How to test for mediation and moderation effects in biomedical research

Christophe Lalanne*

Octobre 2009

1 Overview

1.1 The multiple regression framework: A look into causal pathways

Figure 1 gives an outline of the different situations a researcher may faced when applying a multiple regression model, especially for data coming from an observationnal study or with community samples.

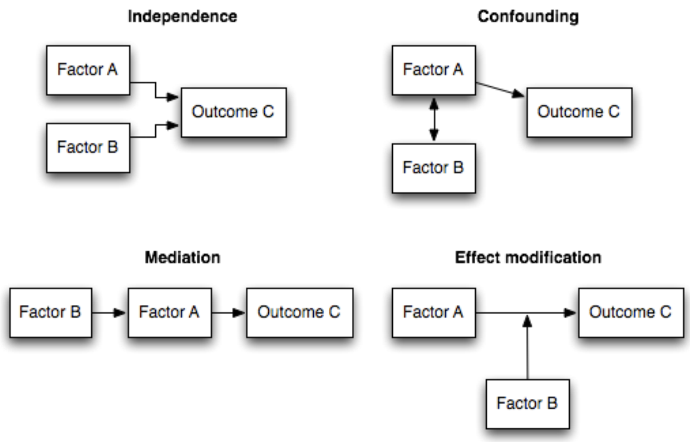


Figure 1: Different scenarios in multiple regression. (Reproduced from *Practical Psychiatric Epidemiology*[10], p. 250)

Several biomedical and psychological[6;14], or social[2] studies rely in fact on such an approach. The interested reader is referred to e.g. Aiken & West[1], Vittinghoff and coll.[15] (Chapter 4), or Rothman & Greenland[13] (Chapter 2) for a thorough discussion of their implications on causal interpretation.

According to Baron & Kenney[2], a mediator must satisfy the following conditions:

*This document is available on www.aliquote.org. Any suggestion for improvement is welcome.

1. the predictor must be significantly correlated with the hypothesized mediator;
2. the predictor must be significantly correlated with the outcome;
3. the mediator must be significantly correlated with the outcome;
4. the impact of the predictor on the outcome is no longer significant after controlling for the mediator.

Such mediated effect are known as indirect effect in *path analysis*[3]. Moderation is usually modelled as a usual linear by linear interaction in a regression model.

1.2 Some examples

To give a clear picture of the aforementioned direct and indirect effects, let's consider the following examples[14]:

- the relationship between neighborhood disadvantage (high rates of poverty, crime, and unemployment) and children's externalizing behavior is *mediated* by the intervening variable of parent-child conflict[11];
- the relationship between neighborhood disadvantage and children's internalizing behavior is *mediated* by mother's perceptions of neighborhood quallity such as cleanliness[4];
- the efficacy of treatment for depression depends on patients' attachment insecurity, such that cognitive-behavioral therapy is more effective than interpersonal psychotherapy for patients who are higher in attachment avoidance[8]: the effect of therapy is thus *moderated* by patients' attachment insecurity.

Finally, Rose[12] nicely noticed that if everyone smoked, lung cancer would appear to be genetic disease.

2 Mediation and interaction effects

2.1 Assessing Mediation effect

In what follows, the response variable, or outcome, will be denoted as y and we deliberately avoid using the "dedicated" term of *dependent* variable.

Data ($n = 200$) were simulated using a script provided by Thomas Fletcher on his homepage. We also make use of his R package **QuantPsyc** which implements methods proposed by MacKinnon and coll.[7].

Briefly, a given sample correlation matrix is used (with cholesky decomposition) to construct the observed data. Figure 2 shows how it looks like and pairwise correlation are below in the lower diagonal part of the table below:

	x	z	y
x	–	0.28	0.17
z	0.31	–	0.10
y	0.21	0.16	–

The partial correlation are shown in the upper-part of the correlation matrix, where the cell (x, y) refers to the correlation of x and y after z has been partialled out (See the R function `partial.r()` in the **psych** package).

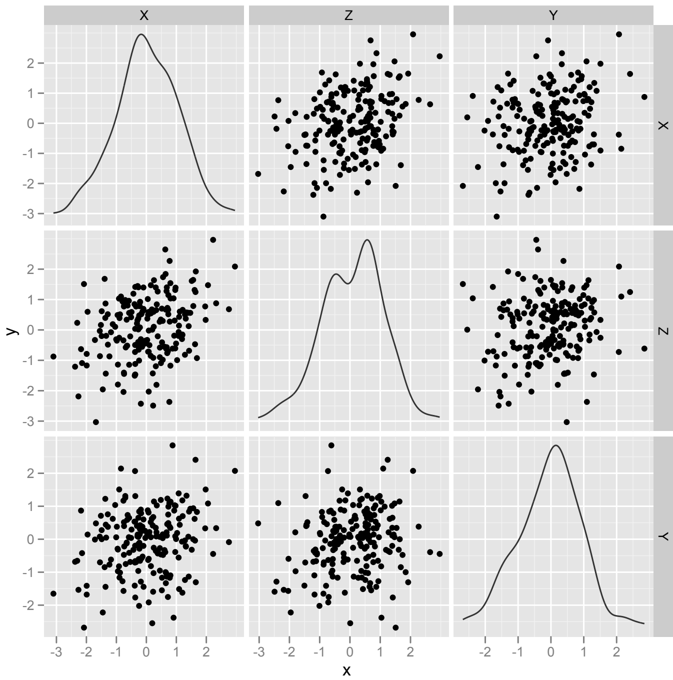


Figure 2: Joint distribution of x, y and z.

We use the (unadjusted) coefficient of determination (R^2) as a measure of effect size and Table 1 summarizes the results obtained when testing the different models. Note that the coefficient of determination can be a useful measure of mediation[5]. As can be seen in the Table, the univariate effects of z and x are significant, when considered separately, and the effect of x remains significant after accounting for the z in the $y \sim x + z$ model while that of z becomes non-significant. Thus, z acts as a mediation variable. This should be distinguished from a confusion effect which would tend to do the reverse (cancelling out the significance of x when z enters the model).

A more detailed output summary is provided by the function `proximal.med()`, see Table 2.

We can use bootstrap to get a more accurate estimate of the standard error for the direct effect. With 1000 replicates, we get a 95% CI of $[0.029; 0.104]$ (compared to the value 0.066 for $x \to y \mid z$ in Table 2). Here, indirect effect of x on y, i.e. the effect of the predictor through the mediator or $x \to z \to y$, is about 1/5 of the direct effect, and its

Table 1: Results for a simple mediation model.

Model	R^2	p	
$y \sim x$	0.045	0.003	
$y \sim z$	0.025	0.025	
$y \sim x + z$	0.054	0.014	(x)
		0.161	(z)

Table 2: Analysis of direct and indirect effects.

Path	Effect	SE
$x \to z$	0.301	0.066
$z \to y \mid x$	0.119	0.067
$x \to y$	0.199	0.063
$x \to y \mid z$	0.163	0.066
$x \to z \to y$	0.036	0.022

SE is estimated at 0.022. Using Aroian or Goodman's method[7], we would get similar SE estimates, namely 0.022 and 0.021.

2.2 Assessing Moderation effect

As previously discussed, moderation stands for the usual interaction effect whereby the effect of a given predictor depends on the value, or level, of another variable. Interaction can be of two kinds: We speak of *qualitative* interaction when the effect of x on y is reversed depending on the level considered for z (we also speak of "crossed" interaction), or *quantitative* when z positiveley or negatively enhances the effect of x on y.

3 Application

The following is inspired from a Stata tutorial that can be found at Stata FAQ, but see also Vittinghohh and coll.[15] (pp. 95–108, and examples on my website). We will be using data from Preacher and coll.[9], which can be downloaded from Stata using the following commands:

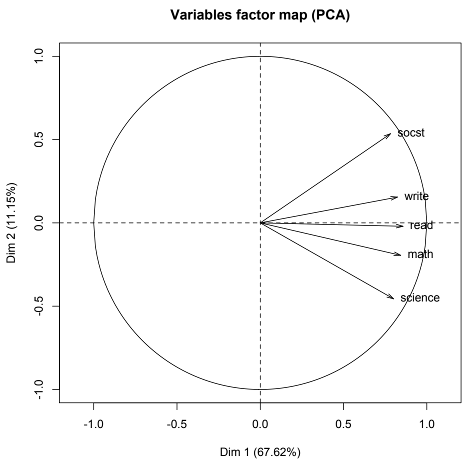
```
. use http://www.ats.ucla.edu/stat/data/hsb2, clear
```

or alternatively using the **foreign** package in R:

```
> require(foreign)
> hsb2 <- read.dta("hsb2.dta")
> head(hsb2)
   id female  race    ses schtyp    prog read write math science socst
1  70   male white    low public general   57    52   41      47    57
```

```
2 121 female white middle public vocation   68    59   53       63    61
3  86   male white   high public  general   44    33   54       58    31
4 141   male white   high public vocation   63    44   47       53    56
5 172   male white middle public academic   47    52   57       53    61
6 113   male white middle public academic   44    52   51       63    61
```

`Science` (y) is the outcome and `math` is the independent variable (x); `read` is a mediator, and `write` and `socst` are moderator variables. A rough PCA indicates that all five variables are correlated and account for 68% of the variance (with an eigenvalue of 3.38) on the first principal component (Figure 3).



	read	write	math	science	socst
read	1.000	0.597	0.662	0.630	0.621
write	0.597	1.000	0.617	0.570	0.605
math	0.662	0.617	1.000	0.631	0.544
science	0.630	0.570	0.631	1.000	0.465
socst	0.621	0.605	0.544	0.465	1.000

Figure 3: Correlation circle for the five variables of interest.

Let's first evaluate a simple model of mediation,

$$\boxed{\text{math}} \rightarrow \boxed{\text{read}} \rightarrow \boxed{\text{science}}$$

The commands

```
hsb2.med1 <- hsb2[,c("math","read","science")]
colnames(hsb2.med1) <- c("x","m","y")
proximal.med(hsb2.med1)
```

yield the results summarized in Table 3.

Table 3: Analysis of direct and indirect effects for the `hsb2` data.

Path	Effect	SE
$x \rightarrow z$	0.725	0.058
$z \rightarrow y \mid x$	0.365	0.066
$x \rightarrow y$	0.667	0.058
$x \rightarrow y \mid z$	0.402	0.073
$x \rightarrow z \rightarrow y$	0.265	0.053

If we test this model with Stata we obtain comparable results. The syntax used is:

```
. rename science y
. rename math x
. rename read m
. global m=r(mean)
. global s=r(sd)
. sureg (m x)(y m x)
```

which gives

```
Seemingly unrelated regression
----------------------------------------------------------------------
Equation          Obs  Parms        RMSE    "R-sq"       chi2        P
----------------------------------------------------------------------
m                 200      1    7.662848    0.4386     156.26   0.0000
y                 200      2    7.133989    0.4782     183.30   0.0000
----------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
m            |
           x |    .724807   .0579824    12.50   0.000     .6111636    .8384504
       _cons |   14.07254   3.100201     4.54   0.000     7.996255    20.14882
-------------+----------------------------------------------------------------
y            |
           m |   .3654205   .0658305     5.55   0.000     .2363951    .4944459
           x |   .4017207   .0720457     5.58   0.000     .2605138    .5429276
       _cons |    11.6155   3.031268     3.83   0.000     5.674324    17.55668
------------------------------------------------------------------------------
```

The Stata tutorial provides more complicated model such that one where the preceding indirect effect is also moderated by `math`.

4 Conclusion

Quoting R. Stewart[10] (p. 251):

> The importance of effect modification is inherently acknowledged in the diagnostic formulation for psychiatry and other medical specialties. In particular the division of identified potential causes into predisposing and precipitating factors acknowledges that single causes are usually insufficient to bring about the outcome and that 'precipitants' may require a 'predisposition' in order to exert their effects (and *vice versa*). However, despite this, statistical analyses for the majority of studies appear to be carried out entirely to distinguish between independence and counfounding, with no consideration of the possibility that causes might interact.

References

[1] L S Aiken and S G West. *Multiple regression: Testing and interpreting interactions*. Newbury park, CA: Sage Publications, 1991.

[2] R M Baron and D A Kenney. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51:1173–1182, 1986.

[3] K A Bollen. Total, direct and indirect effects in structural equation models. In C C Clogg, editor, *Sociological Methodology*, pages 37–69. Washington, DC: American Sociological Association, 1987.

[4] S Deng, V Lopez, M W Roosa, E Ryu, G L Burrell, and J Y et al Tein. Family processes mediating the relationship of neighborhood disadvantage to early adolescent internalizing problems. *Journal of Early Adolescence*, 26:206–231, 2006.

[5] A J Fairchild, D P MacKinnon, M P Taborga, and A B Taylor. r^2 effect-size measures for mediation analysis. *Behavior Research Methods*, 41:486–498, 2009.

[6] William B Jaffee and Thomas J D'Zurilla. Personality, problem solving, and adolescent substance use. *Behavior Therapy*, 40:93–101, 2009. http://www.aliquote.org/pub/Jaffee.behavther2009.pdf.

[7] D P MacKinnon, C M Lockwood, J M Hoffman, S G West, and V Sheets. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7:83–104, 2002.

[8] C McBride, L Atkinson, L C Quilty, and R M Bagby. Attachment as moderator of treatment outcome in major depression: A randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology*, 74:1041–1054, 2006.

[9] K J Preacher, D D Rucker, and A F Hayes. Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1):185–227, 2007. http://www.aliquote.org/pub/Preacher.mbr2007.pdf.

[10] M Prince, R Stewart, T Ford, and M Hotopf, editors. *Practical Psychiatric Epidemiology*. Oxford University Press, 2003.

[11] M W Roosa, S Deng, E Ryu, G L Burrell, J Y Tein, and S et al Jones. Family and children characteristics linking neighborhood context and child externalizing behavior. *Journal of Marriage and Family*, 67:515–529, 2005.

[12] G Rose. Sick individuals and sick populations. *International Journal of Epidemiology*, 14:32–38, 1985.

[13] K J Rothman and S Greenland. *Modern Epidemiology*. Lippincott Williams & Wilkins, 2nd edition, 1998.

[14] E Ryu, S G West, and K H Sousa. Mediation and moderation: Testing relationships between symptoms status, functional health, and quality of life in hiv patients. *Multivariate Behavioral Research*, 44:213–232, 2009. `http://www.aliquote.org/pub/Ryu.mbr2009.pdf`.

[15] E Vittinghoff, D V Glidden, S C Shiboski, and C E McCulloch. *Regression Methods in Biostatistics*. Springer-Verlag, 2005.

$Id: mediation.tex,v 1.0 2009/10/09 chl$