WEEK 2
Path Mediation

Data Analysis for Psychology in R 3

dapR3 Team

Department of Psychology
The University of Edinburgh

Learning Objectives

1. Understand the purpose of mediation models and the conceptual challenges
2. Be able to describe direct, indirect and total effects in a mediation model.
3. Estimate and interpret a mediation model using `lavaan`

Part 1: Introduction to mediation

Part 2: Direct, indirect and total effects

Part 3: Estimating mediation in $lavaan$

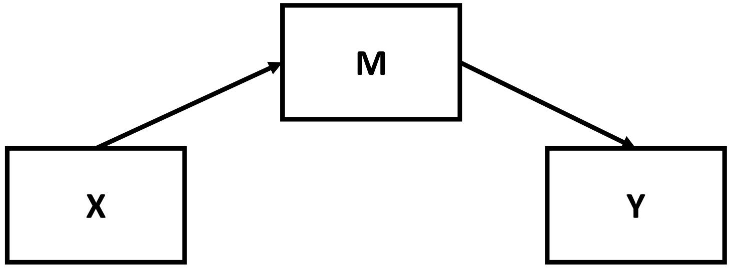
Part 4: Reporting

Mediation

- Is when a predictor X, has an effect on an outcome Y, via a mediating variable M
- The mediator **transmits** the effect of X to Y
- Examples of mediation hypotheses:
    - Conscientiousness (X) affects health (Y) via health behaviours (M)
    - Conduct problems (X) increase the risk of depression (Y) via peer problems (M)
    - Attitudes to smoking (X) predict intentions to smoke (M) which in turn predicts smoking behaviour (Y)
    - An intervention (X) to reduce youth crime (Y) works by increasing youth self-contol (M)

Visualising a mediation model

- In a SEM diagram we can represent mediation as:



Mediation...not to be confused with moderation

- Mediation is commonly confused with **moderation**
- Moderation is when a moderator z modifies the effect of X on Y
  - e.g., the effect of X on Y is stronger at higher levels of Z
  - Also known as an **interaction** between X and Z
- Examples of moderation could be:
  - An intervention (X) works better to reduce bullying (Y) at older ages (Z) of school pupil
  - The relation between stress (X) and depression (Y) is lower for those scoring higher on spirituality (Z)

End of Part 1

Part 1: Introduction to mediation

Part 2: Direct, indirect and total effects

Part 3: Estimating mediation in *lavaan*

Part 4: Reporting

Direct and indirect effects in mediation

- We seldom hypothesise that a mediator completely explains the relation between X and Y
- More commonly, we expect both **indirect effects** and **direct effects** of X on Y
    - The indirect effects of X on Y are those transmitted via the mediator
    - The direct effect of X on Y is the remaining effect of X on Y

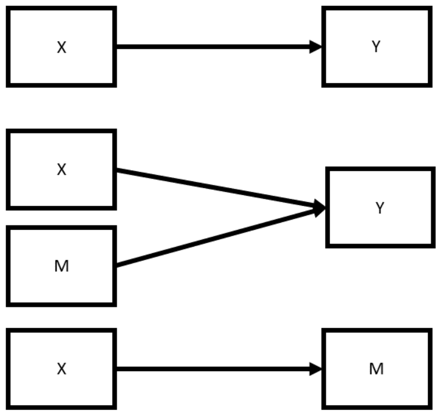
Visualing direct and indirect effects in mediation

Testing mediation

- Traditionally, mediation was tested using a series of separate linear models:

    1. Y~X
    2. Y~X+M
    3. M~X

- May see this referred to as th Baron and Kenny approach.



Traditional methods for mediation

- The three regression models:
  1. Y~X
  2. Y~X+M
  3. M~X

- Model 1 estimates the overall effect of X on Y
- Model 2 estimates the partial effects of X and M on Y

- Model 3 estimates the effect of X on M

- If the following conditions were met, mediation was assumed to hold:

  - The effect of X on Y (eq.1) is significant
  - The effect of X on M (eq.3) is significant
  - The effect of X on Y becomes reduced when M is added into the model (eq.2)

Limitations of traditional methods for mediation

- Low power
- Very cumbersome for multiple mediators, predictors, or outcomes
- You don't get an estimate of the magnitude of the indirect effect
- Much better way: **path mediation model**

BREAK QUIZ

- Quiz question:
    - Which of these hypotheses is a mediation hypothesis?
        - 1) Vocabulary development in childhood follows a non-linear trajecrtory
        - 2) The effects of conscientiousness on academic achievement are stronger at low levels of cognitive ability
        - 3) Poverty affects child behaviour problems through increasing parental stress
        - 4) Earlier pubertal onset increases the risk of antisocial behaviour only in girls and not boys

End of Part 2

Part 1: Introduction to mediation

Part 2: Direct, indirect and total effects

Part 3: Estimating mediation in *lavaan*

Part 4: Reporting

WELCOME BACK

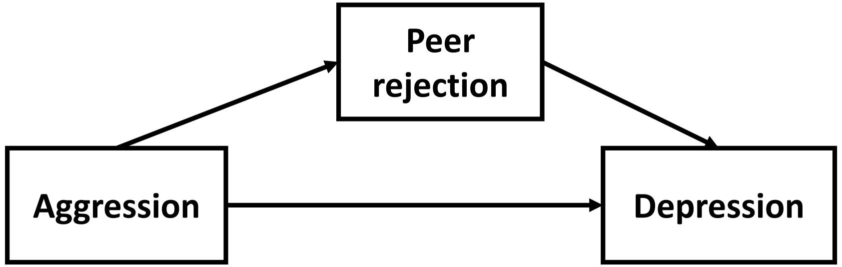
- Welcome back!
- The answer to the quiz question is...
  - Which of these hypotheses is a mediation hypothesis?
    - 1) Vocabulary development in childhood follows a non-linear trajecrtory
    - 2) The effects of conscientiousness on academic achievement are stronger at low levels of cognitive ability
    - 3) **Poverty affects child behaviour problems through increasing parental stress**
    - 4) Earlier pubertal onset increases the risk of antisocial behaviour only in girls and not boys

Testing a path mediation model in lavaan

- Specification
    - Create a lavaan syntax object
- Estimation
    - Estimate the model using e.g., maximum likelihood estimation
- Evaluation/interpretation
    - Inspect the model to judge how good it is
    - Interpret the parameter estimates

Example

- Does peer rejection mediate the association between aggression and depression?



The data

```
slice(agg.data2, 1:10)
```

```
##         Dep       PR      Agg
## 1   1.8730  0.93073  3.36176
## 2   0.5729  0.03075  0.87809
## 3   1.3726  0.54845 -0.23319
## 4  -0.0117  0.29221 -0.60339
## 5  -0.3733  0.65174 -0.05626
## 6  -0.5427 -0.02317 -0.73165
## 7   0.4146  0.64086  0.14327
## 8  -1.0071  0.18180 -0.25944
## 9   1.0511 -0.43144 -0.59311
## 10 -3.2104 -1.93771 -0.51254
```

Mediation Example

- Does peer rejection mediate the association between aggression and depression?

```
model1<-'Dep ~ PR       # Depression predicted by peer rejection
         Dep ~ Agg      # Depression predicted by aggression (the direct effect)
         PR ~ Agg       # Peer rejection predicted by aggression
'
```

- Estimate the model

```
model1.est<-sem(model1, data=agg.data2)
```

The model output

```r
summary(model1.est, fit.measures=T)
```

```
## lavaan 0.6-12 ended normally after 1 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         5
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Model Test Baseline Model:
##
##   Test statistic                               219.872
##   Degrees of freedom                                 3
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.000
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1294.513
##   Loglikelihood unrestricted model (H1)      -1294.513
```

Things to note from the model output

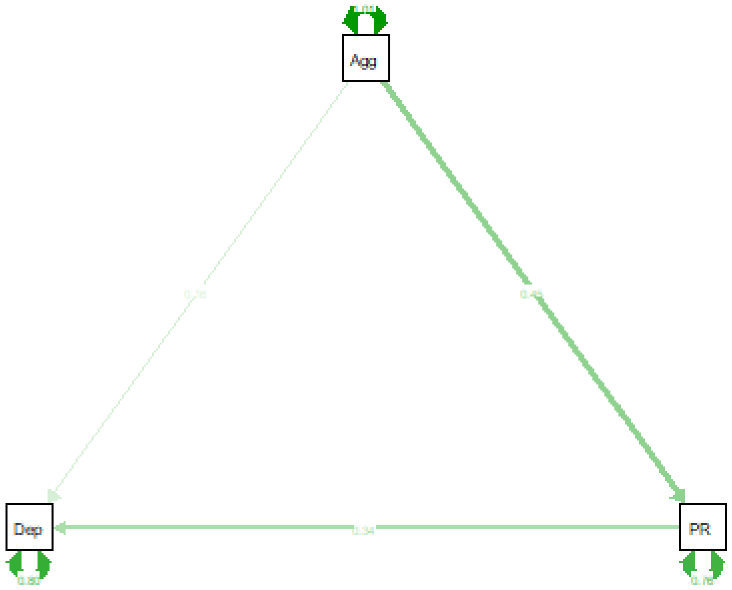
- All three regressions paths are statistically significant
- The model is **just-identified**
    - The degrees of freedom are equal to 0
    - The model fit cannot be tested
    - The model fit statistics (TLI, CFI, RMSEA, SRMR) all suggest perfect fit but this is meaningless

Visualising the model using `semPaths()`

- We can use semPaths() from the semPlot package to help us visualise the model
    - Shows the parameter estimates within an SEM diagram

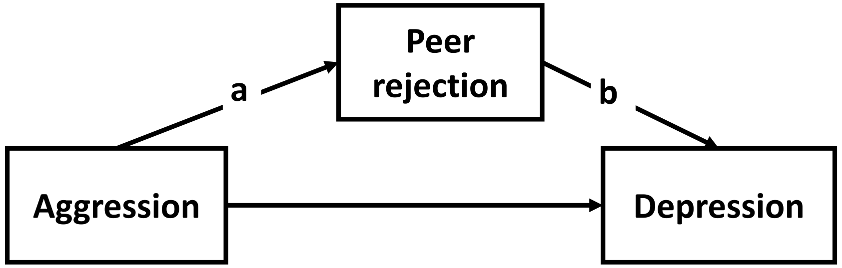
```
library(semPlot)
semPaths(model1.est, what='est')
```

Visualising the model using `semPaths()`



Calculating the indirect effects

- To calculate the indirect effect of X on Y in path mediation, we need to create some new parameters
- The indirect effect of X on Y via M is:
    - $a * b$
    - a = the regression coefficient for M~X
    - b = the regression coefficient for Y~M



Calculating indirect effects in lavaan

- To calculate the indirect effect of X on Y in lavaan we:
- Use parameter labels 'a' and 'b' to label the relevant paths
    - a is for the effect of X on M
    - b is for the effect of M on Y
- Use the ':=' operator to create a new parameter 'ind'
    - 'ind' represents our indirect effect

```
model1<-'Dep~b*PR
         Dep~Agg
         PR~a*Agg
         ind:=a*b
         '
```

Indirect effects in the output

```
model1.est<-sem(model1, data=agg.data2)
summary(model1.est)
```

```
## lavaan 0.6-12 ended normally after 1 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         5
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Dep ~
##     PR          (b)    0.336    0.046    7.308    0.000
##     Agg                0.159    0.045    3.560    0.000
##   PR ~
##     Agg         (a)    0.446    0.039   11.505    0.000
##
```

Statistical significance of the indirect effects

- Default method of assessing the statistical significance of indirect effects assume normal sampling distribution
- May not hold for indirect effects which are the product of regression coefficients
- Instead we can use **bootstrapping**
    - Allows 95% confidence intervals (CIs) to be computed
    - If 95% CI includes 0, the indirect effect is not significant at alpha=.05

Bootstapped CIs for indirect effect in lavaan

```
  model1<-'Dep~b*PR
           Dep~Agg
           PR~a*Agg
ind:=a*b'

model1.est<-sem(model1, data=agg.data2, se='bootstrap') #we add the argument se='bootstrap'
```

Output for bootstrapped CIs

```r
summary(model1.est, ci=T) # we add the argument ci=T to see the confidence intervals in the output
```

```
## lavaan 0.6-12 ended normally after 1 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         5
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                            Bootstrap
##   Number of requested bootstrap draws             1000
##   Number of successful bootstrap draws            1000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|) ci.lower ci.upper
##   Dep ~
##     PR         (b)    0.336    0.046    7.366    0.000    0.247    0.424
##     Agg               0.159    0.050    3.160    0.002    0.063    0.262
##   PR ~
##     Agg        (a)    0.446    0.034   13.122    0.000    0.379    0.510
##
## Variances:
```

Total effects in path mediation

- As well as the direct and indirect effect, it is often of interest to know the **total** effect of X on Y

$$Total = Indirect + Direct$$

Total effects in path mediation

$$Total = a * b + c$$

Total effect in lavaan

```
  model1<-'Dep~b*PR
           Dep~c*Agg          # we add the label c for our direct effect
           PR~a*Agg
ind:=a*b
total:=a*b+c                  # we add a new parameter for the total effect'

model1.est<-sem(model1, data=agg.data2, se='bootstrap') #we add the argument se='bootstrap'
```

Total effect in lavaan output

```r
summary(model1.est, ci=T)
```

```
## lavaan 0.6-12 ended normally after 1 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         5
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                            Bootstrap
##   Number of requested bootstrap draws             1000
##   Number of successful bootstrap draws            1000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|) ci.lower ci.upper
##   Dep ~
##     PR         (b)    0.336    0.045    7.471    0.000    0.246    0.422
##     Agg        (c)    0.159    0.047    3.361    0.001    0.069    0.251
##   PR ~
##     Agg        (a)    0.446    0.034   13.127    0.000    0.378    0.513
##
## Variances:
```

Why code the total effect in lavaan?

- We could have just added up the coefficients for the direct and indirect effects
- By coding it in lavaan, however, we can assess the statistical significance of the total effect
- Useful because sometimes the direct and indirect effects are not individually significant but the total effect is
    - May be especially relevant in cases where there are many mediators of small effect

Interpreting the total, direct, and indirect effect coefficients

- The total effect can be interpreted as the **unit increase in Y expected to occur when X increases by one unit**
- The indirect effect can be interpreted as the **unit increase in Y expected to occur via M when X increases by one unit**
- The direct effect can be interpreted as the **unit increase in Y expected to occur with a unit increase in X over and above the increase transmitted by M**
- **Note**: 'direct' effect may not actually be direct - it may be acting via other mediators not included in our model

Standardised parameters

- As with CFA models, standardised parameters can be obtained using:

```
summary(model1.est, ci=T, std=T)
```

Standardised parameters

```
## lavaan 0.6-12 ended normally after 1 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         5
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                            Bootstrap
##   Number of requested bootstrap draws             1000
##   Number of successful bootstrap draws            1000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|) ci.lower ci.upper
##   Dep ~
##     PR          (b)    0.336    0.045    7.471    0.000    0.246    0.422
##     Agg         (c)    0.159    0.047    3.361    0.001    0.069    0.251
##   PR ~
##     Agg         (a)    0.446    0.034   13.127    0.000    0.378    0.513
##    Std.lv  Std.all
##
##     0.336    0.332
##     0.159    0.162
##
```

BREAK QUIZ

- Time for a pause
- Quiz question
    - If the effect of X on M is b=.30 and the effect of M on Y is b=.10, what is the indirect effect of X on Y?
        - 1) b=.40
        - 2) b=.03
        - 3) b=.30
        - 4) b=.10

End of Part 3

Part 1: Introduction to mediation

Part 2: Direct, indirect and total effects

Part 3: Estimating mediation in *lavaan*

Part 4: Reporting

Welcome back

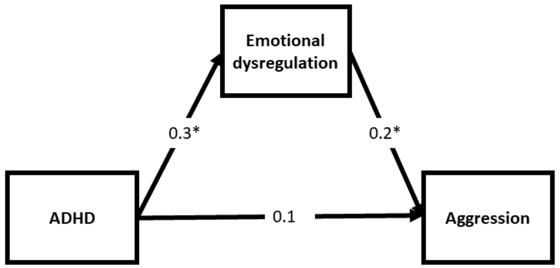
- The answer to the quiz question is...
- Quiz question
    - If the effect of X on M is b=.30 and the effect of M on Y is b=.10, what is the indirect effect of X on Y?
        - 1) b=.40
        - 2) **b=.03**
        - 3) b=.30
        - 4) b=.10

Reporting path mediation models

- Methods/ Analysis Strategy
    - The model being tested
    - e.g. 'Y was regressed on both X and M and M was regressed on X'
    - The estimator used (e.g., maximum likelihood estimation)
    - The method used to test the significance of indirect effects ('bootstrapped 95% confidence intervals')
- Results
    - Model fit (for over-identified models)
    - The parameter estimates for the path mediation and their statistical significance
    - Can be useful to present these in a SEM diagram
    - The diagrams from R not considered 'publication quality' draw in powerpoint or similar

Reporting path mediation models - example of SEM diagram with results

- Include the key parameter estimates
- Indicate statistically significant paths (e.g. with an '*')
- Include a figure note that explains how statistically significant paths (and at what level) are signified



Reporting path mediation models - the indirect effects

- Results
  - The coefficient for the indirect effect and the bootstrapped 95% confidence intervals
  - Common to also report **proportion mediation**:

$$\frac{indirect}{total}$$

- However, important to be aware of limitations:
  - Big proportion mediation possible when total effect is small - makes effect seem more impressive
  - Small proportion mediation even when total effect is big - can underplay importance of effect
  - Should be interpreted in context of total effect
- Tricky interpretation if there are a mix of negative and positive effects involved

Extensions of path mediation models

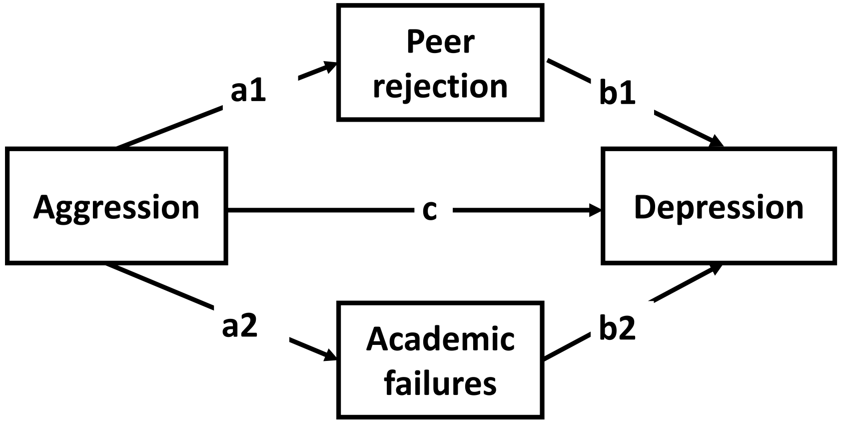
- We can extend our path mediation model in various ways:
  - Several mediators in sequence or parallel
  - Multiple outcomes
  - Multiple predictors
  - Multiple groups (e.g., comparing direct and indirect effects across males and females)
  - Add covariates to adjust for potential confounders

Example: Multiple mediation model

```
model2<-'Dep~b2*Aca
         Aca~a2*Agg
         Dep~b1*PR
         PR~a1*Agg
         Dep~c*Agg

        ind1:=a1*b1
        ind2:=a2*b2

        total=a1*b1+a2*b2+c
'
```



Other path analysis models

- Path mediation models are a common application of path models
    - But they are just one example
- Anything that can be expressed in terms of regressions between observed variables can be tested as a path model
    - Can include ordinal or binary variables
    - Can include moderation
- Other common path analysis models include:
    - Autoregressive models for longitudinal data
    - Cross-lagged panel models for longitudinal data

Making model modifications

- You may want to make some modifications to your initially hypothesised model
    - non-significant paths that you want to trim
    - include some additional paths not initially included
- Remember that this now moves us into exploratory territory where:
    - Model modifications should be substantively as well as statistically justifiable
    - You must be aware of the possibility that you are capitalising on chance
    - You should aim to replicate the modifications in independent data

Cautions regarding path analysis models

- Assumption that the paths represent causal effects is only an assumption
  - Especially if using cross-sectional data
  - Mediation models should ideally be estimated on longitudinal data.
  - X time 1
  - M time 2
  - Y time 3
- The parameters are only accurate if the model is correctly specified

Cautions: Indistinguishable models

Measurement error in path analysis

- Path analysis models use observed variables
    - Assumes no measurement error in these variables
- Path coefficients likely to be attenuated due to unmodelled measurement error
- Structural equation models solve this issue
    - They are path analysis models where the paths are between latent rather than observed variables
    - ...very brief comment on this in the final week

Path analysis summary

- Path analysis can be used to fit sets of regression models
    - Common path analysis model is the path mediation model
    - But very flexible huge range of models that can be tested
- In R, path analysis can be done using the `sem()` function in `lavaan`
- Need to be aware that we aren't *testing* causality but assuming it

End