Course: Multivariate statistics (AUT23)

Chapter 7: Moderation analysis

In this exercise, we will use the data "protest.sav" (Hayes, 2022) which can be downloaded here under "data files and code". Especially, we will focus on the following variables:

- Protest (independent variable): A lawyer protests against gender discrimination (experimental group, dichotomous 0 = no and 1 = yes)
- Like (dependent variable): assessment of the lawyer (scale 1-7)
- Sexism (moderator): perception of sexism as a ubiquitous problem in society (scale 1-7)

7.16.1 Exercise 1: protest with a continuous moderator

We want to test the assumption that when women believe that sexism is a problem in society, they like the lawyer more when he protests sexism than when he doesn't protest.

Start by drawing the regression equations.

- Solution: equation

Now, we want to know if the overall model is significant? Start by importing the data:

- Show the code

Note that there are several ways to center the variables when creating the interaction term.

- Show the code

Now, run the regression model:

- Show the code

How much variance does the model explain? Are there main effects or conditional effects? If yes, what do they look like?

- Solution: interpretation

Is there a moderation effect?

- Show the code

If so, how much variance does this explain and what does this effect mean in general?

- Solution: interpretation

Illustrate the moderation effect graphically and interpret it. First, we can create an interaction plot:

- Show the code

Second, we can provide a Johnson-Neyman plot:

- Show the code

7.16.2 Exercise 2: protest with a dichotomous moderator

Now, divide the moderator into a dichotomous variable (sexism low vs. high) with a median split and recalculate the moderation analysis.

- Show the code

What changes in the output?

- Show the code
- Solution: interpretation

Calculate the moderation analysis again with the variable «x» as the independent variable (it measure the lawyer protests to varying degrees on a scale of 1-7) and the metric moderator. What changes in the output?

- Show the code
- Solution: interpretation

What changes in the graphics?

- Show the code
- Solution: interpretation

Chapter 7: Moderation analysis (answers)

In this exercise, we will use the data "protest.sav" (Hayes, 2022) which can be downloaded here under "data files and code". Especially, we will focus on the following variables:

- Protest (independent variable): A lawyer protests against gender discrimination (experimental group, dichotomous 0 = no and 1 = yes)
- Like (dependent variable): assessment of the lawyer (scale 1-7)
- Sexism (moderator): perception of sexism as a ubiquitous problem in society (scale 1-7)

7.16.1 Exercise 1: protest with a continuous moderator

We want to test the assumption that when women believe that sexism is a problem in society, they like the lawyer more when he protests sexism than when he doesn't protest.

Start by drawing the regression equations.

➢ Solution: equation

The regression equation go as:

$$Y_i = \beta_0 + \beta_1 * Protest_i + \beta_2 * Sexismus_i + \beta_3 * (Protest * Sexismus_i) + \epsilon$$

Now, we want to know if the overall model is significant? Start by importing the data:

➢ Show the code

```
# load the data
library(foreign)
db <- read.spss(file=paste0(getwd(),
                "/data/protest.sav"),
                use.value.labels = F,
                to.data.frame = T)
```

Note that there are several ways to center the variables when creating the interaction term.

➢ Show the code

```
# interaction term
# without centering
db$ProtestXSexism1 = db$protest*db$sexism
# with centering
db$ProtestXSexism2 = (db$protest-mean(db$protest)) * (db$sexism-mean(db$sexism))
# z-standardization
```

```
# db$ProtestXSexism3 = scale(db$protest)*scale(db$sexism)
# view
head(db)
##   sexism liking respappr protest x Sexism_h_t ProtestXSexism1 ProtestXSexism2
## 1   4.25   4.50    5.75       0 4         1              0       0.5914260
## 2   4.62   6.83    5.75       0 6         1              0       0.3390229
## 3   4.62   4.83    5.25       0 4         1              0       0.3390229
## 4   4.37   4.83    4.25       0 5         1              0       0.5095655
## 5   4.25   5.50    2.50       0 3         1              0       0.5914260
## 6   4.00   6.83    4.75       0 3         1              0       0.7619686
```

Now, run the regression model:

- Show the code

```
# regression model (with centering)
m.cent = lm(liking ~ protest + sexism + ProtestXSexism2, data=db)
summary(m.cent)
##
## Call:
## lm(formula = liking ~ protest + sexism + ProtestXSexism2, data = db)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9894 -0.6381  0.0478  0.7404  2.3650
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.79659    0.58679   8.174 2.83e-13 ***
## protest         0.49262    0.18722   2.631  0.00958 **
## sexism          0.09613    0.11169   0.861  0.39102
## ProtestXSexism2  0.83355    0.24356   3.422  0.00084 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9888 on 125 degrees of freedom
## Multiple R-squared:  0.1335, Adjusted R-squared:  0.1127
## F-statistic: 6.419 on 3 and 125 DF,  p-value: 0.0004439

# regression model (without centering)
m.roh = lm(liking ~ protest + sexism + ProtestXSexism1, data=db)
summary(m.roh)
##
## Call:
## lm(formula = liking ~ protest + sexism + ProtestXSexism1, data = db)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9894 -0.6381  0.0478  0.7404  2.3650
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.7062     1.0449   7.375 1.99e-11 ***
## protest         -3.7727     1.2541  -3.008  0.00318 **
## sexism          -0.4725     0.2038  -2.318  0.02205 *
## ProtestXSexism1   0.8336     0.2436   3.422  0.00084 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9888 on 125 degrees of freedom
## Multiple R-squared:  0.1335, Adjusted R-squared:  0.1127
```

```
## F-statistic: 6.419 on 3 and 125 DF,  p-value: 0.0004439
```

How much variance does the model explain? Are there main effects or conditional effects? If yes, what do they look like?

- Solution: interpretation

> The overall model is significant ($p < .001$) and explains 13.3% of the variance. We are allowed to interpret the results of the regression.
>
> There is a significant conditional effect of protest when sexism = 0. It is not a main effect since interaction is also included.
>
> Protest has an effect on the assessment if the moderator (sexism) has the value zero (= a medium level, since mean centered)

Is there a moderation effect?

- Show the code

```
# run the model without the interaction term
m0 = lm(liking ~ protest + sexism, data=db)
# compare the R2
summary(m.cent)$r.squared - summary(m0)$r.squared
## [1] 0.08119242
# get EtaSq
DescTools::EtaSq(m.cent)
##                  eta.sq eta.sq.part
## protest       0.047991546 0.052478399
## sexism        0.005136019 0.005892326
## ProtestXSexism2 0.081192415 0.085672955
```

If so, how much variance does this explain and what does this effect mean in general?

- Solution: interpretation

The interaction of protest and sexism perception is significant. There is a moderation effect.

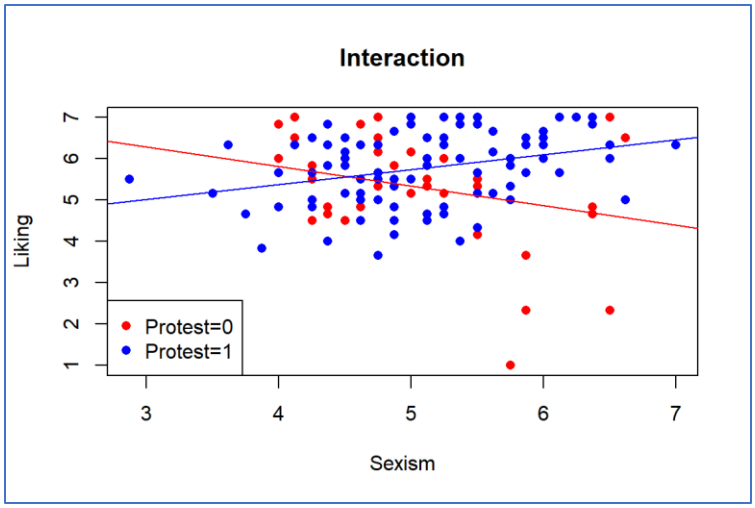
The effect of the protest on the lawyer's assessment varies depending on how strongly the subjects perceive sexism as a problem.

The interaction contributes 8.1% to explaining the variance. The dependent variable is thus better explained if moderation is taken into account.

Illustrate the moderation effect graphically and interpret it. First, we can create an interaction plot:

➢ Show the code

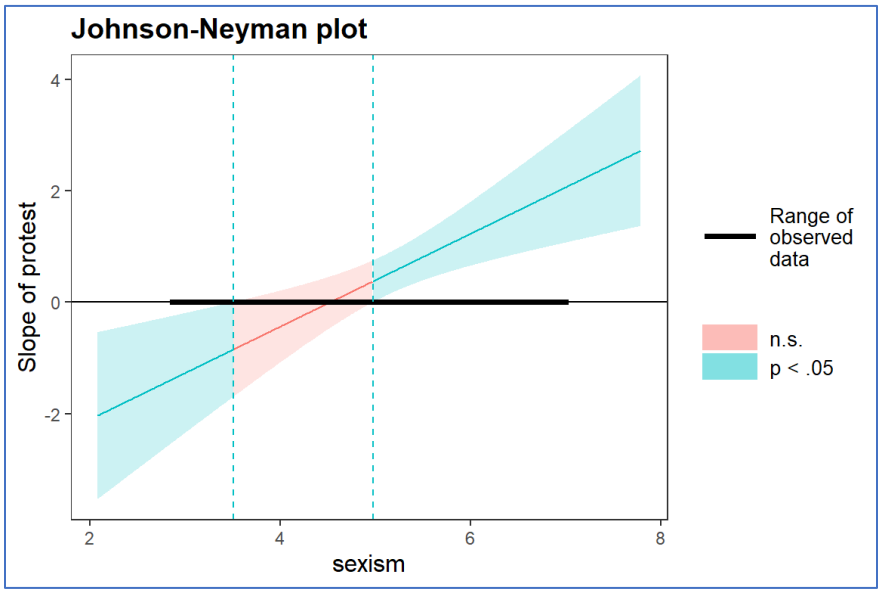
```
# extract the needed coefficients
intercept.p0 = coefficients(m.roh)[1]
intercept.p1 = coefficients(m.roh)[1] + coefficients(m.roh)[2]
slope.p0 = coefficients(m.roh)[3]
slope.p1 = coefficients(m.roh)[3] + coefficients(m.roh)[4]
# interaction plot
par(mfrow=c(1,1))
farben = c("red","blue")
plot(db$sexism, db$liking, main="Interaction",
    col=farben[db$protest+1],pch=16,
    xlab="Sexism",ylab="Liking")
abline(intercept.p0,slope.p0,col="red")
abline(intercept.p1,slope.p1,col="blue")
legend("bottomleft",
      c("Protest=0","Protest=1"),
      col=c("red","blue"),pch=16)
```



Second, we can provide a Johnson-Neyman plot:

➢ Show the code

```
library(interactions)
m.simplified = lm(liking ~ protest*sexism, data=db)
johnson_neyman(m.simplified,"protest","sexism")
## JOHNSON-NEYMAN INTERVAL
##
## When sexism is OUTSIDE the interval [3.51, 4.98], the slope of protest is p < .05.
##
## Note: The range of observed values of sexism is [2.87, 7.00]
```



7.16.2 Exercise 2: protest with a dichotomous moderator

Now, divide the moderator into a dichotomous variable (sexism low vs. high) with a median split and recalculate the moderation analysis.

➢ Show the code

```
# Median split

db$sexism.ms = as.integer(db$sexism>=median(db$sexism))
```

What changes in the output?

➢ Show the code

```
# new model

m3 = lm(liking ~ protest*sexism.ms, data=db)

summary(m3)

##

## Call:

## lm(formula = liking ~ protest * sexism.ms, data = db)

##

## Residuals:

##     Min      1Q  Median      3Q     Max

## -3.9179 -0.6815  0.1263  0.7963  2.0821

##

## Coefficients:

##                Estimate Std. Error t value Pr(>|t|)

## (Intercept)       5.6491     0.2125  26.584  < 2e-16 ***

## protest          -0.1154     0.2634  -0.438  0.66199

## sexism.ms         -0.7312     0.3122  -2.342  0.02074 *

## protest:sexism.ms   1.2090     0.3779   3.199  0.00175 **

## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.9967 on 125 degrees of freedom
```

```
## Multiple R-squared:  0.1195, Adjusted R-squared:  0.09839
## F-statistic: 5.656 on 3 and 125 DF,  p-value: 0.001148
# get EtaSq
DescTools::EtaSq(m3)
##                     eta.sq eta.sq.part
## protest          0.043987025 0.047581194
## sexism.ms        0.002000014 0.002266368
## protest:sexism.ms 0.072096985 0.075686621
# get the coeff
meanvalues = tapply(db$liking, list(db$protest,db$sexism.ms),FUN=mean)
meanvalues
##        0       1
## 0 5.649091 4.917895
## 1 5.533659 6.011489
```

- Solution: interpretation

R-square significantly worse (probably due to loss of information during dichotomization) and the coefficients change slightly.

There are only two Simple Slopes because there are only two moderator levels. There are no Johnson-Neyman values.

Calculate the moderation analysis again with the variable «x» as the independent variable (it measure the lawyer protests to varying degrees on a scale of 1-7) and the metric moderator. What changes in the output?

- Show the code

```
# new model
m4 = lm(liking ~ x*sexism, data=db)
summary(m4)
##
## Call:
## lm(formula = liking ~ x * sexism, data = db)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0451 -0.6128  0.1029  0.7720  1.6583
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  8.65856    1.90366   4.548 0.0000126 ***
## x           -0.90210    0.45129  -1.999    0.0478 *
## sexism      -0.73604    0.36256  -2.030    0.0445 *
## x:sexism     0.21069    0.08563   2.460    0.0152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.009 on 125 degrees of freedom
## Multiple R-squared:  0.09858,    Adjusted R-squared:  0.07695
## F-statistic: 4.557 on 3 and 125 DF,  p-value: 0.004584
# get EtaSq
DescTools::EtaSq(m4)
##               eta.sq eta.sq.part
## x        0.046574596  0.04912969
## sexism   0.006844541  0.00753586
## x:sexism 0.043654300  0.04619148
```

- Solution: interpretation

Now the hypothesis is that the more women believe that sexism is a problem, the more they like the lawyer, the more she protests.

The coefficients change, the explanation of variance overall and through the interaction alone is weaker in each case (but this is simply because it is a completely different and only a simulated variable).

What changes in the graphics?

- Show the code

```
# plots

johnson_neyman(m4,"x","sexism")

## JOHNSON-NEYMAN INTERVAL

##

## When sexism is OUTSIDE the interval [0.20, 5.01], the slope of x is p < .05.

##

## Note: The range of observed values of sexism is [2.87, 7.00]
```

- Solution: interpretation