Causality, Mediation, and Controlling for Post-Treatment Variables

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Multilevel Regression Modeling, 2009

Causality, Mediation, and Controlling for Post-Treatment Variables

Introduction

A substantial amount of work in clinical, social, and personality psychology is based on the analysis of *mediating variables* and their causal role.

Near the end of their Chapter 9, Gelman and Hill abruptly warn of the dangers of attempting to include post-treatment measures, and subsequently assessing causality, in observational settings.

Their examples are interesting and important, and I feel we might profit by exploring them in more explicit numerical detail than we saw in the textbook presentation.

Modeling a Post-Treatment Variable

Imagine T treatment, z parenting quality (mediator), y Child's IQ, PP an unmeasured background variable we might call parenting potential. z is related to the treatment as

$$z = 0.3 + 0.2T + 0.2PP + \nu \quad (1)$$

Then, given a certain level of parenting potential, we would see z at 0.2 units higher for parents receiving the treatment.

The model for estimating y given T is

$$y = \theta T + \epsilon \quad (2)$$

Now, suppose we add z to the model. Then we have

$$y = \theta^* T + \beta z + \epsilon^* \quad (3)$$

The Non-ignorability Problem

Suppose we have two families that are identical in their observed value of z (0.5), but one does not have the treatment, while the other does.

Then although the observed values of z are the same, the potential outcomes on z could be different across treatment groups.

Consequently, given two families with the same intermediate outcome, one that received the treatment actually has lower underlying parenting skills.

Thus, in the regression of y on (T, z), the coefficient of T represents a comparison of families that differ in their underlying characteristics.

A Numerical Example

PP	T	z	z^1	z^0
1	0	.5	.5	.7
1	1	.7	.5	.7
0	0	.3	.3	.5
0	1	.5	.3	.5

Rows 1 and 4 of the table demonstrate how "controlling for z" equates families with different potential.

Simplified Example

Gelman and Hill give a simplified example of this phenomenon in their Figure 9.14. Below I present a slightly modified version of their figure.

z^0	z^1	y^0	y^1	p	parenting.potential
20	20	60	70	0.1	0
20	30	65	80	0.7	1
30	30	90	100	0.2	2

In the above table, we have knowledge that would not normally be available to us, because we know the counterfactuals. Given the above knowledge, how would you characterize the causal impact of z and T?

Regression Analysis

However, the above table translates into the following *observable* picture if we assume that $T = 0$ and $T = 1$ is applied equally often across PP types.

z	y	T	p	PP
20	60	0	0.05	0
20	70	1	0.05	0
20	65	0	0.35	1
30	80	1	0.35	1
30	90	0	0.1	2
30	100	1	0.1	2

If this is all we are allowed to see, how would we assess the effect of T and z?

Regression Analysis

Try filling in this table:

z	T	Mean	p
20	1		
20	0		
30	1		
30	0		

Regression Analysis

Here's what I got

z	T	Mean	p
20	1	70.00	0.05
20	0	64.38	0.40
30	1	84.44	0.45
30	0	90.00	0.10

In the observed data, within levels of z, T doesn't have much effect, and it is contradictory. On the other hand, within levels of T, z has a very substantial effect.

Regression Analysis

Check out what happens when we perform a regression analysis on these data.

```
> data ← read.table("Fig914B.txt",header=TRUE)
> attach(data)

        The following object(s) are masked from package:base :

         T


> lm(y~T+z,weights=p)


Call:
lm(formula = y ~ T + z, weights = p)

Coefficients:
(Intercept)              T              z
     21.98          -1.62           2.16
```