HW5

Carina\_Sun

2024-04-17

### Code

knitr::opts\_chunk$set(echo = TRUE)  
library(mediation)

## Loading required package: MASS

## Loading required package: Matrix

## Loading required package: mvtnorm

## Loading required package: sandwich

## mediation: Causal Mediation Analysis  
## Version: 4.5.0

library(psych)

##   
## Attaching package: 'psych'

## The following object is masked from 'package:mediation':  
##   
## mediate

library(nnet)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ ggplot2::%+%() masks psych::%+%()  
## ✖ ggplot2::alpha() masks psych::alpha()  
## ✖ tidyr::expand() masks Matrix::expand()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ tidyr::pack() masks Matrix::pack()  
## ✖ dplyr::select() masks MASS::select()  
## ✖ tidyr::unpack() masks Matrix::unpack()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tinytex)  
library(rmarkdown)  
library(skimr)  
library(diagram)

## Loading required package: shape

meddata <-read.csv("meddata2.csv", stringsAsFactors = T)  
  
#1.Check the structure of the data using str  
str(meddata)

## 'data.frame': 100 obs. of 4 variables:  
## $ grades : int 6 7 7 8 4 4 9 5 8 4 ...  
## $ self.esteem: int 5 5 7 4 3 4 7 0 7 3 ...  
## $ happiness : int 6 5 4 8 5 7 8 4 7 4 ...  
## $ gender : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...

#2. Get descriptive statistics   
describe(meddata)

## vars n mean sd median trimmed mad min max range skew kurtosis  
## grades 1 100 5.99 1.74 6.0 5.99 1.48 0 10 10 -0.25 0.24  
## self.esteem 2 100 4.86 1.90 5.0 4.84 1.48 0 10 10 0.11 -0.38  
## happiness 3 100 5.23 2.04 5.0 5.20 1.48 0 10 10 0.12 -0.09  
## gender\* 4 100 1.50 0.50 1.5 1.50 0.74 1 2 1 0.00 -2.02  
## se  
## grades 0.17  
## self.esteem 0.19  
## happiness 0.20  
## gender\* 0.05

summary(meddata)

## grades self.esteem happiness gender  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 F:50   
## 1st Qu.: 5.00 1st Qu.: 3.00 1st Qu.: 4.00 M:50   
## Median : 6.00 Median : 5.00 Median : 5.00   
## Mean : 5.99 Mean : 4.86 Mean : 5.23   
## 3rd Qu.: 7.00 3rd Qu.: 6.00 3rd Qu.: 7.00   
## Max. :10.00 Max. :10.00 Max. :10.00

#On the math assessment, the 50 female and 50 male students had a mean score of 5.99 (SD = 1.74). Scores of 5.000, 6.000, and 7.000 represented the 25th, 50th, and 75th percentiles respectively.  
#For their self esteem estimate, they had a mean score of 4.86 (SD = 1.90). Scores of 3, 5, and 6 represented the 25th, 50th, and 75th percentiles respectively.  
#For their happiness level, they had a mean score of 5.23 (SD = 2.04). Scores of 4, 5, and 7 represented the 25th, 50th, and 75th percentiles respectively.  
  
# 3.four steps for mediation analysis  
#a. Step 1: Make sure you assign the model for this to a name called Totaleffect.   
#i. Is this step necessary? If so, explain why, if not explain why.  
Totaleffect = lm(happiness~grades,meddata)  
summary(Totaleffect)

##   
## Call:  
## lm(formula = happiness ~ grades, data = meddata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0262 -1.2340 -0.3282 1.5583 5.1622   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.8572 0.6932 4.122 7.88e-05 \*\*\*  
## grades 0.3961 0.1112 3.564 0.000567 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.929 on 98 degrees of freedom  
## Multiple R-squared: 0.1147, Adjusted R-squared: 0.1057   
## F-statistic: 12.7 on 1 and 98 DF, p-value: 0.0005671

# This Step is NOT necessary, total effect equals to direct effect + indirect effect, which can be calculated. Because the mediation effects, which we are interested here, can be present in the absence of a total effect.  
# Interpretation: The total effect (p<0.001) describes that there is a total effect that happiness has on grades, regardless of the mediation.  
  
# b. Step 2  
fit.mediator=lm(self.esteem~grades,meddata)  
summary(fit.mediator)

##   
## Call:  
## lm(formula = self.esteem ~ grades, data = meddata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.3046 -0.8656 0.1344 1.1344 4.6954   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.49952 0.58920 2.545 0.0125 \*   
## grades 0.56102 0.09448 5.938 4.39e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.639 on 98 degrees of freedom  
## Multiple R-squared: 0.2646, Adjusted R-squared: 0.2571   
## F-statistic: 35.26 on 1 and 98 DF, p-value: 4.391e-08

# Interpretation: Grades have a significant effect on students' self esteem (p<0.001).  
  
#c. Step 3 and Step 4   
fit.dv=lm(happiness~grades+self.esteem,meddata)  
summary(fit.dv)

##   
## Call:  
## lm(formula = happiness ~ grades + self.esteem, data = meddata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7631 -1.2393 0.0308 1.0832 4.0055   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.9043 0.6055 3.145 0.0022 \*\*   
## grades 0.0396 0.1096 0.361 0.7187   
## self.esteem 0.6355 0.1005 6.321 7.92e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.631 on 97 degrees of freedom  
## Multiple R-squared: 0.373, Adjusted R-squared: 0.3601   
## F-statistic: 28.85 on 2 and 97 DF, p-value: 1.471e-10

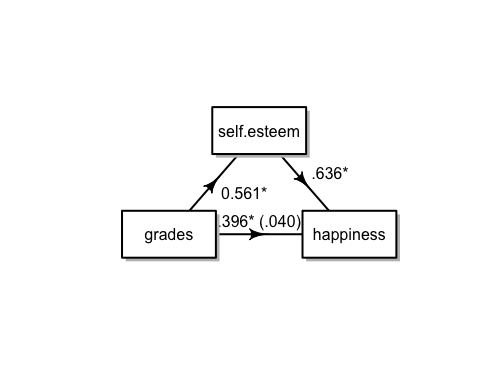
options(scipen = 999)  
# Interpretation: The effect of grades is no longer significant while the mediator self esteem has a significant effect on happiness. Put the mediation in place, the total effect of grades on happiness is explained by the mediator - self esteem.   
  
#4. Run the mediation using the mediation  
results = mediation::mediate(fit.mediator,fit.dv,boot = T, treat = "grades", mediator = "self.esteem" )

## Running nonparametric bootstrap

summary(results)

##   
## Causal Mediation Analysis   
##   
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method  
##   
## Estimate 95% CI Lower 95% CI Upper p-value   
## ACME 0.3565 0.2247 0.54 <0.0000000000000002 \*\*\*  
## ADE 0.0396 -0.2065 0.30 0.752   
## Total Effect 0.3961 0.1540 0.66 0.002 \*\*   
## Prop. Mediated 0.9000 0.4970 2.06 0.002 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Sample Size Used: 100   
##   
##   
## Simulations: 1000

# ACME: The average causal mediation effect, which is path a x path b (step 2 times step 3), is the indirect effect. Here, (0.63 \*0.56) = 0.3565  
# ADE: Average Direct Effect says the direct effect of the IV on the DV, the path c' in step 3+4, which is 0.0396.  
# Total effect is ACME+ADE, which is also step 1.  
  
  
#5. Using the diagram   
data <- c(0, "'0.561\*'", 0,  
 0, 0, 0,  
 "'.636\*'", "'.396\* (.040)'", 0)  
M<- matrix (nrow=3, ncol=3, byrow = TRUE, data=data)  
plot<- plotmat (M, pos=c(1,2),  
 name= c( "self.esteem","grades", "happiness"),  
 box.type = "rect", box.size = 0.13, box.prop=0.5, curve=0)



### Final Conclusion

* The overall model was statistically significant, F(2, 97) = 28.85, p < .001\*\*\*, with an adjusted R-squared of .3601, indicating that approximately 36% of the variance in happiness can be explained by the combined effects of grades and self-esteem. Analyzing the indirect effects, results reveal that self esteem significantly mediates the relationship between grades and happiness, ab=.3565, p<2e-16, and the 95% confidence interval ranged from .2147 to .52., as the direct effect of grades on happiness becomes not significant when self-esteem is included in the model. s
* Bootstrapping: We tested the significance of self esteem (indirect effect) using bootstrapping procedures. Unstandardized mediation indirect effects were computed for each of 100 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized mediation (indirect) effect was .3565, and the 95% confidence interval ranged from .2147 to .52. Thus, the indirect effect was statistically significant, p<2e-16 \*\*\*.
* Mediator: There was a significant effect of the mediator of grades on happiness through self esteem , ab = 0.3565, CI [0.2147, 0.52]. The mediator could account for more than half of the total effect, = .636. Percent mediation is 0.90. or about a 90 % increase over the total effect.