Standardized Simple Mediation Model: A Numerical Example

1,2,3 Anwar Fitrianto and 1,2 Habshah Midi

1 Department of Mathematics, Faculty of Science Universiti Putra Malaysia, Malaysia
2 Laboratory of Applied and Computational Statistics, Institute for Mathematical Research, Universiti Putra Malaysia
3 Department of Statistics, Faculty of Mathematics and Natural Sciences Bogor Agricultural University, Indonesia

Abstract: Mediation models figure out how an effect occurred by hypothesizing a causal sequence. For a simple mediation model, a causal sequence is described in which an independent variable causes the mediator which sequentially causes the dependent variable. In this article, we tried to introduce to use standardized regression coefficient to the involving the simple mediation model since a standardized coefficient will be more meaningful than an unstandardized coefficients. In this article, we show that in simple mediation model, even though standardized regression coefficients are different from the unstandardized coefficients, but the standardized coefficients maintain the order of magnitude of the unstandardized regression coefficients for the simple mediation model.

Key words: Mediation Model · Standardized Coefficient · Standard Error

INTRODUCTION

Mediation modeling is a powerful analytical tool that can be used to explain the nature of the relationship among three or more variables. In addition to that, it can be used to show how a variable mediates the relationship between levels of intervention and outcome. Mediation models also explain how an effect occurred by hypothesizing a causal sequence. The basic mediation model is a causal sequence in which the independent variable \((X)\) causes the mediator \((M)\) which in turn causes the dependent variable \((Y)\), therefore explaining how \(X\) had its effect on \(Y\) ([1] and [2]). In general, a mediational effect is the accounting of the relationship between a predictor variable and an outcome. Research has presented basic measurement approaches for assessing the effect of mediation ([3], [4] and [5]).

There are many definitions of a mediator variable, but according to [6], the definition is a variable that is “causally between two variables and that accounts for the relationship between those two variables”. In contrast to an independent variable \(X\)’s direct effect on dependent variable \(Y\), which is its effect independent of any mediators or statistical controls, \(X\)’s effect can also be indirect via one or more mediators. That is, \(X\)’s effect on \(Y\) may be the result of \(X\) affecting the mediating variable(s), which in turn causally affect \(Y\). In this approach, the relationship between an independent variable and a dependent variable is decomposed into direct and indirect (mediated) effects. The simplest example of mediation is a three-variable recursive model, which provides two causal paths feeding into a single dependent variable; the independent variable and the mediating variable directly affect the dependent variable, while the independent variable directly affects the mediator.

Unstandardized Simple Mediation Models: Evidence for mediation occurs when the relationship between two variables can be partially or totally accounted for by an intervening variable, which is the mediator. The significance of the mediated effect can be tested with a series of regression equations. There are at least three linear regression equations needed to build a simple mediation model (Fig. 1, panel I and panel ii). Those equations are as follows:}

Corresponding Author: Anwar Fitrianto, Department of Mathematics, Faculty of Science Universiti Putra Malaysia, Malaysia.
Fig. 1: Basic simple mediation model

Y = i1 + cX + ε1,  
M = i2 + aX + ε2,  
Y = i3 + c'X + bM + ε3,

where Y is the dependent variable, X is the independent variable, M is the mediating variable or mediator. Unstandardized coefficient c represents the relation between the independent variable to the dependent variable in the first equation, c is the parameter relating the independent variable to the dependent variable adjusted for the effects of mediator, a is unstandardized coefficient of the relationship between X and M and b is the parameter relating the mediator to the dependent variable adjusted for the effects of the independents variable to the mediating variable.

Note that the ε1, ε2 and ε3 represent error variability and i1, i2 and i3 are the intercepts. The intercepts are not involved in the estimation of mediated effects and could be left out of the equations, [7]. Note that, both c and c' are parameters relating the independent variable to the dependent variable, but c' is a partial effect, the adjusted effects of the mediator.

The variance of X is $\sigma_X^2$ and $\sigma_{\epsilon_1}^2$ is the error variance, namely the variance in Y not accounted for by X. Panel ii) shows the model in which the intervening variable is included. In this model, there are three relationships of interest. First, X influences the mediator, M and is denoted by a, a regression coefficient obtained from a simple regression of M on X. The error variance for this regression is denoted by $\sigma_{\epsilon_2}^2$. Second, the relationship between X and Y is examined again, this time with the mediator, M, also included as a predictor of Y.

From this analysis, both the influence of M on Y (path b) and the influence of X and Y (path c', also called the direct effect of X) can be obtained.

There are assumptions required for identifying mediation, including assumptions necessary for the statistical methods used to estimate the strength of the relationship in the regression models. The assumptions for the mediator model are necessary to yield accurate results. Cohen, [8], clearly said that each mediation regression equation requires the usual assumptions for regression analysis. These assumptions are the correct functional form, no omitted influences, accurate measurement and well-behaved residuals. Unfortunately, all those estimation use unstandardized linear regression coefficients. The slope coefficients, c, c', a and b depends on the units in which the predictors and outcome are measured, so that if either or both were measured in different units, the slope coefficients would change.

**Building Standardized Simple Mediation Model:**

The magnitudes of estimated unstandardized slope parameters in multiple regression depend on the scales of the predictor and outcome variables. The unstandardized coefficient does not tell us which independent variable is the strongest predictor in one sample. However, the standardized slope is interpreted as the estimated number of standard deviations of change in the dependent variable for one standard deviation unit change in the independent variable, controlling for other independent variables. This index can be compared across studies in much the same way that standardized mean difference effects are compared.

Another issue in unstandardized linear regression coefficient is that we must be cautious about interpreting the magnitudes of the regression coefficients as indicators of the relative importance of variables in the equation. They are partials, reflecting the variation accounted for when all other predictors are in the model and they may have different measurement units and variances. If they are standardized (beta weights) they become more comparable. Therefore unstandardized regression coefficients cannot be considered to reflect the relative importance of individual variables and the interpretation of standardized beta weights is extremely limited by their context-dependent nature ([8, 9]).

In linear regression, standardized betas are often used to compare strength of prediction across variables. The predictors are placed on a common scale so that each
has the same mean and standard deviation. Variables having larger absolute value of standardized beta weights are considered as stronger predictors in the equation. Though some researchers discourage the use of standardized coefficients, [10] or warn against interpreting them as indicators of variable importance, [11], the most recent publication manual of the American Psychological Association, [12], they encourage their routine application and reporting. Standardized beta weights are especially useful when variables are measured on an arbitrary scale.

**Computing Standardized Coefficients in Linear Regression:** It is well-known that the standardized regression coefficient in a bivariate regression model is the same as the bivariate correlation coefficient between the independent and dependent variables. In multiple regression analysis, standardized regression coefficients are scale free estimates and are related to correlation coefficients, but the relationship is much more complex than in bivariate regression.

Consider a regression model with two independent variables. When two independent variables are included in a multiple regression model, the model can be defined as

\[ y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{4} \]

where the \( y_i \) is the score on the dependent variable of the \( i \)th subject, \( x_{1i} \) and \( x_{2i} \) are the values of the independent variables for the \( i \)th subject, \( \beta_0, \beta_1, \) and \( \beta_2 \) are population regression coefficients and \( \epsilon_i \) is the error term assumed to be normally distributed with mean of zero and constant variance. To obtain standardized regression coefficients, we can transform the variables into standardized form. From the multiple regression equation with two predictors, the mean of the dependent variable is obtained as \( \bar{y} = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 \). The mean-deviation form for the above regression model of equation (4) can be obtained by subtracting \( \bar{y} \) as follows:

\[ y_i - \bar{y} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i - (\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2) \tag{5} \]

\[ = \beta_1 (x_{1i} - \bar{x}_1) + \beta_2 (x_{2i} - \bar{x}_2) + \epsilon_i \]

Then we divide by the standard deviation of \( y \) on both sides and multiply each term on the right side by \( 1 \) in the form of \( \frac{S_{x_i}}{S_y} \), to obtain the following equation:

\[ \frac{y_i - \bar{y}}{S_y} = \beta_1 \frac{S_{x_1}}{S_y} (x_{1i} - \bar{x}_1) + \beta_2 \frac{S_{x_2}}{S_y} (x_{2i} - \bar{x}_2) + \frac{\epsilon_i}{S_y} \tag{6} \]

The Equation (6) is then transformed through transforming its variables of \( \frac{y_i^* - \bar{y}}{S_y} \), as follows:

\[ z_{1i} = \frac{(x_{1i} - \bar{x}_1)}{S_{x_1}}, \quad z_{2i} = \frac{(x_{2i} - \bar{x}_2)}{S_{x_2}} \]. Meanwhile, the respective standard deviations are defined as:

\[ S_{x_1} = \sqrt{\frac{\sum(x_{1i} - \bar{x}_1)^2}{n-1}}, \quad S_y = \sqrt{\frac{\sum(y_j - \bar{y})^2}{n-1}} \]

and \( S_{x_2} = \sqrt{\frac{\sum(x_{2i} - \bar{x}_2)^2}{n-1}} \). Then, the transformed response variable, \( y_i^* \), is written as:

\[ y_i^* = \beta_1^* z_{1i} + \beta_2^* z_{2i} + \epsilon_i^* \tag{7} \]

The relation between the ordinary multiple regression coefficients and the standardized regression coefficients then can be written as:

\[ \beta_j^* = \frac{S_{x_j}}{S_y} \beta_j, \quad (j=1, \ldots, p-1), \] where \( \beta_j^* \) are the standardized regression coefficients in population.

**Accommodating Standardized Coefficients into Simple Mediation Models:** One reason to accommodate the standardized linear regression coefficient is that for variables with no natural metric, a “scale-free” standardized coefficient may be more meaningful than an unstandardized coefficient. The use of standardized coefficients in OLS transforms the independent variable into a variable measured in standard deviation units. This transformation will have huge impact in mediation analysis, in general, since a mediation model consists of several linear regression equations. The advantage to this is that a one standard deviation change is known to be a substantial change relative to the range of the independent variable.

A Simple mediation model involves three regression equations as in Eq. (1), Eq. (2) and Eq. (3) on which it contains both simple and multiple linear regressions. When the beta coefficients are standardized, the interpretation of the mediation analysis will be more meaningful.
Table 1: Summary statistics for unstandardized and standardized simple mediation model for Harris and Rosenthal’s Data.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Mediated Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstandardized Estimates</td>
<td>0.7078</td>
<td>0.8401</td>
<td>0.4257</td>
<td>0.3560</td>
</tr>
<tr>
<td>Standardized Estimates</td>
<td>0.5520</td>
<td>0.6533</td>
<td>0.4250</td>
<td>0.2776</td>
</tr>
<tr>
<td>p value</td>
<td>0.0002</td>
<td>&lt;.0001</td>
<td>0.0152</td>
<td>0.1089</td>
</tr>
<tr>
<td>$\hat{a}(\theta)$</td>
<td>0.1734</td>
<td>0.1580</td>
<td>0.1665</td>
<td>0.2142</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.3047</td>
<td>0.4268</td>
<td>0.3763</td>
<td></td>
</tr>
</tbody>
</table>

**Numerical Example:** To illustrate the importance of employing the standardized betas in simple mediation analysis, we present a numerical example available in [7], that is Harris and Rosenthal’s Data. Illustrations of the data are also available in [13]. The data contains of 40 subjects of teacher expectancies and student achievement. The example is originally based on [14] of four mediation models for how teacher expectancies affect student achievement. They described potential mediational processes for how expectancies about a person’s behavior lead to actual changes in behavior. The dependent variable ($Y$) was the score on a test of basic skills after one semester in the classroom. The independent variable, $X$, is the teacher expectancy based on an intelligence test given to the student the previous year. There are two mediator variables, $M_1$ and $M_2$ to represent social climate and teacher input (the average observer rating of input to the student). For the purpose of this study, we arbitrary use $M_1$ as the mediator. It is hypothesized that the teacher expectancy may lead the teacher to teach more material and more difficult material that may lead to increased student achievement.

RESULTS

Some statistics of the simple mediation analysis of the Harris and Rosenthal’s data is shown in Table 1. The teacher expectancy is significantly related to the score on a test of basic skills ($\hat{c} = 0.7078; se = 0.1734$) providing evidence that there is a statistically significant relation between the independent and the dependent variable.

But since that coefficient is unstandardized coefficient, we must be careful in interpreting the coefficient, as has been mentioned in the previous section. Instead of that, it is wise to interpret it based on the standardized coefficient, which is equal to 0.5520. One unit score increase in teacher expectancy is associated with roughly a half increase of the basic skill score.

There was a statistically significant effect of teacher expectancy on social environment ($\hat{a} = 0.8401; se = 0.1580$) and according to its associated standardized coefficient, it can be seen that one unit score increase in teacher expectancy is associated with about 85% increase of the score of social climate. The relation of the social climate (mediator) on the basic skill score of students is statistically significant ($\hat{b} = 0.4237; se = 0.1665$) when controlling for teacher expectancy. A 1 unit change in the score of social climate is associated with an increase of 0.4250 basic skill score. Meanwhile, the adjusted effect of teacher expectancy does not contribute statistically to the response variable when the mediator is in the model ($\hat{c}' = 0.3518; se = 0.2141$). There is a drop in the value of $\hat{c}' = 0.3518$ compared to $\hat{c} = 0.7078$.

The unstandardized estimate of the mediated effect is equal to $\hat{a}\hat{b} = 0.3560 = \hat{c} - \hat{c}' = 0.7078 - 0.3518$. Meanwhile, the standardized mediated effect of teacher expectancy through social climate is equal to $\hat{a}'\hat{b}' = 0.2776 = \hat{c}' - \hat{c}'' = 0.5520 - 0.2744$ score of basic skill. In another word, even though standardized regression coefficients are different from the unstandardized coefficients, but the standardized coefficients maintain the order of magnitude of the unstandardized regression coefficients for the simple mediation model.

ACKNOWLEDGMENTS

The authors are grateful to University Grants Scheme by Universiti Putra Malaysia for awarding the research grant.
REFERENCES


