




Some computational descriptions of moderation analysis

Pier-Olivier Caron ^a  , Philippe Valois ^{b,c}  & Alexandre Gellen-Kamel ^b 

^aDépartement des Sciences humaines, Lettres et Communications, Télé-Université – Université du Québec

^bLaboratoire des Sciences Appliquées du Comportement, Département de Psychologie, Université du Québec à Montréal


^cCentre d'études troubles obsessionnels-compulsifs et tics, Centre de recherche de l'Institut universitaire en santé mentale de Montréal

Abstract ■ Moderation analysis is getting more and more popular as a statistical analysis in psychology and in other social sciences. However, there are very few detailed accounts of the computations performed within the model. Articles are more often focusing on explaining moderation conceptually rather than mathematically. Thus, the purpose of the current paper is to introduce the computation within moderation analysis accompanied with examples with R. Firstly, three moderation models will be described: a continuous variable with another continuous variable, two groups and then three groups. Then, two ways to analyze moderation (regression and analysis of variance) are presented. We will show examples using R and the code to implement all computations presented is offered for the readers to implement it themselves as well as a script to carry a complete example.

Keywords ■ Moderation analysis, linear model, data modelling. **Tools** ■ R.

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ No reviewer

 pocaroon19@gmail.com

 [10.20982/tqmp.16.1.p009](https://doi.org/10.20982/tqmp.16.1.p009)

Introduction

Moderation analysis is a popular statistical analysis in which a dependent variable (y) is predicted by an independent variable (x), but the relation between both is altered by a third variable, the moderating variable (w)¹. The moderator interacts as to modify the direction or the strength of the relation between x and y (Hayes, 2013). For instance, the effect of social media on self-esteem can be different among men and women; here, the variable sex would be a moderator. Another example is the relation between mathematics and intelligence which may be moderated by reading abilities: being intelligent lead to better results in mathematics, which may be magnified or diminished by reading abilities either by being able to answer more questions or understanding them more easily. (Surely, intelligence would be correlated to reading, but we will not go any deeper into this complex topic.)

There has been an increasing amount of papers dedicated to explaining moderation conceptually. However,

like we discussed previously for another popular statistical analysis (Caron & Valois, 2018), very few details the computation within. Readers could nevertheless be interested in implementing moderation on their own to simulate, to study or to teach the statistical properties of moderation models, notably by applying theoretical concepts (Lemay & Laporte, 2018). Following a pedagogical perspective, this paper adopts a computational standpoint to help the readers to understand and apply the modelling within moderation analysis.

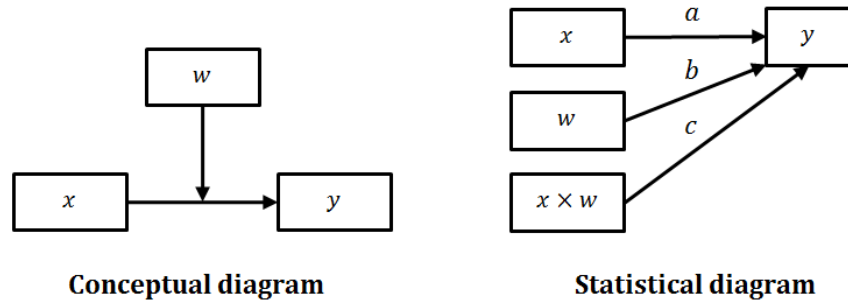
Moderation

Quantitatively, moderation analysis is a one-step general linear model analysis (either an analysis of variance or a regression, depending on the variables at hand) in which a dependent variable, y , is predicted by three variables, the independent variable, x , the moderator variable, w , and their product, $x \times w$. In terms of an analysis of variance, the product $x \times w$ is an interaction. Mathematically, the

¹The letter w was chosen to represent the moderation variable. In a previous article (Caron & Valois, 2018), the letter m have already used for a mediation variable. To distinguish them clearly, the letter w is used in a similar fashion to Hayes (2013).



Figure 1 ■ Conceptual and statistical diagram of moderation analysis.



relation is expressed as:

$$y = ax + bw + c(x \times w) + e_y \tag{1}$$

where a , b and c are the regression coefficients (partial correlation) relating their corresponding term to the dependent variable. It would be more appropriate to label these parameters : $a_{xy|w, x \times w}$, $b_{wy|x, x \times w}$ and $c_{xy|x, w}$, as they are partial correlations controlling for other terms in Equation 1. For the remaining of this paper, we will keep the more parsimonious labels of a , b and c . The terms ax and bw in Equation 1 are main effects; the interaction, $c(x \times w)$ is the effect of interest. Previous works showed that it is important to include the main effects in the analysis (Aiken & West, 1991; Cohen, 1978; Cronbach, 1987). It might be worth to note that, despite using the same coefficient naming convention compared to mediation models (Caron & Valois, 2018), the parameters are not the same. Also note that the intercept was not included in Equation 1 but could be in practice.

A moderation model is depicted in Figure 1. Left panel shows the conceptual diagram of moderation models (how researchers illustrate moderation in a path diagram). The right panel depicts the statistical diagram and explicitly shows the regression coefficients. It illustrates more clearly the moderation effect as the product of the two variables x and w . As more variables are added, the graph tends to be crowded, which may partially explain the preference for the conceptual diagram. Another way to point how moderation operates is by rewriting Equation 1 as to emphasis the key role of w :

$$y = (a + cw)x + bw + e_y \tag{1b}$$

which shows that w influences the effect of x .

Conceptually, x and w can be continuous or nominal variables; the computation is slightly different depending on the scale. Thus, the present paper focus on three different computational descriptions of moderation models.

These cases include a situation with a continuous independent variable and (I) a continuous moderator, (II) a categorical moderator representing two groups and (III) a categorical moderator divided in three groups. It is worth to note that choosing which variable is the independent or the moderator variable is arbitrary and is more a theoretical matter rather than a computational one. These cases will help the readers to generate data to fit the model and shows how to carry a moderation analysis. The R code to generate a moderation model is presented as well as a complete example. For the sake of simplicity, descriptions will be developed with standardized coefficients. Thereafter, the computation to unstandardize data is described.

Computational description I: Two continuous variables

The simplest scenario of moderation is the case with two continuous variables. Like it was illustrated in the right panel of Figure 1, there are three independent variables that predicts the dependent variable, x , w and their product, $x \times w$. Firstly, the variables x and w must be generated according to a normal distribution with a mean of 0 and a variance of 1; $x \sim \mathcal{N}(0, 1)$ and $w \sim \mathcal{N}(0, 1)$. If a correlation between x and w has to be included, then the moderator must be generated using the equation:

$$w = dx + e_w \tag{2}$$

where d is the standardized regression coefficient between x and w , e_w is the error in w (i.e., $\text{var}(e_w)$ is the variance of the residual). Residual error has a mean of 0 and, to keep the variance at 1, the error variance, $\text{var}(e_w)$, is set to :

$$\text{var}(e_w) = 1 - d^2 \tag{3}$$

so that w is normally distributed, $w \sim \mathcal{N}(0, 1)$. Secondly, once x and w are generated, the interaction is computed, as previously mentioned, by the product of both variables.



Listing 1 ■ R code to generate a moderation model with two continuous independent variables.

```

#Generate data for two continuous variables
randModerationCont <- function(n = 1000, a = .1, b = .1, c = .1, d = .1) {
  #n is the number of observation
  #a is the standardized regression coefficient of x over y
  #b is the standardized regression coefficient of w over y
  #c is the standardized regression coefficient of x * w over y
  #d is the standardized regression coefficient of x over w

  #foul proof n and standardized regression coefficient
  if(n <= 0 | n %% 1 > 0 | n == Inf) stop("n_should_be_an_integer_greater_than_0")
  if(-1 > a | a > 1) stop("a_should_be_a_standardized_regression_coefficient")
  if(-1 > b | b > 1) stop("b_should_be_a_standardized_regression_coefficient")
  if(-1 > c | c > 1) stop("c_should_be_a_standardized_regression_coefficient")
  if(-1 > d | d > 1) stop("d_should_be_a_standardized_regression_coefficient")

  x <- rnorm(n)
  w <- d * x + sqrt(1 - d^2) * rnorm(n)
  xw <- x * w
  y <- a * x + b * w + c * xw + sqrt(1 - (a^2 + b^2 + c^2 * (1 + d^2) + 2 * a * b *
    d)) * rnorm(n)

  return(as.data.frame(cbind(x, w, y)))
}

set.seed(42)
example1 <- randModerationCont(n = 10^7, a = .2, b = .4, c = .5, d = .2)

```

Table 1 ■ Variance-covariance matrix of simulated data from the computation description I

Variables	x	w	x × w	y
x	1.000			
w	.200	1.000		
x × w	.000	.001	1.041	
y	.280	.440	.521	1.000

Note. Obtained from the cov() function

The variance of the interaction term is: ²

$$\text{var}(xw) = 1 + d^2 \tag{4}$$

Thus, if there is a correlation between x and w, then the var(xw) ≥ 1. Finally, as for linear model, generating y is carried by the following equation:

$$y = ax + bw + cxw + e_y \tag{5}$$

²The general case is $\text{var}(xw) = E(x^2w^2) - E(xw)^2 = \text{cov}(x^2, w^2) + E(x^2)E(w^2) - (\text{cov}(x, w) + E(x)E(w))^2$ (Casella & Berger, 2002). Acknowledging that x^2 and w^2 are both χ^2 -distributed with 1 degree of freedom, their expectations are 1 for each, thus their product, $E(x^2)E(w^2) = 1$. The covariance between two χ^2 -distributed variables is given by $2\text{cov}(x, w)^2$. Finally, the expectation of x and w are both 0 in the standardized case, and thus cancel out. Only $\text{cov}(x, w)^2$ remains. Since in this paper $d = \text{cov}(x, w)$, this yield $(2d^2) + (1) - (d)^2 = 1 + d^2$.

Table 2 ■ Variance-covariance matrix of simulated data from the computation description II

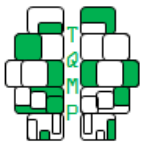
Variables	x	w	x × w	y
x	1.000			
w	.000	.090		
x × w	.900	.000	.900	
y	.650	.036	.630	1.000

Note. Obtained from the cov() function

in which a, b and c are standardized coefficients. The term xw is the interaction term. Coefficients represent the effect of their corresponding variables to the dependent variable. The term e_y is the residual error and its variance is computed by

$$\text{var}(e_y) = \sqrt{1 - (a^2 + b^2 + c^2(1 + d^2) + 2abd)} \tag{6}$$

so that $y \sim \mathcal{N}(0, 1)$. If there is a correlation between x and w, the term $1 + d^2$ weights the variance attributed to the in-



Listing 2 ■ R code to unstandardize data modelled with moderation.

```
unstandModeration <- function (data, x.mean = 0, x.sd = 1, w.mean = 0, w.sd = 1,
  y.mean = 0, y.sd = 1) {

  #Use this function with the result of randModerationCont, randModeration2gr or randModeration3gr function.
  #x.mean, w.mean, y.mean should be replaced by the intended mean of each variable.
  #As example, x.mean = 100 for IQ as the independent variable.
  #x.sd, w.sd, y.sd should be replaced by the intended sd of each variable.
  #As example, x.sd = 15 for IQ as the independent variable.

  if (missing(data)) {stop("Data_is_missing!_Tell_Picard!")}

  data$x <- x.mean + data$x * x.sd
  if (is.factor(data$w) == FALSE) {data$w <- w.mean + data$w * w.sd}
  data$y <- y.mean + data$y * y.sd

  return (data)
}
```

teraction and the term $2abd$ accounts for their covariance (because the sum of two correlated variables is the sum of their variance plus twice their covariance). Listing 1 shows the code to model a moderation in a data set.

The code presented in Listing 2 unstandardized the three variables by multiplying the desired standard deviation, σ_x , and adding the mean, μ_x , to the standardized, x_{std} , to yield the unstandardized variable, x_{unstd} , as followed:

$$x_{unstd} = \sigma_x x_{std} + \mu_x. \tag{7}$$

Listing of Appendix B also shows an example of moderation models with the coefficients, $a = .20$, $b = .40$, $c = .50$, and $d = .20$ with $n = 10^7$. The variance-covariance matrix is showed in Table 1. As expected, all the variances (diagonal cells) are approximately equal to 1, except for the interaction variable, $x \times w$, which, because of the correlation between x and w , i.e., parameter d , has a variance slightly higher than 1 (see Equation 4). Remember that covariances between the independent variables and the dependent variables (last row of the matrix) are different from those implemented, because they originally are partial coefficients and are now expressed as simple coefficients, which explains why they are slightly higher.

Computation description II: Two groups

A second instance of moderation can be due to groups, that is, being part of a group or another alters the strength of the relationship between two variables. A common example could be the difference between men and women, smokers and non-smokers, etc., among a plethora of variables. In this example, the grouping variable is considered

the moderator (i.e., w) whereas it could also be the independent variable (i.e., x).

Like the computational description I, x is generated using a normal distribution, $x \sim \mathcal{N}(0, 1)$. Then, w is generated to have a given number of subjects in either group. There are several ways to generate vectors of 0 and 1 in R. A simple one is to repeat (function `rep()`) the number of 0 and 1 ones m_i times each, m being the number of subjects in group i . It is simpler to have an equal number of subjects in each group though this is not mandatory. Because its domain is limited to 0 and 1, an interesting property of the grouping variable is that its variance is $\text{var}(w) = \frac{m_i(n-m_i)}{n^2}$. This equation is the same as the variance of a binomial distribution, $p(1-p)$, where $p = \frac{m}{n}$. Knowing the variance of w is useful to compute the error variance of y , e_y . It is worth to note that correlating x and w is complex when there is non-normal distribution. Thus, this topic is not covered herein, but the interested readers may look for more advance statistical materials on the matter (Casella & Berger, 2002). That said, the interaction is still computed as the product of x and w and its variance is p . The variable y is computed as follow:

$$y = ax + bw + cxw + e_y. \tag{8}$$

Because the sum of three correlated random variables is

$$\begin{aligned} \text{var}(ax+by + cz) &= a^2\text{var}(x) + b^2\text{var}(y) + \\ & c^2\text{var}(z) + \\ & 2ab \text{cov}(x, y) + 2ac \text{cov}(x, z) + \\ & 2bc \text{cov}(y, z), \end{aligned} \tag{9}$$

and that the covariances between x and w , and also, w and



Table 3 ■ Variance-covariance matrix of simulated data from the computation description III

Variables	x	w_1	w_2	$x \times w_1$	$x \times w_2$	y
x	1.002					
w_1	-.001	.210				
w_2	.001	-.060	.160			
$x \times w_1$.301	.000	.000	.301		
$x \times w_2$.198	.000	.000	.000	.198	
y	.399	.080	-.066	.180	.118	1.009

Note. Obtained from the `cov()` function

$x \times w$ are approximately 0, the error variance can be calculated as

$$\text{var}(e_y) = a^2 + b^2 \text{var}(w) + c^2(1 - p) + 2ac(1 - p), \quad (10)$$

which is then multiplied to a Gaussian random distribution of parameters $\mathcal{N}(0, 1)$, so that $y \sim \mathcal{N}(0, 1)$.

Listing of Appendix B shows all the R code to implement the current computation description with two groups and a continuous variable. The example uses the same parameters as the previous computational description, that is, $a = .20$, $b = .40$, $c = .50$, and with $n = 10^7$. There is no correlation between x and w (no d parameter). The first group has 1 000 000 subjects and the other has 9 000 000 subjects. The variance-covariance matrix is showed in Table 2. As expected, the variances of x and y are approximately equal to 1. The other two variances are $\frac{m}{n}$ and $\frac{n-m}{n}$. Like the previous example, the relation between the independent variables and the dependent variables are different from those implemented.

Computational description III: Three groups

A third and slightly more complicated moderation model is the case with a continuous variable and three groups because it requires more computation. If the first two computational descriptions necessitated three independent variables, this computation description needs five, that are, one continuous variable, two grouping variables and two interactions. It also implies some inevitable covariances that have to be taken into account.

Here, we present a slight digression to explain dummy coding briefly. When the grouping variable is composed of three or more groupings, groups cannot be distinguished only with 0 and 1 on a single variable, $k - 1$ variables are necessary (where k is the number of group (Cohen, Cohen, West, & Aiken, 2003)). The participants of the first group are represent by 1 on the first dummy variable. The second group is referred as 1 on the second variable. All other values are 0. The last group does not need to be referred by a variable, because participants not in group 1 or 2 are logically in group 3. The third group is used to compare group 1 and 2 in the linear model.

The rational within the modelling is the same as previously. We need five parameters, a, b_1, b_2, c_1 and c_2 , where indices refer to the effect of the corresponding group. There are also two implicit parameters, b_3 and c_3 which are set to 0 (herein, the third group is the reference group in dummy coding). Like previous models, the first step is to generate a random variable, x , normally distributed $x \sim \mathcal{N}(0, 1)$. Then, two vectors, w_1 and w_2 of length n , indicate which subject is related to which group. Subjects in neither w_1 or w_2 are logically in the third group. From these two vectors, we can compute two moderator variables, $x \times w_1$ and $x \times w_2$. The variance of the interaction is the probability to be in that group: $p_i = \frac{m_i}{n}$. We can now compute y by the following equation:

$$y = ax + b_1w_1 + b_2w_2 + c_1xw_1 + c_2xw_2 + e_y \quad (11)$$

where e_y is calculated by

$$e_y = a^2 + b_1^2 \text{var}(w_1) + b_2^2 \text{var}(w_2) + c_1^2 p_1 + c_2^2 p_2 - 2p_1 p_2 + 2ap_1 + 2ap_2 \quad (12)$$

and where the three last terms refer to non-null covariances between variables (which can be seen in Table 3).

Listing of Appendix C shows an example to generate data according to computational description III. The parameters are $a = .20$, $b_1 = .30$, $b_2 = -.30$, $c_1 = .40$ and $c_2 = .40$. The size of each group is $m_1 = 3 \times 10^5$, $m_2 = 2 \times 10^5$ and $m_3 = 5 \times 10^5$ to reach $n = 10^6$. In this case, the moderating effect is not different between groups 1 and 2, but both groups have the same different moderating effect compared to the third group, where there is no influence of the moderator. Table 3 shows the variance-covariance matrix from the simulated data given the previously mentioned parameters. There is a covariance between w_1 and w_2 which is equal to $p_1 p_2$. Unfortunately, the variance of y will always be slightly higher than 1, because of small covariances between x and w_1 and w_2 . Like the previous example, the relation between the independent variables and the dependent variables are different from those implemented.

**Table 4** ■ Descriptive analysis of the complete example

Variables	Mean	SD	Min	Max	Skew	Kurtosis	SE
Study time	1.99	.33	1.01	3.07	.05	.01	.01
Locus of control	19.86	4.09	7.00	34.00	-.05	.25	.17
Grade	75.06	5.02	55.66	92.04	-.01	.84	.21

Note. Obtained with the function `describe()` from `psych` package, SD : Standard deviation, Min : Minimum observed value, Max : Maximum observed value, Skew : Skewness, SE : Standard error of the mean.

Listing 3 ■ Moderation analysis code in R.

```
# Herein, x is the independent variable, y is the dependent variable and w is the moderator

# Computational description I: Two continuous variables
res1 <- lm(y~x*w, data = example1)

# Computational description II: Two groups
res2 <- aov(y~x*w, data = example2)

# Computational description III: Three groups
res3 <- aov(y~x*w, data = example3)
```

The analysis

Until now, we have described three moderation models. We have not yet described how to detect such an effect in a data set. Like it was previously mentioned, moderation analysis, in its simplest form, is based on the general linear model with two independent variables and an interaction effect to predict a dependent variable. As such, it is equivalent to a regression or an analysis of variance. In R, the linear regression is carried with the `lm()` function and the analysis of variance is carried with the `aov()` function. To declare an interaction, the command “*” (product sign) specifies the interaction between the two variables. Both yield the same regression coefficients, however, the `aov()` function has a more appropriate Type I error rate if there is a grouping variable, because of the omnibus test for the said variable. Listing 3 shows the code to implement the moderation analysis. Once the analysis is carried out, the moderation effect is deemed significant if the interaction term, or the coefficient c , is significant (p -value below a specified threshold, usually 0.05).

A complete example

A complete and hypothetical example with two continuous variables as described previously is provided in this section. The appendix shows the code to implement the complete example (make sure you have already run the two functions described earlier: `randModerationCont()` in Listing 1 and `unstandardize()` in Listing 2). Five

hundred and eighty-one pupils were asked to answer (a) the Nockiwi-Strickland locus of control scale (Nowick & Strickland, 1973), (b) a self-reported questionnaire on time passed to study per week, and (c) a mathematic test at the end of the semester.

The Nowick-Strickland test (1973) assesses the perception of control on events that affect one’s life on a continuous scale from 0 (internal) to 40 (external). A score of 0 means an individual perceives having control on events in his own life and a score of 40 means that an individual perceives the events in his life are controlled by external phenomena (i.e., luck, chance). To help for the interpretation, scores are reversed, so that 0 shows a high external locus of control and that 40 means a high internal locus of control. Then, the self-reported questionnaire assesses the number of hours by week passed to study. The last test consists of a hundred questions on mathematic matter and the grade obtained may be range between 0 and 100. The hypothesis suggests that the effect of time by week studying (the independent variable, x , labelled StudyTime) on the grade’s math test (the dependent variable, y , labelled Grade) is moderated by an individual’s locus of control (the moderator, w , labelled Locus).

Table 4 shows the descriptive analysis using the `describe()` function from the `psych` package (Revelle, 2018) and Figure 2 presents histograms with density curve showing a normal distribution of data for the three variables using the `hist()` function. Table 5 show the variance-covariance matrix with the function `cov()` and

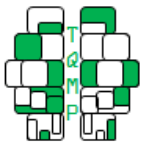


Figure 2 ■ Distribution of complete example’s variables using the hist () function.

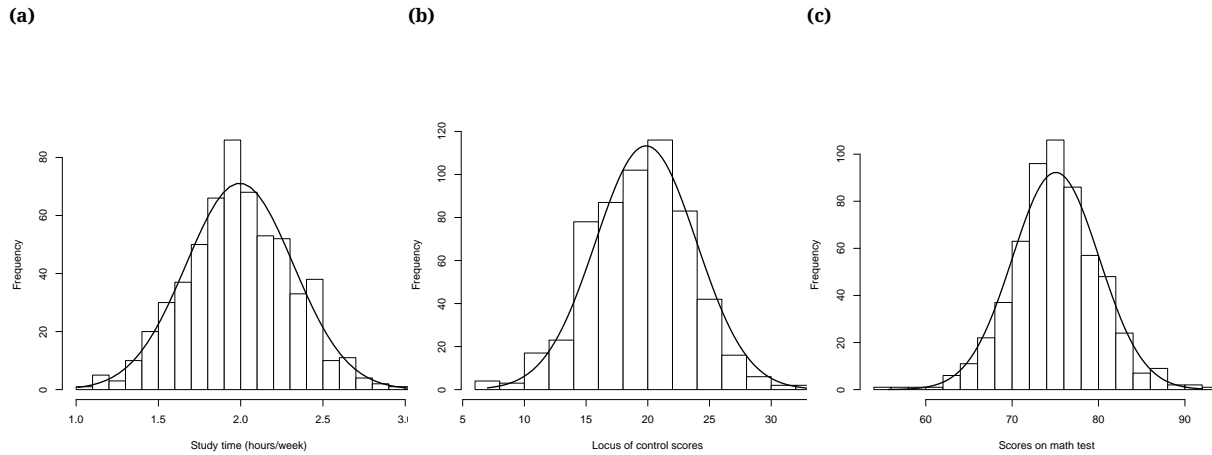


Table 5 ■ Variance-covariance and correlation matrices of the complete example

Variables	Variance-covariance			Correlation		
	Study	Locus	Grade	Study	Locus	Grade
Study time	.106			1.000		
Locus of control	.024	16.739		.018	1.000	
Grade	.795	.251	25.249	.485	.012	1.000

Note. Obtained with the function cov () and the function cor () respectively.

correlation matrix with the cor () function.

Table 6 details the moderation analysis processed with a one-step regression model of StudyTime interacting with Locus on Grade (using the lm () function and “ * ” for the interaction). The regression model is significant, $F(3, 577) = 207.60, p < .001$, and the interaction effect is also significant, $\beta = 2.03, p < .001$. This result suggests a moderation effect of locus of control, that the effect of study time on grade is influenced by the effect locus of control. As an interaction is easier to interpret with a graph, Figure 3 helps visualizing the effect of the moderator Locus on the relationship between independent variable StudyTime and dependent variable Grade using the interact_plot () function from the interaction package (Long, 2019). The three moderator’s levels of Locus

are defined with a standard deviation of 1, the mean and a standard deviation of -1. Remember, these results come from generated data provided in Listing 1 for pedagogical purposes.

Discussion

The purpose of the current paper was to introduce the computation within moderation models. We detailed the conceptual and statistical diagram of moderation. We also explained three moderation models with a continuous variable with another continuous variable, two groups and then three groups. We presented some examples using R and offered the code for readers to implement it themselves. We hope this work will encourage statistical research, the promotion of good methodological practices

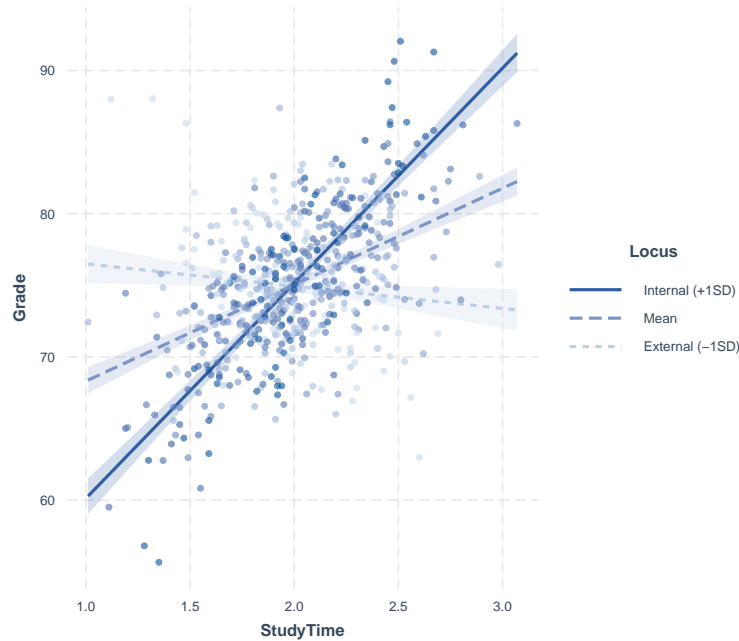
Table 6 ■ Regression models of the moderation analysis of the complete example.

Estimate	SE	t value	p-value
Intercept	141.54	4.56	31.05 <.001
$G \sim ST$	-33.49	2.26	-14.79 <.001
$G \sim L$	-4.03	.22	-18.19 <.001
$G \sim ST \times L$	2.03	.11	18.45 <.001

Note. SE : Standard error of the mean, G : Grade, ST : StudyTime, L : Locus



Figure 3 ■ Visualisation of moderation analysis of the complete example using `interact_plot()` function from `interaction` package.



and help teaching of more advanced statistical technics to students.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression : Testing and interpreting interactions*. Newbury Park, CA: Sage.

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

Casella, G., & Berger, R. L. (2002). *Statistical inference*. Pacific Grove: Duxbury.

Cohen, J. (1978). Partialled products are interactions; partialled power in moderated regression analyses. *Psychological Bulletin, 85*, 858–866. doi:10.1037/0033-2909.85.4.858

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cronbach, L. J. (1987). Statistical tests for moderator variables : Flaws in analyses recently proposed. *Psychological Bulletin, 102*, 414–417. doi:10.1037/0033-2909.102.3.414

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

Lemay, C., & Laporte, W. (2018). Compte rendu du colloque “L’enseignement des statistiques en sciences humaines et sociales”. *The Quantitative Methods for Psychology, 14*, 159–166. doi:10.20982/tqmp.14.3.p159

Long, J. A. (2019). Interactions: Comprehensive user-friendly toolkit for probing interactions (Version 1.0). Retrieved from <https://cran.r-project.org/package=interactions>

Nowick, S., & Strickland, B. (1973). A locus of control scale for children. *Journal of Consulting and Clinical Psychology, 40*, 148–154. doi:10.1037/h0033978

Revelle, W. (2018). *Psych: Procedures for personality and psychological research*. USA: Evanston (IL). Retrieved from <https://CRAN.R-project.org/package=psych>

Appendix A: R code for the complete example 1.

```
# Information
# StudyTime refers to hours of study by week.
```




```
# Locus refers to the Nowicki–Strickland test. Remember, scores are reversed for interpretation purpose.
# A low score means a high external locus of control and a high score means a high internal locus.
# Grade refers to the score obtained on the mathematic test.
# Make sure you have already run two functions described earlier : randModerationCont() in Listing 1 and
# unstandardize() in Listing 4.

# Generate data
set.seed(42)
example1 <- randModerationCont(n = 581, a = .5, b = 0, c = .5, d = 0)

# Unstandardize data. x is StudyTime, w is Locus and y is Grade.
example1 <- unstandModeration(example1, x.mean = 2, x.sd = .33, w.mean = 20, w.sd =
  4, y.mean = 75, y.sd = 5)

# For proper data
colnames(example1) <- c("StudyTime", "Locus", "Grade")
example1$StudyTime <- round(example1$StudyTime, digits = 2)
example1$Locus <- round(example1$Locus, digits = 0)
example1$Grade <- round(example1$Grade, digits = 2)

# Histogram : StudyTime
h.st <-hist(example1$StudyTime, breaks=20, col="white", xlab="Study_time_(hours/
  week)")
xfit.st <-seq(min(example1$StudyTime),max(example1$StudyTime),length=50)
yfit.st <-dnorm(xfit.st,mean=mean(example1$StudyTime),sd=sd(example1$StudyTime))
yfit.st <- yfit.st*diff(h.st$mids[1:2])*length(example1$StudyTime)
lines(xfit.st, yfit.st, col="black", lwd=2)

# Histogram : Locus
h.mo <-hist(example1$Locus, breaks=10, col="white", xlab= "Locus_of_control_scores"
  )
xfit.mo <-seq(min(example1$Locus),max(example1$Locus),length=100)
yfit.mo <-dnorm(xfit.mo,mean=mean(example1$Locus),sd=sd(example1$Locus))
yfit.mo <- yfit.mo*diff(h.mo$mids[1:2])*length(example1$Locus)
lines(xfit.mo, yfit.mo, col="black", lwd=2)

# Histogram : Grade
h.Gr <-hist(example1$Grade, breaks=20, col="white", xlab= "Scores_on_math_test")
xfit.Gr <-seq(min(example1$Grade),max(example1$Grade),length=100)
yfit.Gr <-dnorm(xfit.Gr,mean=mean(example1$Grade),sd=sd(example1$Grade))
yfit.Gr <- yfit.Gr*diff(h.Gr$mids[1:2])*length(example1$Grade)
lines(xfit.Gr, yfit.Gr, col="black", lwd=2)

# Descriptive analysis
require(psych)
describe(example1)

# Covariance and correlation matrices
cov(example1)
cor(example1)

# Results
```



```
res1 <- lm(Grade ~ StudyTime * Locus, data = example1)
summary(res1)
```

Interpretation

*# For one point added on the score of Locus, the effect of StudyTime on Grade is:
(see result in the intersection between "Estimate" and "StudyTime:Locus").*

Plot moderation

```
require(interactions)
interactions::interact_plot(model = res1,
                             pred = StudyTime,
                             modx = Locus,
                             plot.points = TRUE,
                             colors = "Blues",
                             point.size = 1.5,
                             interval = TRUE,
                             modx.labels = c("External_(-1SD)", "Mean", "Internal_
(+1SD)"))
```

Appendix B: R code to generate a moderation model with a continuous variable and two groups.

```
randModeration2gr <- function(n = 1000, a = .1, b = .1, c = .1, m = c(.5, .5)) {
  #n is the number of observation
  #a is the standardized regression coefficient of x over y
  #b is the standardized regression coefficient of w over y
  #c is the standardized regression coefficient of x * w over y
  #m is a vector of two proportions with a sum of 1 for n by group

  #foul proof n, m and standardized regression coefficient
  if(n <= 0 | n %% 1 > 0 | n == Inf) stop("n_should_be_a_integer_greater_than_0")
  if(-1 > a | a > 1) stop("a_should_be_a_standardized_regression_coefficient")
  if(-1 > b | b > 1) stop("b_should_be_a_standardized_regression_coefficient")
  if(-1 > c | c > 1) stop("c_should_be_a_standardized_regression_coefficient")
  if(sum(m) != 1 | length(m) != 2) stop("m_should_be_a_vector_of_two_proportions_
with_a_sum_of_1")

  m1 <- n * m[1]
  m2 <- n * m[2]
  x <- rnorm(n)

  #generate the grouping variable
  w <- c(rep(0, m1), rep(1, m2))
  xw <- x * w
  var.w <- m1 * m2/n^2 #variance of w
  p <- m1/n

  ey <- a^2 + b^2 * var.w + c^2 * (1 - p) + 2 * a * c * (1 - p)
  y <- a * x + b * w + c * xw + sqrt(1 - ey) * rnorm(n)
  res <- as.data.frame(cbind(x, w, y))
  res$w <- factor(res$w, levels = c(0, 1))

  return(res)
```



```
}

set.seed(42)
example2 <- randModeration2gr(n = 10^7, a = .2, b = .4, c = .5, m = c(.1, .9))
```

Appendix C: R code to generate a moderation model with a continuous variable and three groups; The third group is the referent group.

```
randModeration3gr <- function(n = 1000, a = .1, b1 = .1, b2 = .1, c1 = .1, c2 = .1,
  m = c(.33, .33, .34)) {
  #n is the number of observation
  #a is the standardized regression coefficient of x over y
  #b1 is the standardized regression coefficient of w1 over y
  #b2 is the standardized regression coefficient of w2 over y
  #c1 is the standardized regression coefficient of x * w1 over y
  #c2 is the standardized regression coefficient of x * w2 over y
  #m is a vector of three proportions with a sum of 1 for n by group

  #foul proof n, m and standardized regression coefficient
  if(n <= 0 | n %% 1 > 0 | n == Inf) stop("n_should_be_a_integer_greater_than_0")
  if(-1 > a | a > 1) stop("a_should_be_a_standardized_regression_coefficient")
  if(-1 > b1 | b2 > 1) stop("b1_should_be_a_standardized_regression_coefficient")
  if(-1 > b1 | b2 > 1) stop("b2_should_be_a_standardized_regression_coefficient")
  if(-1 > c1 | c1 > 1) stop("c1_should_be_a_standardized_regression_coefficient")
  if(-1 > c2 | c2 > 1) stop("c2_should_be_a_standardized_regression_coefficient")
  if(sum(m) != 1 | length(m) != 3) stop("m_should_be_a_vector_of_three_proportions_
    with_a_sum_of_1")

  m1 <- n * m[1]
  m2 <- n * m[2]
  m3 <- n * m[3]
  p1 <- m1/n
  p2 <- m2/n
  var.w1 <- p1 * (1 - p1)
  var.w2 <- p2 * (1 - p2)
  x <- rnorm(n)

  #generate the grouping variable
  w1 <- c(rep(1, m1), rep(0, m2 + m3))
  w2 <- c(rep(0, m1), rep(1, m2), rep(0, m3))
  xw1 <- x * w1
  xw2 <- x * w2

  ey <- a^2 + b1^2 * var.w1 + b2^2 * var.w2 + c1^2 * p1 + c2^2 * p2 - 2 * p1 * p2 +
    2 * a * p1 + 2 * a * p2
  y <- a * x + b1 * w1 + b2 * w2 + c1 * xw1 + c2 * xw2 + sqrt(1 - ey) * rnorm(n)
  w <- w1 + w2 * 2

  res <- as.data.frame(cbind(x, w, y))
  res$w <- factor(res$w, levels = c(0, 1, 2))

  return(res)
```



}

```
set.seed(42)
example3 <- randModeration3gr(n = 1000000, a = .2, b1 = .3, b2 = -.3, c1 = .4, c2 =
  .4, m = c(.3, .2, .5))
```

Open practices

📄 The *Open Material* badge was earned because supplementary material(s) are available on the [journal's web site](#).

Citation

Caron, P.-O., Valois, P., & Gellen-Kamel, A. (2020). Some computational descriptions of moderation analysis. *The Quantitative Methods for Psychology*, 16(1), 9–20. doi:[10.20982/tqmp.16.1.p009](https://doi.org/10.20982/tqmp.16.1.p009)

Copyright © 2020, Caron, Valois, and Gellen-Kamel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 12/10/2018 ~ Accepted: 10/12/2019