Lab 9: Mediation Analysis

SDS358: Applied Regression Analysis

Michael J. Mahometa, Ph.D.

.

> "The great masters do not take any Model quite so seriously as the rest of us. They know that it is, after all, only a model, possibly replaceable."
>
> *C.S. Lewis (The Discarded Image)*

Introduction

The basic idea of Lab is as follows: Answer a research question with the provided dataset. Each week, that research question (and data) will change depending on the topic we've covered the prior class days. Once we're done with Lab, you'll have a Lab Assignment, that will look a lot like the Lab: a research question you'll need to answer given some data. In Lab, you'll learn the procedure for answering the research question. For the Lab Assignment, you'll do that procedure for a grade (independently).

To help answer the research question, we'll follow some basic steps that we'll repeat throughout the semester:

- Reflect on the Question: Figure out the variables of interest, and the technique that's required.
- Analyze the Data: Perform the steps required for the technique.
- Draw Conclusions: Use the information that you got from the prior step to answer the research question in a concise, logical manner.

Let's get started:

Primary Research Question:

Amongst High School students, while controlling for social studies score, is there a signficant indirect effect of reading score on the relationship of math score to science knowledge? How does the effect of math score on science knowlege change with the mediator of reading score?

Step1: Reflect on the Question:

Download the syntax and data files from Canvas.

Let's load in our SDSRegressionR package so that we can use some of it's functions later:

Remember we need the *newest* version of the package for this lab.

```
#Load our class package
library(SDSRegressionR)
```

Next, we'll load in the data. Be sure to use the basic file structure we talked about the first Lab: Put your syntax in a folder specific to this Lab. Then, make a "data" folder in that same place - use lowercase. If you do that, then all of this syntax will work like a charm.

```
hs <- read_csv("data/hsKnowledge.csv")
```

Check the Data:

To make sure that we're working with the right data, and that we're all looking at it the same way, we'll answer some basic questions about the data before moving on:

1. How many observations are in the dataset for the model?

2. Of the frist 10 observations, how many were Female?
3. Of the first 10 students, how many had a science score above 50?

These questions can be answered simply by looking at the dataset once it's loaded in:

```
View(hs)
```

Check the Variables of Interest

Let's find the variables that we need to answer the primary research question:

1. Which variable tells us the science score of the student?
    - What type of variable is this?

2. What is the "variable(s) of interest" for this model?
    - What is the mediator for this model?

3. You'll need two models here (at least) for the mediation, what are they?

Again, these can be answered by looking at the dataframe, and with the help of the *names()* function. Also, the codebook for the data frame is our friend. You can open this in R or Excel. Remember, R is case-sensitive.

```
names(hs)
```

```
##  [1] "id"      "female"  "ses"     "schtyp"  "prog"    "read"    "write"
##  [8] "math"    "science" "socst"   "honors"  "awards"  "cid"
```

Reflect on the Method

The last part of Reflect on the Question asks about the method or technique we'll use.

1. We will use Mediation in a Regression Model to answer this regression. Why?

2. We'll need to run at least two regression models. Why?

3. We might need to use two techniques to discover "significance."" Why?

4. Bootstrapping will help us to find the confidence interval of the indirect effect. How?

Step2: Analyze the Data

In this step, we'll run the provided syntax and answer some questions about the output to help us prepare for the final step.

Here's the syntax you'll need (from the .R syntax file):

```
#### Here is the R script you will use:  (remember that # indicates a comment) ####
#Lab9: Mediation Models

library(SDSRegressionR)

#Load Data:
hs <- read_csv("data/hsKnowledge.csv")
names(hs)

#Full model to look for "baddness"
full <- lm(science ~ math + read + socst, data=hs)
residFitted(full)
cooksPlot(full, key.variable="id", print.obs=TRUE, sort.obs=TRUE)
threeOuts(full, key.variable="id")

#Take out the Cook's D folks:
g_hs <- hs %>%
  filter(id %not in% c(150, 167))

#Run the FULL model first:
#Path B and Cprime
cprime <- lm(science ~ math + read + socst, data=g_hs)
summary(cprime)
lmBeta(cprime)

#Keep ALL the model data...
all_data <- modelData(cprime)

#Total effect model (no mediator(s))
pathc <- lm(science ~ math + socst, data=all_data)
summary(pathc)
lmBeta(pathc)

#Path A
patha <- lm(read ~ math + socst, data=all_data)
summary(patha)
lmBeta(patha)

#Indirect Effect (Multiply paths a and b - your choice how)
ind <- summary(patha)$coef["math", "Estimate"] *
  summary(cprime)$coef["read", "Estimate"]

#Sobel test
se <- sqrt((summary(cprime)$coef["read", "Estimate"]^2 *
              summary(patha)$coef["math", "Std. Error"]^2) +
              (summary(patha)$coef["math", "Estimate"]^2 *
                 summary(cprime)$coef["read", "Std. Error"]^2))
```

```
z <- ind/se
z
(1 - pnorm(z)) * 2

#In practice: use a package to get the job done (with Bootstrapped CIs)
library(psych)
#NOTICE: use of our "good dataset"
med_science <- psych::mediate(science ~ math + (read) - socst, data=g_hs, n.iter = 1000)
med_science
print(med_science, short=FALSE)
```

All quesitons are in reference to models run *after* outlier removal.

Question 1

What is the Total Effect for the Moderation Model?

Question 2

What is the value of path a in the Moderation Model?

Question 3

What is the value of path b in the Moderation Model?

Question 4

The Direct Effect is _____, while the indirect effect is _____.

Question 5

The Sobel test shows a _____ Indirect Effect, z = _____, p = _____.

Question 6

A Bootstrapped Indirect Effects showed a 95% Confidence Interval between _____ and _____.

Quesiton 7

Draw out the final model representing the Mediation:

Step3: Draw Conclusions

The final step is for us to Draw Conclusions. We'll take the syntax we've been given from Analyze the Question, run it, then examine the output. The questions from the prior step help set us up for the Draw Conclusions part.

We'll "fill in the blanks" in a canned paragraph for the Lab. For the Lab Assignment, you'll need to come up with a similar paragraph all on your own (please don't steal mine).

Primary Research Question:
Amongst High School students, while controlling for social studies score, is there a signficant indirect effect of reading score on the relationship of math score to science knowledge? How does the effect of math score on science knowlege change with the mediator of reading score?

> Our models tested the meditation of Reading score on the effect of Math score on Science score in a sample of 300 high school student. The Total Effect of Math score on Science score was $b =$ _____ (Standardized Beta: _____, t(_____) = _____, p < 0.05), was reduced to $b =$ _____ (Standardized Beta: _____, t(_____) = _____, p < 0.05) with the inclusion of Reding score into the model. We found a significant Indirect Effect of for Reading score of $b =$ _____, (Sobel z-score _____, p = _____).
>
> Using a standard Bootstrapping procedure of 1000 samples, we found the 95% Confidence Interval of the Indirect Effect to be between _____ and _____, also indicating significance of the Indirect Effect. The effect of Math score on Science score was _____ mediated, as the Direct Effect was still _____. Finally, the Indirect Effect could account for _____% of the Total Effect.

Lab Assignment

Now, with the tools at your disposal (the R syntax from Lab, and the logic of proceeding through the three steps of answering the research question), you'll have a Lab Assignment to complete (independently). For now, the Lab Assignment is to be completed in Canvas. It will follow the basic structure, and lead to the same place - answering the research question with a concise paragraph as in Draw Conclusions.

Good Luck!