Introduction to Mediation and Moderation

**Solutions Exercises #1**

The instructions asked you to:

Use the dataset **hsbdemo.csv** to complete the following tasks:

**The .R script to run these exercises is available:**

Intro Mediation and Modearation OPerra exercise n1.R.

The solutions are illustrated in what follows.

1. Run a simple mediation model where Low SES influences Science scores through Reading scores.

To open the dataset use the command “read.csv” (change the path to the file accordingly to where you located the dataset csv file):

d <- read.csv("E:/Teaching/Mod\_Med/hsbdemo.csv")

In the rest of the exercise the dataset is going to be called “d” for dataset. I used “d” mostly to save typing.

Before running the analyses you will also have to load and run the PROCESS macro in your R session. Once you have done that, you can specify the simple model requested in this way:

process (data=d,y="science",x="lowinc",m="read",total=1,normal=1,

model=4,seed=83582)

The “data” option is used to specify the dataset to be used. The options “y”, “x”, and “m” are used to specify the dependent, independent, and mediator variables respectively. The names of the variables in dataset “d” are used between quotation marks. Thus, the model specify an effect from LowInc to Math, which is mediated by Read.

The other options in the command above are:

**model**🡪 Invokes different types of pre-set models and equations for the mediation.

**model =4** 🡪 estimates a model based on equations:

*M = im + aX +eM*

*Y = iy +cX +bM + eY*

and thereby provides parameters for paths “a”, “b”, and “c”, together with standard regression statistics such as *R2* for each of the equations.

**total**🡪 Invokes an estimation and test of the total effect from the predictor to the outcome.

**normal**🡪Invokes a Sobel test assuming the underlying distribution of the indirect effect is normal.

**seed**🡪 Allows to specify a specific seed, which ensures reproducibility of the results, when needed.

The main test of mediation is the test of the indirect effect. The Sobel test is displayed under the heading “Normal theory test for indirect effect(s)”:

Normal theory test for indirect effect(s):

Effect se Z p

read -3.0352 1.0266 -2.9566 0.0031

The *z* score of this test is estimated being equal to -3.03, equivalent to a *p* value equal to .003. Since the *p* value is less than .05, we can reject the null hypothesis that the indirect effect is equal to 0, and thus accept there is a substantial indirect effect.

1. Test the significance of the indirect effect of the previous model using the Sobel test and the Bootstrapping approach (use at least 10,000 draws). Compare the results of the two approaches.

The Sobel test for the simple mediation model specified above has been presented in response to exercise 1 before, whereby the PROCESS options required a normal test (normal=1) and the output displayed the “Normal theory test for indirect effect(s)”.

The *z* score of this test was estimated being equal to -2.96, equivalent to a *p* value equal to .003. Since the *p* value is less than .05, we can reject the null hypothesis that the indirect effect is equal to 0, and thus accept there is a substantial indirect effect.

The output of the exercise above also reported the Bootstrap Confidence Intervals of this (and other) parameter(s), but the number of bootstraps is 5,000 by default. To change this and run 10,000 bootstrap draws write:

process (data=d,y="science",x="lowinc",m="read",total=1,**boot=10000**,

model=4,seed=83582)

The output of this command includes this information on the indirect effects:

Indirect effect(s) of X on Y:

Effect BootSE BootLLCI BootULCI

read -3.0352 0.9298 -4.8551 -1.1391

Since the 95% Confidence Intervals (CI) of this effect do not cross 0, we can reject the null hypothesis and accept that the indirect effect is a substantial one. Note that PROCESS will estimate the 95% CI by default, but narrower or larger intervals can be selected, using the “conf” option. For example, “conf=90*”* would request 90% CIs.

It is also possible to use a different method to estimate CI of the parameter, notably Montecarlo Methods, which can be requested using the option “mc”.

The indirect effect coefficient is interpreted as indicating that participants from lower SES backgrounds are expected to show a reduction in Science scores equivalent to approximately 3 units as a result of the effect of Low SES on Reading scores, which in turn affect Science scores.

If you wanted to report this effect in a standardised metric, you could add the option “stand=1” to the PROCESS command. In this case, since the predictor is dichotomous, PROCESS will report the “partially” standardised effect, which means that the predictor is kept in its dichotomous form, and only the results of the outcome Y are standardised. The output of PROCESS with this option will indicate the indirect effect is equivalent to approximately a 0.31 standardised units reduction in Science scores for low SES participants as an effect of the low SES on Reading scores and, in turn, of Reading scores on Science scores.

1. Use the model in exercise 2, but control for the effects of gender (female) and school type.

The PROCESS macro easily allows to include covariates of different nature in the model. In this case the two covariates are dichotomous. Variable “female” is a dummy variable with 1 indicating the participant is identified as female, and 0 if the participant is identified as male (the data considered only two options). Variable “schtype” takes value 1 if the participant attends a public school, whereas take value 2 if the participant attends a private school. For ease of interpretation, it is better to recode this variable as a dummy variable. I created a new variable “private” that indicated if the participant attended a private school:

d$private <-ifelse(d$schtyp==2, 1, 0)

To control for covariates in the PROCESS macro, just add option “cov” and list the names of the variables to be included as covariates:

process (data=d,y="science",x="lowinc",m="read", **cov=c("female", "private")**, total=1, boot=10000, stand=1, model=4, seed=83582)

This option means that the estimation of the model is run while also controlling for the regression of the mediator on the covariates, and the regression of the outcome on the covariates. Note that in this model the covariates are not assumed to influence the predictor.

1. Compare the results of the indirect effect in the exercise 3 model and the model in exercise 2.

The results concerning the indirect effect in exercise 3 are as follows:

Indirect effect(s) of X on Y:

Effect BootSE BootLLCI BootULCI

read -2.8498 0.9446 -4.7109 -0.9700

Partially standardized indirect effect(s) of X on Y:

Effect BootSE BootLLCI BootULCI

read -0.2878 0.0933 -0.4677 -0.1007

The results show that there is relatively little change in the estimated coefficient of the indirect effect Low SES 🡪Reading🡪Science. The raw coefficient was -3.03 in the unadjusted model, and it is -2.85 in this adjusted model. The standardised coefficients is reduced from -0.31 in the unadjusted model to -0.29 in the adjusted one.

1. Test a serial mediation model where the effect of Low SES on Science scores is mediated by Low SES on Reading scores, Low SES on Math scores, as well as the path Low SES🡪Reading🡪Math🡪Science.

To run this model we need to add a further mediator in the “m” option. Furthermore, we need to specify this is a serial multiple mediation model (see the recorded presentations, or the references provided with them). A similar model is specified by instructing PROCESS to use model template n.6, i.e. “model=6”. Note that in this instance the order in which the covariates are listed specifies the order of influence. If we want Reading scores to influence Math scores, we will have to list the covariates in this order. Finally, the option “contrast” also allows to request tests of the null hypotheses of no significant differences between the specific indirect effects from predictor to outcome through the different mediators.

The command should thus be:

process (data=d,y="science",x="lowinc",**m=c("read","math")**,cov=c("female", "private"), total=1, boot=10000, stand=1, **model=6**, **contrast=1**, seed=83582)

In the following table I have mapped the indirect effects onto their estimated coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coeff | 95% CI LB | 95% CI UB | Std Coeff. |
| Ind1: lowinc->read->science | -1.71 | -3.11 | -0.53 | -0.17 |
| Ind2: lowinc->math->science | -0.56 | -1.66 | 0.43 | -0.06 |
| Ind3: lowinc->read->math->science | -1.14 | -2.18 | -0.33 | -0.11 |

The cumulative indirect effects appear significant (slightly confusingly, this is labelled as the “total” indirect effect in the output): the cumulative indirect effect is estimated as being -3.41 (95% CI -5.46 to -1.30). The result indicate significant indirect effects through Reading scores, as well as through reading and then, in turn, thorough math scores. However, the indirect path through Math is not statistically significant (the 95% CI of this parameter cross zero, thus we cannot reject the null hypothesis that this coefficient is equal to zero).

However, note that the tests of differences between these coefficients are all not significant: these are dubbed (C1) , (C2), and (C3) in the output, and inspection of the output shows they all have 95% CI intervals that cross zero.

**A final note of caution**: I have developed these exercises to provide worked examples of the analyses and the outputs. The models described make assumptions regarding the causal links between variables. Although the results of the analyses are consistent with these assumptions, statistical models alone cannot provide proof of the causal claims made in the models. In the presentation I highlighted that mediation model make assumptions about processes of influence from one variable to another that necessitate time to unfold. Mediation models thus entail longitudinal data. Indeed, the estimation of mediation effects based on cross-sectional data may be justified in some stringent conditions: when these are not met, the estimated mediation effects can be biased and misleading (see Cole & Maxwell, 2003). Furthermore, even if longitudinal data can show that some processes preceded others in time, justifying claims about causal links between variables requires further arguments.

**Reference:**

Cole, D.A, & Maxwell, S.E. (2003). Testing mediational models with longitudinal data: Question and tis in the use of Structural Equation Modeling. *Journal of Abnormal Psychology, 112(4)*, 558-577. https://doi.org/10.1037/0021-843X.112.4.558