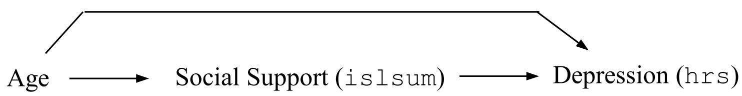
Testing Mediation with Regression Analysis Examples

SPSS

Below I use an SPSS macro developed Andrew Hayes (see Hayes & Rockwood, 2017) to test an indirect effect using the Yale social support data. The Process macro can be downloaded from: https://haskayne.ucalgary.ca/CCRAM/resource-hub. In this example, age (`age`) is the predictor, social support (`islsum`) is the mediator, and depression (`hrs`) is the final outcome. Here is a picture of the model:



After downloading the macro, save it in a known location that you can specify the exact path for. Create a new syntax file and either open your data set or add a `get file='`*`your data set location and file name here`*`'` command to the beginning of the syntax file to specify the location of your data file.[1] Then add the following commands (this was tested with Process version 4.3 syntax), replacing your variable names for my variable names for X (initial predictor), Y (final outcome), and M (mediator) to the syntax file:

```
cd "c:\jason\temp".

insert file='C:\Jason\SPSSWIN\macros\process.sps'.
execute.
process y = hrs
 / x = age
 / m = islsum
 /total=1
 /boot=10000
 /seed=10000
 /model=4
 /stand=1.
execute.
```

Make sure that the `insert file` command points to the exact location of the process macro where you saved it. Then, highlight the entire syntax in the syntax window, and run.

Output

The first section of the output (marked by lines of asterisks) gives each of the direct regression coefficients depicted in the diagram above and will be the same as those you would obtain with the usual regression command in SPSS. The bootstrap tests of the indirect effect are found in the final section under the heading "`TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y`" and then under the subheading "`Indirect effect(s) of X on Y:`", where `Effect` gives the average estimate for indirect effect from the bootstrap samples, `BootSE` gives the standard error estimate, and `BootLLCI` and `BootULCI` are 95% confidence limits. If the 95% confidence limits include zero, the indirect effect test is not significant.[2]

```
Run MATRIX procedure:

**************** PROCESS Procedure for SPSS Version 4.3.1 ****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2022). www.guilford.com/p/hayes3

**************************************************************************
Model  : 4
    Y  : hrs
    X  : age
    M  : islsum

Sample
```

[1] Mac locations have no drive letter and forward slashes, '/Users/your subfolders'
[2] Note that these limits are "percentile" limits which do not involve a bias correction ("accelerated confidence limits"), as the bias-corrected limits may have slightly elevated Type I error rates (Fritz, Taylor, & MacKinnon, 2012; Hayes & Scharkow, 2013).

```
Size:  301

Custom
Seed:     10000

**************************************************************************
OUTCOME VARIABLE:
 islsum

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .0277      .0008     5.2515      .2304     1.0000   299.0000      .6316

Model
              coeff         se          t          p       LLCI       ULCI
constant    12.6950      .6110    20.7788      .0000    11.4927    13.8973
age          -.0046      .0096     -.4800      .6316     -.0236      .0143

Standardized coefficients
         coeff
age     -.0277

**************************************************************************
OUTCOME VARIABLE:
 hrs

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .3329      .1109    30.5137    18.5762     2.0000   298.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant    20.7583     2.3023     9.0162      .0000    16.2274    25.2892
age          -.0588      .0232    -2.5304      .0119     -.1045     -.0131
islsum       -.7825      .1394    -5.6133      .0000    -1.0569     -.5082

Standardized coefficients
            coeff
age        -.1383
islsum     -.3067

************************** TOTAL EFFECT MODEL ****************************
OUTCOME VARIABLE:
 hrs

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .1298      .0168    33.6273     5.1205     1.0000   299.0000      .0244

Model
              coeff         se          t          p       LLCI       ULCI
constant    10.8243     1.5460     7.0014      .0000     7.7819    13.8668
age          -.0552      .0244    -2.2629      .0244     -.1032     -.0072

Standardized coefficients
         coeff
age     -.1298


************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI       c_cs
     -.0552      .0244    -2.2629      .0244     -.1032     -.0072     -.1298

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI      c'_cs
     -.0588      .0232    -2.5304      .0119     -.1045     -.0131     -.1383

Indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
islsum      .0036      .0081     -.0134      .0196

Completely standardized indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
```

```
islsum      .0085      .0188     -.0307      .0461

********************** ANALYSIS NOTES AND ERRORS ***********************

Level of confidence for all confidence intervals in output:
  95.0000

Number of bootstrap samples for percentile bootstrap confidence intervals:
  10000

------ END MATRIX -----
```

R mediation Package

You will need to install the `mediation` package first, using `install.packages("mediation")`. Also, given that two models are test with different variables, they may result in different sample sizes, so use a listwise deletion function with all of the variables you are using to eliminate missing data on any of the variables used in either of the models. I use the `dplyr` package and `filter` function for this.

```
> cat("\014")  #clear output
> rm(d) #clear active frame from previous analyses
> if(!is.null(dev.list())) dev.off(dev.list()["RStudioGD"])  #clear plots

> library(haven)
> d = read_sav("c:/jason/spsswin/da2/yale.sav")

#listwise deletion necessary to make sample sizes for the two regressions match
> library(dplyr)
> d <-filter(d,age != 'NA' & islsum!='NA' & hrs!='NA')

library(mediation)
#specify two separate models, one predicting the mediator, m, and one predicting the outcome, y
# covariates can be included in either model
mmodel <-lm(islsum ~ age, data = d)
ymodel <- lm(hrs ~ islsum + age, data = d)
summary(mmodel)

Call:
lm(formula = islsum ~ age, data = d)

Residuals:
     Min       1Q   Median       3Q      Max
-11.3203  -1.3943   0.3047   1.8372   3.6242

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
(Intercept) 12.694977   0.610957   20.78 <0.0000000000000002 ***
age         -0.004625   0.009635   -0.48               0.632
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.292 on 299 degrees of freedom
Multiple R-squared:  0.0007701,      Adjusted R-squared:  -0.002572
F-statistic: 0.2304 on 1 and 299 DF,  p-value: 0.6316

summary(ymodel)
Call:
lm(formula = hrs ~ islsum + age, data = d)

Residuals:
    Min      1Q  Median      3Q     Max
-10.799  -4.199  -0.759   3.901  16.592

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept) 20.75830    2.30234   9.016 < 0.0000000000000002 ***
islsum      -0.78251    0.13940  -5.613         0.0000000454 ***
age         -0.05879    0.02323  -2.530               0.0119 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.524 on 298 degrees of freedom
Multiple R-squared:  0.1109, Adjusted R-squared:  0.1049
F-statistic: 18.58 on 2 and 298 DF,  p-value: 0.00000002495

#this statement removes any prior medtest result to avoid problems if there are errors and multiple attempts
rm(medtest)
```

```
#request the boostrap indirect test, default is 1000 samples (usually ok),
#"perc" requests standard percentiles or use "bca" is for bias-corrected version (not recommended)
medtest <- mediate(mmodel, ymodel, treat = "age", mediator = "islsum", boot.ci.type = "perc", data = d)
summary(medtest)
```

The output reports the "Estimate", which is the average indirect coefficient of the bootstrap samples, the 95% confidence intervals, and the p-value for significance test. The indirect coefficient is reported for the row labeled "ACME", which stands for average causal mediation effects, and the direct effect of X predicting Y controlling for the mediator is labeled "ADE", which stands for average direct effect. The total effect is the two of these effects added together. The proportion mediated attempts to capture the portion of the total effect that is due to the mediation effect (details of the computations of this quantity vary).

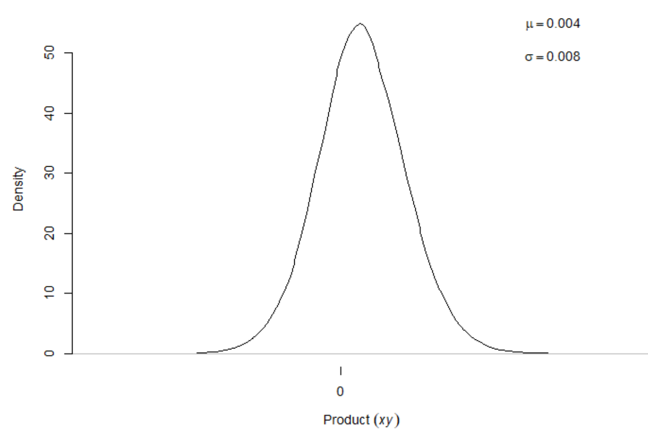
```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

                Estimate 95% CI Lower 95% CI Upper p-value
ACME             0.00376     -0.01067         0.02   0.642
ADE             -0.05894     -0.10446        -0.01   0.010
Total Effect    -0.05518     -0.10348        -0.01   0.024
Prop. Mediated  -0.05913     -0.87580         0.24   0.662

Sample Size Used: 301
Simulations: 1000
```

R RMediation

Alternatively, the `RMediation` package (see Tofighi & MacKinnon, 2011) to estimate the confidence limits with the Monte Carlo method, which performs similarly to the percentile bootstrap method in most cases but may have slightly better power for indirect effects across multiple direct paths when effect sizes are more modest (Tofighi & MacKinnon, 2015). Values from prior regressions (see outputs from the two `lm` models above) need to be input into the `medci` function, where `mu.x` and `mu.y` refer to the a and b paths and `se.x` and `se.y` refer to their respective standard errors.

```
> #tofighi and mackinnon (2011) package has been removed from CRAN
> library(RMediation)
> #plug in results for mu.x (a path) and mu.y (b path) and their SEs from the mediation regressions
> medci(mu.x=-0.004625,mu.y=-0.78251,se.x=0.009635,se.y=0.13940,rho=0,alpha=.05,type="MC")
$`95% CI`
       2.5%       97.5%
-0.01146218  0.01926891

$Estimate
[1] 0.003611272

$SE
[1] 0.007722864

$`MC Error`
[1] 0.00000007722864

> #request a plot of the resampling for the product term using plot=TRUE
> medci(mu.x=-0.004625,mu.y=-0.78251,se.x=0.009635,se.y=0.13940,rho=0,alpha=.05,type="MC", plot=TRUE)
```

```
$`97.5% CI`
       2.5%       97.5%
-0.01129172  0.01922726

$Estimate
[1] 0.003624395

$SE
[1] 0.0076663

$`MC Error`
[1] 0.000000076663
```



R Process Macro

To use the R process macro (Hayes, 2022), follow the same instructions as those described in the moderation example handout "Regression Moderation Examples: SPSS and R" for this class. Download and open the process.R macro script and run it. Then you need these statements for the mediation model test after reading in your data, of course:

```
> #R process macro-first open process.R script and run it
> #seed helps replicate the exact bootsrap results, and can be any number you desire
> #boot=1000 sets number of bootstrap samples (usually 1000 adequate) stand=1 standardized results
> process(data=d,y="hrs",x="age",m="islsum",total=1,boot=1000,seed=10000,model=4,stand=1)


********************* PROCESS for R Version 4.3.1 ********************

           Written by Andrew F. Hayes, Ph.D.  www.afhayes.com
   Documentation available in Hayes (2022). www.guilford.com/p/hayes3

***********************************************************************

Model : 4
    Y : hrs
    X : age
    M : islsum

Sample size: 301

Custom seed: 10000


***********************************************************************
Outcome Variable: islsum

Model Summary:
          R      R-sq       MSE         F       df1       df2         p
     0.0277    0.0008    5.2515    0.2304    1.0000  299.0000    0.6316

Model:
             coeff        se         t         p      LLCI      ULCI
constant   12.6950    0.6110   20.7788    0.0000   11.4927   13.8973
age        -0.0046    0.0096   -0.4800    0.6316   -0.0236    0.0143

Standardized coefficients:
        coeff
age   -0.0276

***********************************************************************
Outcome Variable: hrs

Model Summary:
          R      R-sq       MSE         F       df1       df2         p
     0.3329    0.1109   30.5137   18.5762    2.0000  298.0000    0.0000

Model:
             coeff        se         t         p      LLCI      ULCI
constant   20.7583    2.3023    9.0162    0.0000   16.2274   25.2892
age        -0.0588    0.0232   -2.5304    0.0119   -0.1045   -0.0131
islsum     -0.7825    0.1394   -5.6133    0.0000   -1.0569   -0.5082

Standardized coefficients:
           coeff
age      -0.1383
islsum   -0.3067
************************ TOTAL EFFECT MODEL ***************************
Outcome Variable: hrs

Model Summary:
          R      R-sq       MSE         F       df1       df2         p
     0.1298    0.0168   33.6273    5.1205    1.0000  299.0000    0.0244

Model:
             coeff        se         t         p      LLCI      ULCI
constant   10.8243    1.5460    7.0014    0.0000    7.7819   13.8668
age        -0.0552    0.0244   -2.2629    0.0244   -0.1032   -0.0072

Standardized coefficients:
        coeff
age   -0.1298

***********************************************************************
Bootstrapping progress:
  |>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>| 100%
```

```
************ TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y ************

Total effect of X on Y:
     effect        se         t         p      LLCI      ULCI      c_cs
    -0.0552    0.0244   -2.2629    0.0244   -0.1032   -0.0072   -0.1298

Direct effect of X on Y:
     effect        se         t         p      LLCI      ULCI     c'_cs
    -0.0588    0.0232   -2.5304    0.0119   -0.1045   -0.0131   -0.1383

Indirect effect(s) of X on Y:
          Effect    BootSE  BootLLCI  BootULCI
islsum    0.0036    0.0076   -0.0131    0.0181

Completely standardized indirect effect(s) of X on Y:
          Effect    BootSE  BootLLCI  BootULCI
islsum    0.0085    0.0176   -0.0301    0.0417

******************** ANALYSIS NOTES AND ERRORS ************************

Level of confidence for all confidence intervals in output: 95

Number of bootstraps for percentile bootstrap confidence intervals: 1000
```

References

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate behavioral research, 47*(1), 61-87.

Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour research and therapy*, *98*, 39-57.

Hayes, A.F. (2022) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, Third Edition*. Guilford Press.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter?. *Psychological science, 24*(10), 1918-1927.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods, 7*(4), 422.

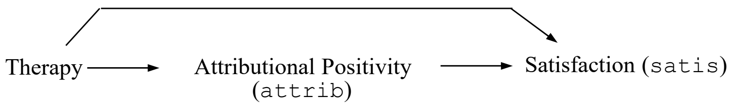
Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. mediation: R Package for Causal Mediation Analysis. https://cran.r-project.org/web/packages/mediation/vignettes/mediation.pdf.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692-700.

Tofighi, D., & MacKinnon, D. P. (2015). Monte Carlo Confidence Intervals for Complex Functions of Indirect Effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-12

Example Mediation Write-up

Sadly, my example above was not significant ☹, so I did another example that would be ☺ for use in an example write-up. The output from the macro is included again here to see where the results came from. The example was an investigation of the hypothesis that therapy affects satisfaction by affecting attributional positivity.



Steps are the same as above, and here are the modified command lines.

```
cd "c:\jason\temp".

insert file='C:\Jason\SPSSWIN\macros\process.sps'.
execute.
process y = satis
 / x = therapy
 / m = attrib
 /total=1
 /boot=10000
 /seed=10000
 /model=4
 /stand=1.
execute.

**************************************************************************
OUTCOME VARIABLE:
 attrib

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4595      .2111      .6676     7.4940     1.0000    28.0000      .0106

Model
             coeff         se          t          p       LLCI       ULCI
constant     -.3536      .2184    -1.6191      .1166     -.8009      .0938
therapy       .8186      .2990     2.7375      .0106      .2060     1.4311

Standardized coefficients
             coeff
therapy      .9056

**************************************************************************
OUTCOME VARIABLE:
 satis

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5566      .3098      .6112     6.0605     2.0000    27.0000      .0067

Model
             coeff         se          t          p       LLCI       ULCI
constant     -.1843      .2185     -.8437      .4063     -.6327      .2640
therapy       .4334      .3221     1.3455      .1897     -.2275     1.0944
attrib        .4039      .1808     2.2337      .0340      .0329      .7749

Standardized coefficients
             coeff
therapy      .4773
attrib       .4021

************************** TOTAL EFFECT MODEL ****************************
OUTCOME VARIABLE:
 satis

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4270      .1823      .6982     6.2421     1.0000    28.0000      .0186
```

```
Model
              coeff         se          t          p       LLCI       ULCI
constant     -.3271      .2233    -1.4649      .1541     -.7846      .1303
therapy       .7640      .3058     2.4984      .0186      .1376     1.3904

Standardized coefficients
             coeff
therapy      .8414

************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI       c_ps
      .7640      .3058     2.4984      .0186      .1376     1.3904      .8414

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI      c'_ps
      .4334      .3221     1.3455      .1897     -.2275     1.0944      .4773

Indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
attrib      .3306      .1702      .0377      .7004

Partially standardized indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
attrib      .3641      .1856      .0455      .7677
```

Write-up

Regression analysis was used to investigate the hypothesis that social support mediates the effect of age on depression. Results indicated that therapy was a significant predictor of attributional positivity, B = .82, SE = .30, 95%CI[.21,1.43], β = .91, p = .01, and that attributional positivity was a significant predictor of satisfaction, B = .404, SE = .181, 95%CI[.03,.77], β = .40, p = .03. Approximately 31% of the variance in satisfaction was accounted for by the predictors (R^2 = .31). The indirect effect was tested using a percentile bootstrap estimation approach with 10000 samples (Shrout & Bolger, 2002), implemented with the PROCESS macro Version 4.2 beta (Hayes, 2017).[3] These results indicated the indirect coefficient was significant, B = .33, SE = .17, 95%CI[.04,.70], partially standardized β = .36.[4] Receiving therapy was associated with satisfaction scores that were approximately .33 points higher as mediated by attributional positivity. Because therapy was no longer a significant predictor of satisfaction after controlling for the mediator, attributional positivity, B = .43, SE = .32, 95%CI[-.23,1.09], β = .47, p = .19, the results are consistent with full mediation.

Note: I am always cautious about using causal language and so saying something like the results "support the mediational hypothesis" or are "consistent with the mediational hypothesis" is preferrable. This applies to any application, even experimental, but especially when data are cross-sectional and nonexperimental.

---

[3] Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach, third edition*. Guilford Publications.

[4] Only the partially standardized coefficient was produced by PROCESS in this example because the predictor was dichotomous. In this case, PROCESS standardizes the outcome variable and mediator but not predictor variable because it is dichotomous. The partially standardized indirect coefficient therefore represents the standard deviation difference in y in the two therapy groups as mediated by attributional positivity.