation effect, the three methods are used with the unstandardized data in order to demonstrate the non-necessity of using standardized dataset. Table 7 summarizes the results. All tests yield the same outcome (regardless whether data were standardized or not). The Baron & Kenny test suggests a significant partial mediation. The Sobel test shows a significant mediation, $z = -5.445, p < 0.001$, for both dataset. Finally, the bootstrap BCa confidence intervals had a lower limit of 1.043 and an upper limit of 1.754. The confidence interval does not include 0 and, therefore, the indirect effect is deemed significant. We could interpret the results as the number of weekly positive social interaction partially mediate the effect of weekly alcoholic beverage intake on depression by reducing the later scores, but these data were generated using the code provided in this article.

Discussion

The purpose of the current paper was to introduce the computation within mediation analysis. Firstly, we detailed the parameters in the conceptual diagram and labelled them appropriately. We then showed some examples using R and gave the code for the readers to implement it themselves. We hope this work will encourage statistical research in the analysis of mediation models and help the reader to better understand them.

Authors' note

We would like to thank Denis Cousineau and an anonymous reviewer for their commentaries on an earlier version of this manuscript.

References

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research : conceptual, strategic, and statistical consid-

Figure 3 ■ Distribution of the three variables using the `hist()` function.

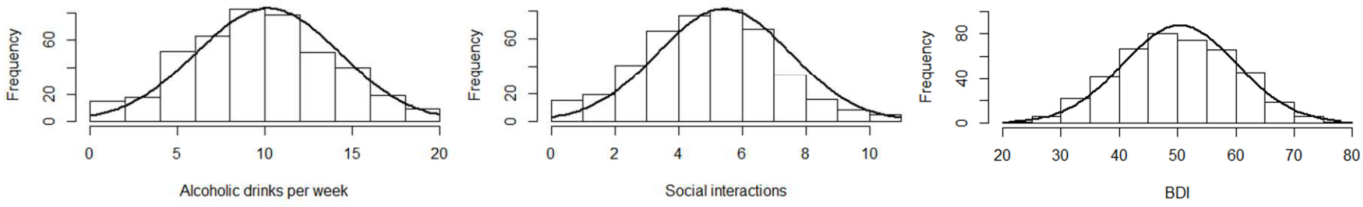


Table 4 ■ Descriptive analysis of the complete example

Variable	Mean	SD	Min	Max	Skew	Kurtosis	SE
Alcoholic intake	10.16	4.10	0	20	.01	-.35	.20
Positive social interaction	5.43	2.10	0	11	.04	-.06	.10
BDI score	50.39	9.79	21	78	-.04	-.49	.47

Note. SD : Standard deviation, Min : Minimum observed value, Max : Maximum observed value, Skew : Skewness, SE : Standard error.

Table 5 ■ Variance-Covariance [a] and correlation [b] matrices of the complete example

Variance-Covariance	Correlation					
Variables	AI	PSI	BDI	AI	PSI	BDI
AI	16.799			1.000		
PSI	3.179	4.395		0.370	1.000	
BDI	8.664	-4.538	95.804	0.216	-0.221	1.000

Note. [a] obtained with the function `cov()`, [b] obtained with the function `cor()`, PSI : Positive social interaction, AI : Positive social interaction, BDI : BDI score

erations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-ii. *San Antonio*, *78*(2), 490–8.

Efron, B. & Tibshirani, R. (1979). *An introduction to the bootstrap*. New York (NY): Chapman & Hall.

Hayes, A. F. (2013). *Introduction to mediation, moderation and conditional process analysis*. New York (NY): Guildford.

Howell, D. C. (2012). *Statistical methods for psychology*. Belmont: Wadsworth.

Kane, L. & Ashbaugh, A. R. (2017). Simple and parallel mediation : a tutorial exploring anxiety sensitivity, sensation seeking, and gender. *The Quantitative Methods for Psychology*, *13*, 148–165. doi:10.20982/tqmp.13.3.p148

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Assessing moderated mediation hypotheses : theory, methods and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227.

Revelle, W. (2017). *Psych: procedures for personality and psychological research*. USA: Evanston (IL). Retrieved from https://CRAN.R-project.org/package=psych

Citation

Caron, P.-O. & Valois, P. (2018). A computational description of simple mediation analysis. *The Quantitative Methods for Psychology*, *14*(2), 147–158. doi:10.20982/tqmp.14.2.p147

Received: 19/12/2017 ∼ Accepted: 22/02/2018

Table 7 and Listings 3 to 7 follows.

Table 6 ■ Regression models for the first two step of mediation analysis

Regression model	Intercept	AI estimate	PSI estimate	P value
PSI ∼ AI	3.504	0.189	-	<.001
BDI ∼ AI + PSI	50.855	0.824	-1.629	<.001

Note. PSI : Positive social interaction, AI : Positive social interaction, BDI : BDI score

Table 7 ■ Empirical results

Baron & Kenny			
Outcome	Partial mediation		
Sobel test	Values		
z	-5.445		
p	< 0.001		
Bootstrap	Lower	ab	Upper
95% BCa CI	1.043	1.408	1.754

Listing 3 ■ Code for Sobel test.

```
SobelTest <- function(x, y, m, data, alpha = 0.05) {
  # x is the column name or number of the predictor in data
  # m is the column name or number of the mediator in data
  # y is the column name or number of the dependent variable in data
  # data is a data.frame or a matrix that contain columns with the names of x, y and m.
  # If x, m or y are missing, data[,1] will be used for x, data[,2] will be
  # used for m and data[,3] will be used for y.

  if(missing(data)) {
    stop("There's no data")
  }

  if(is.data.frame(data) != TRUE & is.matrix(data)) {
    d <- as.data.frame(data)
  } else if (is.data.frame(data) != TRUE & is.matrix(data) != TRUE) {
    stop('"data" should be a matrix or a data.frame')
  }

  if(missing(x)){x <- data[,1]} else if (is.numeric(x) == TRUE) {x <- data[,x]}
    else {x <- data[,match(x, table = colnames(data))]}
  if(missing(m)){m <- data[,1]} else if (is.numeric(m) == TRUE) {m <- data[,m]}
    else {m <- data[,match(m, table = colnames(data))]}
  if(missing(y)){y <- data[,1]} else if (is.numeric(y) == TRUE) {y <- data[,y]}
    else {y <- data[,match(y, table = colnames(data))]}

  step1 <- lm(formula = m ~ x)
  step2 <- lm(formula = y ~ x + m)

  a <- step1$coefficient[2]
  SEa <- coef(summary(step1))[2, 2]
  b <- step2$coefficient[3]
  SEb <- coef(summary(step2))[3, 2]
  SE <- sqrt(a^2*SEb^2 + b^2*SEa^2)
  z <- a*b/SE
  p <- 1-pnorm(z)
  sig <- qnorm(1-alpha/2) < abs(z)
  return(list(z = z, p = p, sig = sig))
}
```

Listing 4 ■ Code for the bootstrap method. This code requires the function provided in Listing 5

```
BootTest <- function(x, y, m, data, alpha = 0.05, R = 5000) {
  # Warning: This function can be excessively slow with high replication values and high sample sizes
  # x is the column name or number of the predictor in data
  # m is the column name or number of the mediator in data
  # y is the column name or number of the dependent variable in data
  # data is a data.frame or a matrix that contain columns with the names pf x, y and m.
  # If x, m or y are missing, data[,1] will be used for x, data[,2] will be
  # used for m and data[,3] will be used for y.
  # R is the number of replication

  if(missing(data)) {
    stop("There's no data")
  }

  if(is.data.frame(data) != TRUE & is.matrix(data)) {
    d <- as.data.frame(data)
  } else if (is.data.frame(data) != TRUE & is.matrix(data) != TRUE) {
    stop('"data" should be a matrix or a data.frame')
  }

  if(missing(x)){x <- data[,1]} else if (is.numeric(x) == TRUE) {x <- data[,x]}
    else {x <- data[,match(x, table = colnames(data))]}
  if(missing(m)){m <- data[,1]} else if (is.numeric(m) == TRUE) {m <- data[,m]}
    else {m <- data[,match(m, table = colnames(data))]}
  if(missing(y)){y <- data[,1]} else if (is.numeric(y) == TRUE) {y <- data[,y]}
    else {y <- data[,match(y, table = colnames(data))]}

  d <- as.matrix(cbind(x, m, y))

  # Compute the indirect effect for the Bca.boot function
  indirect <- function(data, indice) {
    d <- data[indice,]
    b <- solve(t(d[,1:2])%*%d[,1:2])%*%t(d[,1:2])%*%d[,3]
    a <- solve(t(d[,1])%*%d[,1])%*%t(d[,1])%*%d[,2]
    ab <- a*b[2]
    return(ab)
  }

  res <- BCa.boot(data = d, stat = indirect, R = R)
  sig <- 0 < prod(sign(res$BCaCI))
  return(list(ab = round(res$estimate,3), CI = round(res$BCaCI, 3), sig = sig))
}
```

Listing 5 ■ Code for the bootstrap method (contd.).

```
# Bootstraping function with the bias corrected and accelerated boostrap interval (BCa)
BCa.boot = function(data, stat, R = 5000, alpha=0.05){
  # data is the data to bootstrap    # stat is the function to bootstrap
  # R is the number of replication    # alpha is significance threshold
  data <- as.matrix(data)
  n <- dim(data)[1]
  N <- 1:n
  res <- rep(0,R)
  zj <- rep(0,n)
  est <- stat(data,indice=N)

  M <- max(R,n)
  for (i in 1:M){
    if(i<=R){
     id <- sample(n, replace = TRUE)
     res[i] <- stat(data=data,indice=id)
   }
   if(i<=n){
     J <- N[1:(n-1)]
     zj[i] <- stat(data[-i,],J)
   }
  }
  z0 <- qnorm(sum(res < rep(est,R))/R)
  zc <- qnorm(c(alpha/2,1-alpha/2))
  L <- mean(zj)-zj
  a <- sum(L^3)/(6*sum(L^2)^1.5)
  adj.alpha <- pnorm(z0+zc)/(1-a*(z0+zc))
  limits <- quantile(res,adj.alpha)
  CI <- c(limits[[1]], limits[[2]])
  return(list(estimate = est, BCa=limits, BCaCI = CI))
}
```

Listing 6 ■ General function for power analysis of indirect effect tests.

```
PowerMediation  <- function(MediationTest, a = .25, b = .6, c = .0, n = 40, R =
    5000, alpha = 0.05){
  # Warning : This function can be excessively slow with high replication values and high sample sizes ,
  # especially with bootstrap
  # MediationTest = SobelTest.R or BaronKenny.R or BootTest.R or any function returning an output
  # labelled sig indicating if the result is significant (TRUE or FALSE)

  SIG <- 0
  for(j in 1:R){
    data <- GenerateMediationData(n=n, a=a, b=b, c=c)
    RES <- MediationTest(data=data, alpha=alpha)
    SIG <- RES$sig + SIG
  }
  Power <- round(SIG/R,3)
  return(list(Power=Power))
}
```

Listing 7 ■ Script to run the complete example. `Alcool` refers to Alcoholic drinks per week, `PosSocial`, to Positive social interactions, and `BDI`, to Beck Depression Inventory (t score)

```
#Complete example
set.seed(20180201)
Example <- GenerateMediationData(n = 429, a = .40, b = -0.35, c = .25, mean.x = 10,
     sd.x = 4, mean.m = 5, sd.m = 2, mean.y = 50, sd.y = 10)
colnames(Example) <- c("Alcool", "PosSocial", "BDI")

Example <- ceiling(Example)
Example$Alcool <- ifelse(Example$Alcool < 0, 0, Example$Alcool)
Example$PosSocial <- ifelse(Example$PosSocial < 0, 0, Example$PosSocial)

require(psych)
describe(Example)

# Histograms 1
h.al <-hist(Example$Alcool, breaks=10, col="white", xlab="Alcoholic␣drinks␣/␣week")
xfit.al <-seq(min(Example$Alcool),max(Example$Alcool),length=100)
yfit.al <-dnorm(xfit.al,mean=mean(Example$Alcool),sd=sd(Example$Alcool))
yfit.al <- yfit.al*diff(h.al$mids[1:2])*length(Example$Alcool)
lines(xfit.al, yfit.al, col="black", lwd=2)
# Histograms 2
h.PS <-hist(Example$PosSocial, breaks=10, col="white", xlab= "Social␣int/ion")
xfit.PS <-seq(min(Example$PosSocial),max(Example$PosSocial),length=100)
yfit.PS <-dnorm(xfit.PS,mean=mean(Example$PosSocial),sd=sd(Example$PosSocial))
yfit.PS <- yfit.PS*diff(h.PS$mids[1:2])*length(Example$Posocial)
lines(xfit.PS, yfit.PS, col="black", lwd=2)
# Histograms 3
h.BDI <-hist(Example$BDI, breaks=10, col="white", ylim = c(0,100), xlab= "BDI")
xfit.BDI <-seq(min(Example$BDI),max(Example$BDI),length=100)
yfit.BDI <-dnorm(xfit.BDI,mean=mean(Example$BDI),sd=sd(Example$BDI))
yfit.BDI <- yfit.BDI*diff(h.BDI$mids[1:2])*length(Example$BDI)
lines(xfit.BDI, yfit.BDI, col="black", lwd=2)

# Covariance and correlation matrices
cov(Example)
cor(Example)

#Regression step 1
Results1 = lm(Example[,2]~Example[,1])
summary(Results1)
#Regression step 2
Results2 = lm(Example[,3]~Example[,1]+Example[,2])
summary(Results2)

# Baron and Kenny with default alpha = 0.05
BaronKenny(x = "Alcool", m = "PosSocial", y = "BDI", data = Example)
# Sobel test with default alpha = 0.05
SobelTest(x = "Alcool", m = "PosSocial", y = "BDI", data = Example)
# Bootstrap with default alpha = 0.05
set.seed(20180206)
BootTest(x = "Alcool", m = "PosSocial", y = "BDI", data = Example)
```