**INTRODUCTION TO MODERATED REGRESSION**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

This document assumes you are, at least somewhat, familiar with linear regression, sometimes called multiple regression. Linear regression, as you will recall, is utilized to ascertain whether various predictors, such as self-esteem and IQ, are related to some outcome, such as the motivation of research candidates.

Sometimes, you might want to explore whether other characteristics or conditions might affect these relationships or associations. For example, if their supervisors are helpful, research candidates might tend to be motivated regardless: thus, in these circumstances, self-esteem might not be as related to the motivation of research candidates. In contrast, if their supervisors are unhelpful, self-esteem might be strongly related to the motivation of research candidates. The following graph represents this possibility.



As this graph shows

* when supervisors are not supportive—corresponding to the broken line—self-esteem is strongly related to motivation of candidates. A small increase in self-esteem greatly enhances motivation
* when supervisors are supportive— corresponding to the unbroken line—self-esteem is not quite as strongly related to motivation of candidates. A small increase in self-esteem modestly enhances motivation
* thus, supervisor support affects or changes the association between self-esteem and motivation
* researchers usually write that “support from supervisors **moderates** the association between self-esteem and motivation”

This document shows you how to assess whether one variable, such as support from supervisors, moderates the association between other variables. This approach is sometimes called moderated regression analysis. In essence, you merely need to

* transform all the variables so their mean is zero. For example, you can convert all the variables to z scores.
* construct a new column or variable in your data file that is merely the product of your moderator and predictor—in this instance, support from supervisors and self-esteem
* conduct a typical linear regression analysis, except include this new column as well
* determine whether or not this new column is significant and, if so, construct a figure that resembles the previous graph.

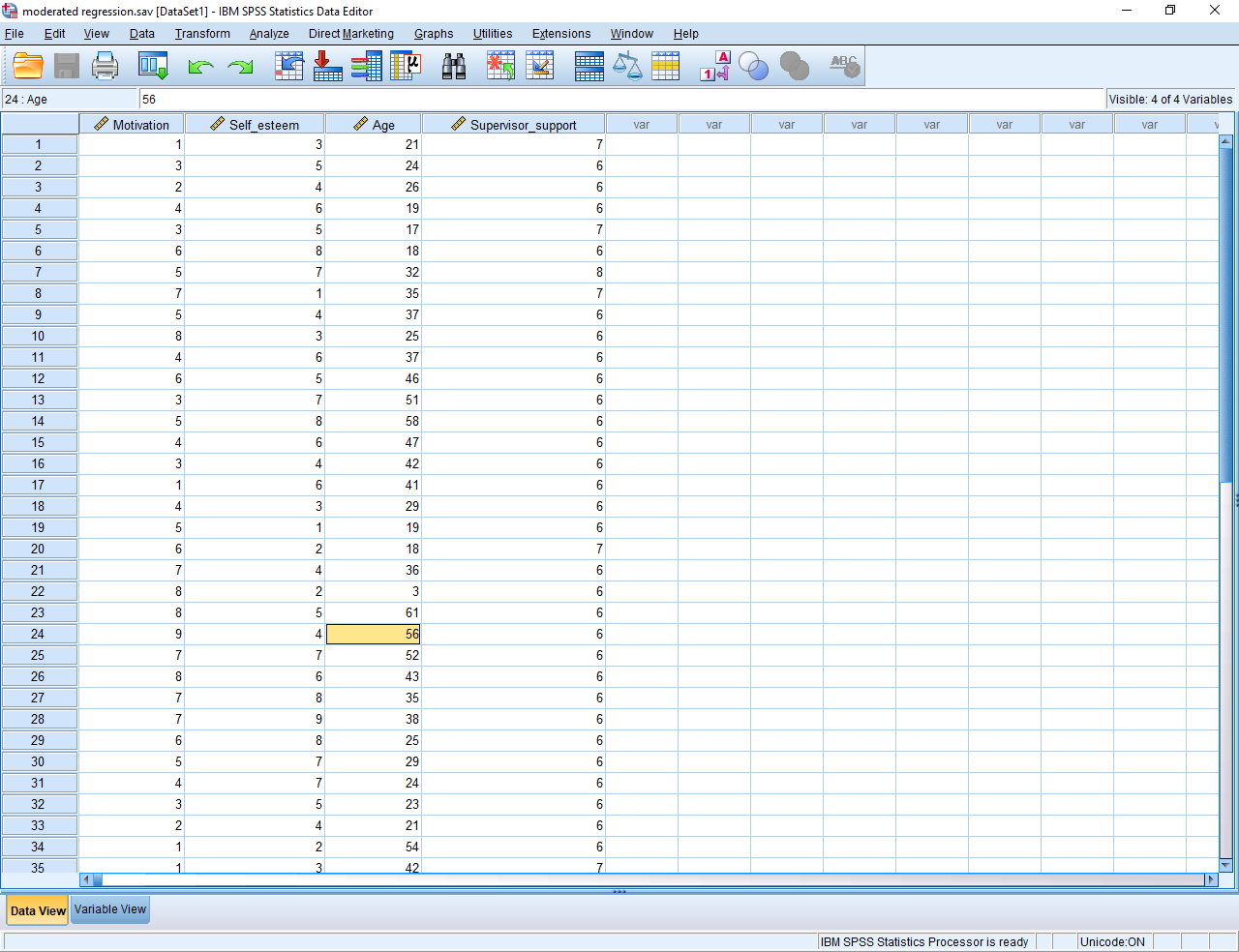
|  |
| --- |
| **A simple example** |

**Example**

To introduce moderated regression, consider this example. Suppose you want to predict which research candidates are likely to feel motivated. To investigate this topic, a researcher administers a survey to 500 research candidates. This survey includes questions that assess

* motivation, such as “On a scale of 1 to 10, how motivated do you feel”
* self-esteem, such as “On a scale of 1 to 10, to what extent do you feel proud of who you are”
* age
* and supervisor support, such as “On a scale of 1 to 10, how supportive is your supervisor”

An extract of the data appears in the following screen. Like most data files, each row corresponds to one person. Each column corresponds to a separate characteristic, called a variable.

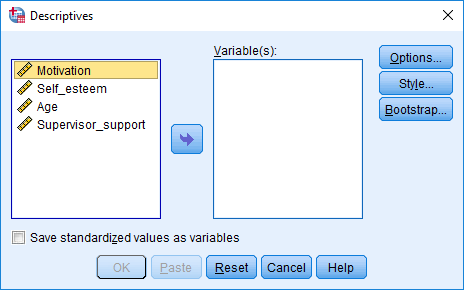


To conduct moderated regression, you merely need to include a few adjustments to linear regression. Therefore, any statistical package that conducts linear regression can be used to conduct moderated regression as well. This example utilises SPSS. If you use another package, such as R or Stata, perhaps follow these examples anyway.

**Transform the variables so the means are zero**

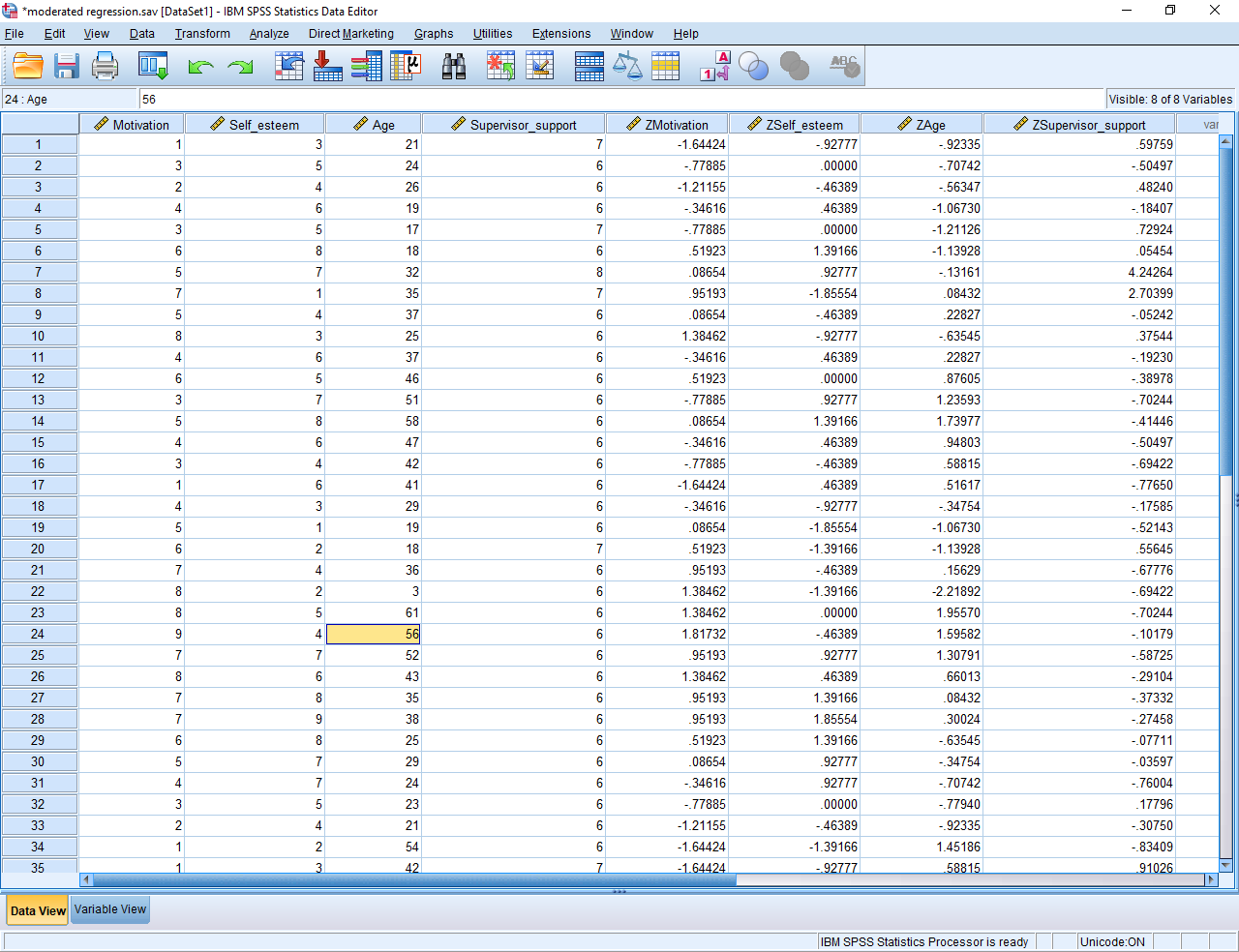
First, you need to transform the variables to set the means to zero; otherwise, the results are hard to interpret. You can apply several approaches to achieve this goal. Perhaps the simplest approach, however, is to convert the variables to standardized scores, sometimes called z scores. Specifically

* to convert a score to a z score, subtract the mean of this variable and then divide by the standard deviation
* or, in SPSS, choose “Descriptive Statistics” and then “Descriptives” from the “Analyze” menu, to generate the following screen
* then, highlight all the variables and press the arrow—to insert the variables into the box called Variables
* tick the box alongside “Save standardized values as variables” and press OK



In the data file will be a new series of columns, representing the z scores of each variable. The following spreadsheet illustrates these z scores. Although not vital to your understanding

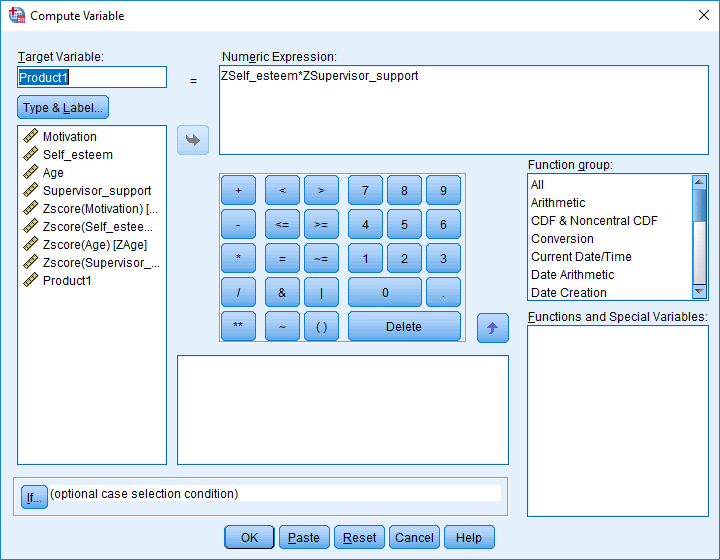
* z scores about 0 correspond to values that exceed the mean
* z scores below 0 correspond to values that are below the mean
* z scores close to 0 correspond to values that resemble mean



**Construct a product term**

The next step is to construct a new column that represents the product of your predictor, in this instance self-esteem, and your moderator, in this instance supervisor support. In SPSS, to construct this new column

* choose “Compute Variable” from the “Transform” menu
* in the box called “Target variable”, simply enter a name to label this new column, such as “Product1”
* in the box called “Numerical expression”, multiply the relevant z score variants of your independent variable and moderator, such as enter “zSelf\_esteem \* zSupervisior\_support”
* press OK
* the following screen illustrates these activities



This procedure will generate a new column in your data file called Product1—or whatever you labelled this column.

**Conduct the linear regression**

After you create this product term, you are ready to conduct the moderated regression analysis. That is, conduct a linear regression as you would usually but utilize the standardized variables and include this product term as well. Specifically, in SPSS for example

* choose “Regression” and then “Linear” from the “Analyze” menu
* the dependent variable in this example is “zMotivation”
* the independent variables in this example are “zself\_esteem”, “zsupervisor\_support”, zAge, and Product 1
* you can, in principle, include more than one product terms at a time
* importantly, whenever you include a product term, also include the two variables from which this product term was derived, such as self-esteem and supervisor support.
* press OK to generate output that resembles the following table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | .083 | .158 |  | .527 | .602 |
| Zscore(Self\_esteem) | .410 | .160 | .410 | 2.554 | .016 |
| Zscore(Age) | .070 | .157 | .070 | .447 | .658 |
| Zscore(Supervisor\_support) | -.016 | .160 | -.016 | -.100 | .921 |
| Product1 | -.341 | .162 | -.329 | -2.103 | .044 |
| a. Dependent Variable: Zscore(Motivation) | | | | | | |

**Interpret the output**

To interpret the table of coefficients, merely determine whether the significance or p value associated with the product term is significant. In particular

* if this p value is greater than .05 and thus not significant, you would conclude the moderator does not significantly affect the association between the predictor and outcome
* if this p value is less than .05 and thus significant, you would conclude the moderator significantly affects the association between the predictor and outcome
* in this example, the product term is significant. Thus, supervisor support does moderate or affect the association between self-esteem and motivation.

**If the product term is significant, construct two equations**

The previous example shows that supervisor support does moderate or affect the association between self-esteem and motivation. But, the p value is not especially informative. This p value does not clarify whether the positive association between self-esteem and motivation strengthens or subsides as supervisor support increases. To resolve this uncertainty, the simplest method is to derive a graph from the B values. To construct this graph, you can

* utilize the Excel file “Graph creation”, available on this webpage
* apply the following instructions—a procedure that demands more time but clarifies your understanding

You may recall, from your reading about multiple regression, that B values can be used to generate an equation. In this example, the equation is

|  |
| --- |
| Motivation = .083 + .41 x Self-esteem + .07 x Age - .16 x Supervisor support - .34 x Product1 |

Note these labels actually refer to the standardized variables and not the raw values. Furthermore, Product 1 is merely Self-esteem x Supervisor support. Therefore, the updated equation is

|  |
| --- |
| Motivation = .083 + .41 x Self-esteem + .07 x Age - .16 x Supervisor support - .34 x Self-esteem x  Supervisor support |

We are now interested in the association between self-esteem and motivation at two or more levels of supervisor support. In particular, we would like to explore these relationships when all other variables, such as Age, are average. Because these variables are standardized, the average of every variable is 0. Therefore, we substitute the other variables with 0. The equation thus becomes

|  |
| --- |
| Motivation = .083 + .41 x Self-esteem + .07 x 0 - .16 x Supervisor support - .34 x Self-esteem x  Supervisor support  = .083 + .41 x Self-esteem - .16 x Supervisor support - .34 x Self-esteem x  Supervisor support |

To construct this graph, first estimate the level of motivation when self-esteem and supervisor support equals 1: you will understand why soon. Remember that self-esteem and supervisor support are standardized variables, and thus 1 represents above average values. In this instance, when self-esteem and supervisor support equals 1

|  |
| --- |
| **When self-esteem = 1 and supervisor support = 1**  Motivation = .083 + .41 x 1 - .16 x 1 - .34 x 1 x 1 = -.007 |

Now estimate level of motivation

* when self-esteem equals 1 but supervisor support equals - 1
* when self-esteem equals -1 but supervisor support equals 1
* both self-esteem and supervisor support = -1

|  |
| --- |
| **When self-esteem = 1 and supervisor support = - 1**  Motivation = .083 + .41 x 1 - .16 x - 1 - .34 x 1 x - 1 = .993 |

|  |
| --- |
| **When self-esteem = -1 and supervisor support = 1**  Motivation = .083 + .41 x 1 - .16 x - 1 - .34 x 1 x - 1 = -.147 |

|  |
| --- |
| **When self-esteem = -1 and supervisor support = - 1**  Motivation = .083 + .41 x 1 - .16 x - 1 - .34 x 1 x - 1 = -.507 |

To reiterate, you have now estimated motivation at two levels of self-esteem and two levels of supervisor support. This information is sufficient to construct a graph—a graph that will resemble this figure.



To generate this graph

* open Microsoft Excel, as the screen below demonstrates
* enter words and numbers that resemble this example
* however, instead of “self-esteem” and “supervisor support”, specify the name of your moderator and independent variable.
* in addition, instead of the numbers in these cells, specify the numbers that you calculated.
* for example, in cell B2, specify the number you calculated when the independent variable is -1 and the moderator is -1.
* to create the graph, then highlight all the cells from A1 to C3.
* choose “Insert”, “Chart”, and “Line”. You then might need to choose a few more options to create a suitable graph.



|  |
| --- |
| **Dichotomous predictors and moderators** |

Usually, when researchers conduct moderated regression, the predictors and moderators are numerical. Yet, if either or both of these variables are categorical—and comprise only two categories, such as gender—you can still conduct moderated regression analysis. You could proceed as usual, besides three minor amendments. The following table outlines these amendments

|  |  |
| --- | --- |
| Amendment | Detail |
| Code dichotomous variables as 1 and 0 | * Do not code gender, for example, as 1 and 2. Instead, 1 could represent males and 0 could represent females. |
| When you construct the graph, do not insert 1 and -1 into the equations. | * Instead, for these dichotomized variables, insert the two numbers that will appear in the standardized version of this variable. * For example, the original version of gender might be coded as 1 and 0 to represent males and females respectively * The standardized variable might include the numbers 0.7667 and -.07667, representing males and females respectively * Enter these numbers into the formulas * The reason is the two lines you generate will correspond to these two categories-in this instance, males and female respectively |
| Use these category labels when constructing the graph | * For this dichotomous variable, replace “z = 1” and “z = -1” with the corresponding categories, such as “males” and “females”. |

|  |
| --- |
| **The rationale that underpins moderated regression** |

The rationale that underpins moderated regression is not easy to explain. But, you might gain some insight from this example.

Observe the following table closely. When supervisors are supportive—the data above the black line—self-esteem is strongly associated with motivation. That is, each increase in self-esteem corresponds to an increase in motivation. When supervisors are not supportive —the data below the black line—self-esteem is not strongly associated with motivation.

|  |  |  |  |
| --- | --- | --- | --- |
| Motivation | Self esteem | Support from supervisor | Product of self-esteem and support from supervisor |
| 5 | 1 | 5 | 5 |
| 6 | 2 | 5 | 10 |
| 7 | 3 | 5 | 15 |
| 8 | 4 | 5 | 20 |
| 9 | 5 | 5 | 25 |
| 7 | 1 | 1 | 1 |
| 7 | 2 | 1 | 2 |
| 7 | 3 | 1 | 3 |
| 7 | 4 | 1 | 4 |
| 7 | 5 | 1 | 5 |

Therefore, in this example, support from supervisors changes the relationship between self-esteem and motivation. The moderation should be significant.

The last column presents the product of self-esteem and support from supervisor. For both motivation and this product, notice how the differences between the numbers above the line are greater than the differences between the numbers below the line. That is, when the moderation is significant, this product term demonstrates a similar pattern, and therefore should correlate, with the dependent variable. This illustration may provide some preliminary insight into why a significant product term may reflect moderation, even after controlling the independent variable and moderator.