SAS Macros for Testing Statistical Mediation in Data with Binary Mediators or Outcomes

By: Srichand Jasti, William N. Dudley, and Eva Goldwater

Jasti, S., Dudley, W. N., & Goldwater, E. (2008). SAS macros for testing statistical mediation in data with binary mediators or outcomes. Nursing Research, 57(2), 118-122.

Abstract:

Background: Statistical mediation is an important tool in behavioral health sciences, but it has been confined primarily to continuous variables. As prevention studies become increasingly common, more often the mediator or outcome is binary. Recent work by D. P. MacKinnon and J. H. Dwyer (1993) has explicated the steps necessary to estimate models for mediation when the mediator or the outcome is binary.

Objective: To report the release of a set of SAS macros used to implement the statistical analyses required to analyze data with binary and continuous-level data.

Approach: A brief introduction to the methodology of mediation analysis in the presence of a binary outcome, mediator, or both is provided. The macros are tested on a sample of 84 participants who were experiencing pain. It is hypothesized that the relationship between pain and fatigue is mediated by sleep disturbance.

Results: The relationship between pain and fatigue was mediated by the presence of sleep disturbances, and the amount of mediation was 23.34%.

Discussion: The SAS macros are available for download without charge from the second author's Web site. Instructions are provided in an included technical manual.

Article:

Statistical mediation can be a useful statistical technique for understanding complex relationships among three or more variables. Baron and Kenny (1986) set the stage for a precise definition of mediation and statistical methods for testing mediation, both of which are used commonly today. This statistical process has been well-articulated and supported by considerable methodological work by MacKinnon and others (MacKinnon, 1994; MacKinnon & Dwyer, 1993; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The process has been well-described conceptually by Bennett (2000). Programming that is helpful to estimating mediation models has been published by Preacher and Leonardelli (2001) and Dudley, Benuzillo, and Carrico (2004). In general, the statistical analyses that are used to test mediation are very straightforward. However, when the mediator or outcome is binary, the programming becomes more complex.

The purposes of this presentation are to briefly introduce the reader to the concepts underlying mediation modeling, to examine how the modeling must be altered to allow for binary data, to illustrate how the computations differ with binary data, and to provide an example of the application of programming to correctly estimate these models.

In Baron and Kenny's (1986) terms, a mediator is a variable that accounts for the correlation between two other variables. For instance, Williamson and Schulz (1992) demonstrated that physical dysfunction mediated the relationship between pain and depression in patients suffering from chronic pain. This work indicated that the oft-observed correlation between pain and depression was due in large part because the patients' pain interfered with their ability to function and that this inability to function was related directly to depression. In related work in nursing research, mediation was employed in a study of pain and functioning in older adults (Bennett, Stewart, Kayser-Jones, & Glaser, 2002). Mediation analyses have been used in the study of childhood suicide (Chang, Lin, & Lin, 2007) and in the study of adolescent risk behavior (Knauth, Skowron, & Escobar, 2006). Finally, in recent works by Dudley and others (Barsevick, Whitmer, Nail, Beck, & Dudley, 2006; Beck, Dudley, & Barsevick, 2005; Dudley, 2003; Dudley, Beck, & Barsevick, 2003), the use of mediation modeling was demonstrated in the study of symptom clusters in cancer treatment.

The analyses used in this research focus on continuous-level outcomes and mediation. However, in much of nursing research, variables are often inherently binary (such as a single symptom in a checklist of whether a treatment was administered or not) or can be dichotomized (as in a measure that has clinical cutoff scores). Mediation models with these types of data cannot be performed with currently available utilities.

Statistical Analyses for Test of Mediation

Mediation analysis is used in situations where it is required to test whether an observed relationship between an independent variable (X) and a dependent variable (Y) is true or whether a mediator variable (M) accounts for the relationship between them. Let us begin by considering the model in Figure 1(a). In the absence of M, the relationship between X and Y can be estimated using the simple linear regression model shown in Equation 1. The coefficient c (the total effect) denotes the effect of the independent variable X on the dependent variable Y.

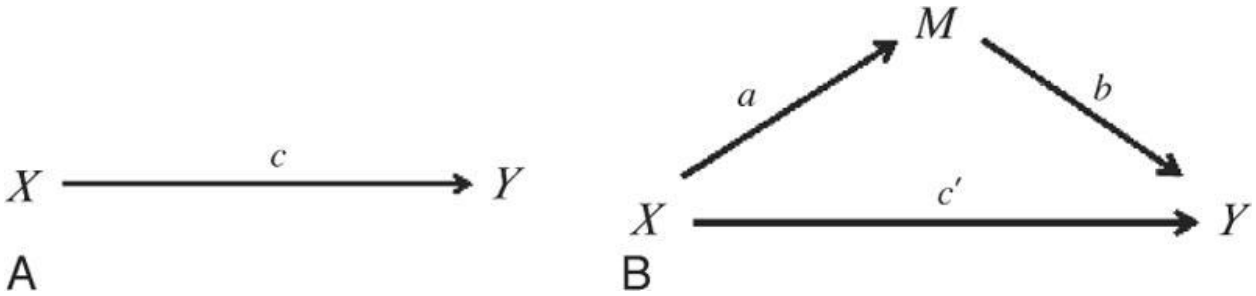


FIGURE 1. Mediation model.

$$Model\ 1:\ Y = a_0 + cX + \varepsilon_1 \qquad (1)$$

Next, test if a given variable, M, mediates the total effect; see Figure 1(b). In this figure, there are three variables, X, Y, and M. The path coefficients a, b, and c' are estimated by fitting two linear regression models, as shown in Equation 2.

$$\text{Model 2: } Y = a_1 + c'X + bM + \varepsilon_2 \quad (2)$$

$$\text{Model 3: } M = a_2 + aX + \varepsilon_3$$

The estimate c' is computed by regressing X on Y while controlling for M. This estimate is the direct effect. An estimate for the mediated effect is calculated by subtracting the direct effect from the total effect, as shown in Equation 3.

$$\text{Amount of Mediation} = c - c' = ab \quad (3)$$

Sobel Test

The Sobel test assesses whether this estimate of mediation due to variable M is statistically significant. As shown in Equation 4, the Sobel test involves dividing the effect of mediation by its standard error to arrive at a z score that is used to compute the p value.

$$\text{Sobel test} = \frac{c - c'}{\sigma_{c-c'}} = \frac{ab}{\sigma_{ab}} \quad (4)$$

$$\text{where } \sigma_{ab} = \sqrt{\sigma_b^2 a^2 + \sigma_a^2 b^2}$$

When the mediator and outcome are measured on a continuous level, the analyses can be accomplished with linear regression as implemented in standard statistical packages such as SPSS or SAS.

As discussed earlier, this process has been well-articulated and supported (Dudley et al., 2004; Preacher & Leonardelli, 2001). However, when the outcome or mediator is a binary measure, the analyses become more complex. The problem lies in the scaling of the coefficients derived from logistic models used when either the outcome or the mediator is binary in nature. Furthermore, as the coefficients are in different scales, Equation 3 no longer holds. Simulation studies show that ab is a better estimate of mediation than c-c' (MacKinnon & Dwyer, 1993).

Mediation in the Presence of Binary Variables

With this introduction to mediation in the simplest case of all the variables being continuous, next considered are complicated scenarios, where one or more of the variables involved are binary in nature. Linear regression is not appropriate for fitting equations in which the dependent variable is binary. Logistic regression is one way to model a binary outcome. The complication is that the scale in logistic regression is not constant across models, as is the case with linear regression (MacKinnon & Dwyer, 1993). Therefore, coefficient estimates cannot be combined from different models, as required by Equation 3, to estimate the mediation effect. In cases where either one or both of the M and the independent variable (Y) are binary, path coefficients are estimated using models with different scales. For example, in a case where Y is a binary variable and M is a continuous variable, the coefficients c, c', and b are computed using logistic regression, whereas the estimate for coefficient a is computed using a linear regression. Using these coefficients in the Sobel test will result in an incorrect test because the coefficients are not in the same scale. MacKinnon and Dwyer (1993) suggest a method of standardization that brings logistic regression coefficients estimated using different regression models into the same scale. Once the coefficients have been standardized, the estimate of mediation and the Sobel test are calculated using Equations 3 and 4 with the standardized coefficients.

In an example case in which the independent variable X and the mediator M are continuous and the dependent variable Y is binary, the models to estimate the coefficients are shown in Equation 5.

$$
\begin{aligned}
&\textit{Model 1 (Logistic)}: \\
&\quad Y = a_0 + \underline{c}X + \varepsilon_1 \\
&\textit{Model 2 (Logistic)}: \\
&\quad Y = a_1 + \underline{c}'X + \underline{b}M + \varepsilon_2 \\
&\textit{Model 3 (Linear)}: \\
&\quad M = a_2 + aX + \varepsilon_3
\end{aligned}
\qquad (5)
$$

In the above models, the coefficients c, c', and b are derived from different logistic models and cannot be combined with each other or with the coefficient from Model 3. Hence, we need to standardize the logistic coefficients by dividing them by the standard error of the model from which those coefficients arose. For example, the coefficient c needs to be standardized by the standard error of Model 1, and coefficients c', and b need to be standardized using the standard error of Model 2.

The standard error of Model 1 is computed as follows:

$$\sigma(Y)_1 = \sqrt{\sigma^2(Y)} \qquad (6)$$

$$= \sqrt{\underline{c}^2\sigma^2(X) + \sigma^2(\varepsilon_1)}$$

$$where,\ \sigma^2(\varepsilon_1) = \frac{\pi^2}{3}$$

Similarly, the standard error of Model 2 is computed as follows:

$$\sigma(Y)_2 = \sqrt{\sigma^2(Y)} \qquad (7)$$

$$= \sqrt{\underline{c}'^2\sigma^2(X) + \underline{b}^2\sigma^2(M) + 2\underline{c}'\underline{b}\sigma(X,M) + \sigma^2(\varepsilon_2)}$$

$$where, \sigma^2(\varepsilon^2) = \frac{\pi^2}{3}$$

Standardized coefficients are computed by dividing each coefficient by the standard error of its model.

$$Std(\underline{c}) = \frac{\underline{c}}{\sigma(Y)_1} \qquad (8)$$

The coefficients from Model 2 are standardized as

$$Std(\underline{c}') = \frac{\underline{c}'}{\sigma(Y)_2}, \text{ and}$$

$$Std(\underline{b}) = \frac{\underline{b}}{\sigma(Y)_2} \quad (9)$$

These standardized coefficients are then used to compute the estimate of mediation and the Sobel test, according to Equations 3 and 4.

Approach

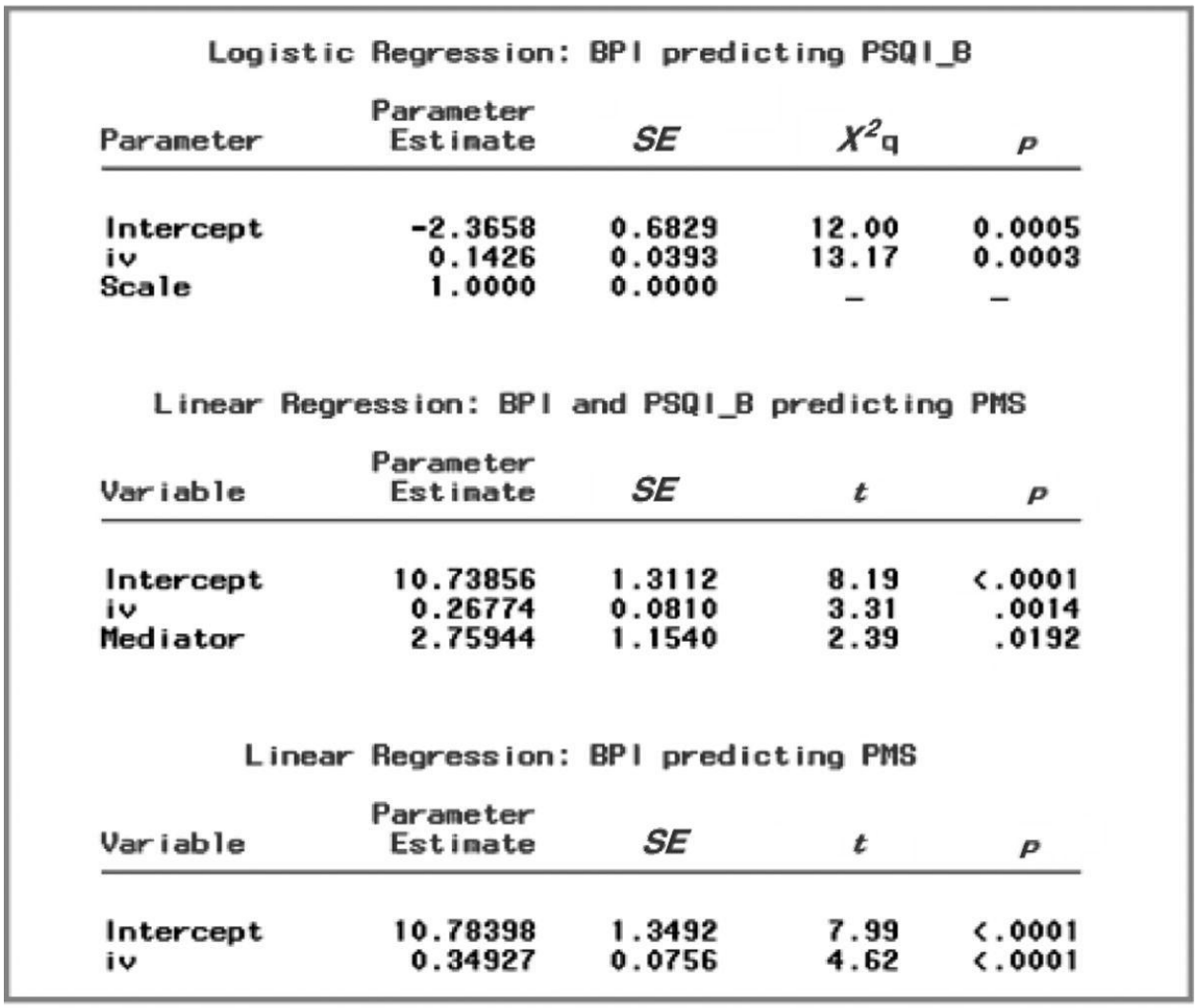
The mixed mediation macro was tested on the pain, sleep disturbance, and fatigue data (Beck et al., 2005). These data consist of 84 patients with cancer who were experiencing pain.

All participants completed the Brief Pain Inventory short form, the Pittsburgh Sleep Quality Index (PSQI) questionnaire, and the fatigue subscale of the Profile of Mood States (PMS) questionnaire. All of the variables are in continuous scale. The PSQI variable was dichotomized (PSQI_B) to provide an example of mixed mediation where M is a binary variable. Participants that scored under the median PSQI value were coded as 0, and those above the median were coded as 1. It is hypothesized that sleep disturbance (PSQI_B) mediates the effects of pain (Brief Pain Inventory) on fatigue (PMS). The following section describes the output obtained from the SAS mixed mediation macro.

Results

The first section of the output contains frequencies and correlations which are used by the macro to determine which variables are dichotomous and which variables have more than two levels of measurement, in which case those variables are treated as continuous. The next three outputs tabs contain three regression outputs either from proc GENMOD (used for logistic regression) or proc REG (used for linear regression). The choice of procedure is based on the presence of varying combinations of binary variables.

The regressions from the previous output are summarized in the next three reports which are highlighted in the top box of Figure 2.



Logistic Regression: BPI predicting PSQI_B

Parameter	Parameter Estimate	*SE*	X^2q	*p*
Intercept	-2.3658	0.6829	12.00	0.0005
iv	0.1426	0.0393	13.17	0.0003
Scale	1.0000	0.0000	_	_

Linear Regression: BPI and PSQI_B predicting PMS

Variable	Parameter Estimate	*SE*	*t*	*p*
Intercept	10.73856	1.3112	8.19	<.0001
iv	0.26774	0.0810	3.31	.0014
Mediator	2.75944	1.1540	2.39	.0192

Linear Regression: BPI predicting PMS

Variable	Parameter Estimate	*SE*	*t*	*p*
Intercept	10.78398	1.3492	7.99	<.0001
iv	0.34927	0.0756	4.62	<.0001

Establishing Mediation
Paths *a*, *b*, and *c*

a	sa	Std *a*	Std sa	*b*	sb	*c*	sc
0.1426	0.0393	0.0685	0.0189	2.7594	1.154	0.3493	0.0756

FIGURE 2. Regression output and standardized coefficients. PSQI_B =Pittsburgh Sleep Quality Index B; PMS = Profile of Mood States; BPI = Brief Pain Inventory.

The first logistic regression output shows that the PSQI_B is a significant predictor of PMS. This regression corresponds with Model 3 in Equation 2. Similarly, the next two regressions correspond to Model 2 in Equation 2 and Model 1 in Equation 1, respectively. All the explanatory variables are statistically significant in both regressions.

Now, the Sobel, Goodman I, and Goodman II tests are conducted to see if sleep disturbance mediates the effect between pain and fatigue. However, the logistic regression estimates and linear regression estimates are not in the same scale and hence cannot be used directly in the abovementioned tests. Therefore, the coefficients and the standard deviations must be standardized using formulae similar with Equations 6 and 8. Because there was only one logistic regression performed to estimate the effect of path a, only this value needs to be standardized. The bottom box in Figure 2 provides the standardized estimates. Each estimate is followed by its standard error and by the standardized estimates if necessary. The estimate of path a before standardization is -0.143, and after standardization, it is -0.069.

Using the standardized coefficients, it is possible to conduct the tests for mediation. The output for this section is shown in the top box in Figure 3.

Sobel Test

Indirect Effect	*SE*	Test statistic	*p*
0.1890366	0.0946743	1.9967055	.045857

Goodman I

Indirect Effect	*SE*	Test statistic	*p*
0.1890366	0.0971485	1.9458521	.051673

Goodman II

Indirect Effect	*SE*	Test statistic	*p*
0.1890366	0.09213359	2.05176629	.040192

Percent Mediated

Percent of the total effect that is mediated	Ratio of the indirect to the direct effect
23.34	0.7060577

FIGURE 3. Tests and measures of mediation.

The Sobel test and the Goodman II test show that the mediation is statistically significant, and the Goodman II test is borderline significant. Hence, it can be concluded that the mediation is statistically significant, suggesting that the relationship between pain and fatigue is mediated by sleep problems. The lower box in Figure 3, under the percent of total effect that is mediated, is used to get an estimate of the amount of mediation. In the case shown here, the percent mediated is 23.34%; that is, sleep disturbance accounted for 23.34% of the effect of pain causing fatigue. Yet, another measure of effect of the mediator is the ratio of indirect to the direct effect. The measure in this study was 0.706, suggesting that the mediated (indirect) effect is 70.6% of the direct effect of pain on fatigue.

Discussion

There is a set of SAS macros to allow for mediation analyses when either the outcome or the mediator is binary in nature. These macros will run on SAS Version 8.2 or higher. Readers interested in testing this type of mediation are encouraged to download the files and provide the authors with feedback on the usefulness of the programming for their use. The macros, with an instruction manual for how to modify the macros to meet specific needs, and example files are available from http://www.uncg.edu/hhp/oor/ .

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173-1182.

Barsevick, A. M., Whitmer, K., Nail, L. M., Beck, S. L., & Dudley, W. N. (2006). Symptom cluster research: Conceptual, design, measurement, and analysis issues. Journal of Pain and Symptom Management, 31(1), 85-95.

Beck, S. L., Dudley, W. N., & Barsevick, A. (2005). Pain, sleep disturbance, and fatigue in patients with cancer: Using a mediation model to test a symptom cluster. Oncology Nursing Forum, 32(3), E48-E55.

Bennett, J. A. (2000). Mediator and moderator variables in nursing research: Conceptual and statistical differences. Research in Nursing & Health, 23(5), 415-420.

Bennett, J. A., Stewart, A. L., Kayser-Jones, J., & Glaser, D. (2002). The mediating effect of pain and fatigue on level of functioning in older adults. Nursing Research, 51, 254-265.

Chang, H. J., Lin, M. F., & Lin, K. C. (2007). The mediating and moderating roles of the cognitive triad on adolescent suicidal ideation. Nursing Research, 56, 252-259.

Dudley, W. N. (2003). Cancer-related fatigue, depression, and function: The test of a mediation model. Paper presented at the Seventh National Conference on Cancer Nursing Research, San Diego, CA.

Dudley, W. N., Beck, S., & Barsevick, A. (2003). Mediation: Concepts, statistical tests, and application to cancer fatigue. Paper presented at the 7th National Conference on Cancer Nursing Research, San Diego, CA.

Dudley, W. N., Benuzillo, J. G., & Carrico, M. S. (2004). SPSS and SAS programming for the testing of mediation models. Nursing Research, 53, 59-62.

Knauth, D. G., Skowron, E. A., & Escobar, M. (2006). Effect of differentiation of self on adolescent risk behavior. Nursing Research, 55, 336-345.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention research. NIDA Research Monograph, 139, 127-153.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. Evaluation Review, 17(2), 144-158.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7(1), 83-104.

Preacher, K. J., & Leonardelli, G. J. (2001). Calculation for the Sobel test: An interactive calculation tool for mediation tests. Retrieved February 12, 2007, from http://quantrm2.psy.ohio-state.edu/kris/sobel/sobel.htm

Williamson, G. M., & Schulz, R. (1992). Pain, activity restriction, and symptoms of depression among community-residing elderly adults. Journal of Gerontology, 47(6), P367-P372.