May 19, 2008

This is a draft document to assist researchers. Please do not cite or quote without the author's permission. I thank both Nathaniel R. Herr and Craig Nathanson who suggested corrections to a prior document. Readers might benefit by consulting the page by Nathaniel Herr: http://www.nrhpsych.com/mediation/logmed.html

Mediation with Dichotomous Outcomes

David A. Kenny

University of Connecticut

This note is on the testing of mediation using logistic regression and is largely based on a paper by MacKinnon and Dwyer (1993). The interested reader should consult their paper for more details. Consider the equations:

$$Y = cX + E_1$$

$$M = aX + E_2$$

$$Y = bM + c'X + E_3$$

where X is the causal variable, Y the outcome, and M the mediator and each is a dichotomy. The coefficients in the equations are estimated by logistic regression.

The usual decomposition of effects c' + ab = c. The problem is that when a variable is used as a predictor in logistic regression, it has a different scale from when it is an outcome variable. The mediation equations need to be rewritten to show the need as follows:

$$Y' = cX + E_1$$

$$M' = aX + E_2$$

$$Y'' = bM + c'X + E_3$$

M', Y', and Y" refers to the “new” M and Y variables differs from the scale of original M and Y. They differ because the error variances are fixed to one.

To make the coefficients comparable across equations, we multiply each coefficient by the standard deviation of the predictor variable and divide by the standard deviation of the outcome variable. The standard deviation of M and X can be computed in the ordinary way. However, the variances of outcome variables in the three above equations are as follows

$$Y': c^2V(X) + \pi^2/3$$

$$M': a^2V(X) + \pi^2/3$$

$$Y'': c'^2V(X) + b^2V(M) + 2bc'C(X,M) + \pi^2/3$$

where $\pi = 3.1416....$ These variances would be square rooted and used in the standardization.

Note also that if either M or Y is measured at the interval level of measurement, ordinary multiple regression coefficients can be used when each is the outcome variable in the analysis and there would be no need to rescale the coefficients from that equation.

An alternative and equivalent estimation strategy (and perhaps easier) is to use as the variances:

$$Y': V(Y'^*) + \pi^2/3$$

$$M': V(M'^*) + \pi^2/3$$

$$Y'': V(Y''^*) + \pi^2/3$$

where Y'*, M'*, and Y"* means the variance in the predicted scores.

After each coefficient has been standardized (multiplied by the standard deviation of the predictor variable and divided by the standard deviation of the outcome variable), the decomposition can be computed. Unlike multiple regression, ab + c' only approximately equals c.

If a probit regression analysis were used, the same procedure would be used, but instead we would substitute 1 for π^2 /3.

Illustration

We use the Morse et al. (1994) data that were reanalyzed by Kenny, Kashy, and Bolger (1988). The variables are

X: condition (0 = control; 1 = treated)

Y: days of stable housing in a month (0 < 15 days; 1 > 14 days)

M: housing services received (0 < 1 contact; 1 greater than or equal to 1 contact)

The SPSS output from Step 1 is:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	treatment	.852	.400	4.531	1	.033	2.344
	Constant	-.223	.254	.775	1	.379	.800

a Variable(s) entered on step 1: treatment.

The SPSS output from Step 2 is:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	treatment	1.165	.413	7.936	1	.005	3.205
	Constant	-1.078	.289	13.860	1	.000	.340

a Variable(s) entered on step 1: treatment.

The SPSS output from Step 3 and 4 is:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	treatment	.561	.427	1.727	1	.189	1.752
	hc_dic	1.346	.451	8.907	1	.003	3.844
	Constant	-.566	.286	3.914	1	.048	.568

a Variable(s) entered on step 1: treatment, hc_dic.

Term	Pred.	Outcome	Estimate	SE	Z	s^2Pred.	s^2Out.	Stan.Coef.
a	X	M	1.165	0.413	7.936	3.624	0.246	0.304
b	M	Y	1.346	0.451	8.907	3.892	0.234	0.330
c'	X	Y	0.561	0.427	1.727	3.892	0.246	0.141
c	X	Y	0.852	0.400	4.530	3.469	0.246	0.227

Note that ab + c' equals 0.241 which is approximately equal to c or 0.227.

References

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233-265). Boston, MA: McGraw-Hill.

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

Morse, G. A., Calsyn, R. J., Allen, G., & Kenny, D. A. (1994). Helping homeless mentally ill people: What variables mediate and moderate program effects? *American Journal of Community Psychology, 22*, 661-683.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington DC: American Sociological Association.