TCSM week 6: moderation

- Next week: One last tutorial covering all topics
- Before the exam on Monday: One last optional tutorial

OUTLINE

- Moderation in regression
- Moderation in SEM
  - With a little mediation to spice it up
  - Comparing nested models in lavaan
- Modifying models
  - Modification indices

What is moderation?

A moderator is:

- a third variable that affects the relation between two other variables.

- E.g. the relation between ‘being rejected’ and ‘problematic social media use’ might be different for teens low- vs high in ‘narcissism’.

  Hawk, S. T., van den Eijnden, R. J., van Lissa, C. J., & ter Bogt, T. F. (2019). Narcissistic adolescents' attention-seeking following social rejection: Links with social media disclosure, problematic social media use, and smartphone stress. Computers in Human Behavior, 92, 65-75.

Moderation, how to study it?
(Baron & Kenny)

Case 1: IV dichotomous, Moderator dichotomous

Case 1: 2x2 ANOVA

Case 2: IV continuous, mod dichotomous

Case 2: test correlation for each level of the moderator

Case 3: IV dichotomous, mod continuous

Case 3: include interaction effect

Case 4: IV continuous, Mod Continuous

Case 4: include interaction effect

Interaction with continuous moderator

Continuous moderators

- Does the relationship between X and Y depend on the value of moderator Z?
- https://utrecht-university.shinyapps.io/cj_moderation/

Insert Web Page

This app allows you to insert secure web pages starting with https:// into the slide deck. Non-secure web pages are not supported for security reasons.

Please enter the URL below.

https://	utrecht-university.shinyapps.io/cj_moderation/

Note: Many popular websites allow secure access. Please click on the preview button to ensure the web page is accessible.

Interaction with continuous moderator in lavaan

- Center all predictors (generally a good practice)
- Calculate “interaction term”: X * Z
- Add interaction term to your data, e.g.
  `data$int <- data$X * data$Z`
- Include int, X and Z as predictors of your outcome

- A nice trick:

```
sem_data <- model.matrix(~X*Z, data)
head(sem_data)
```

(Intercept)	x	z	x:z
1	2	5	10
1	10	7	70
1	1	9	9
1	1	10	10
1	6	4	24
1	5	2	10

Interaction with continuous moderator in lavaan

- Center all predictors (generally a good practice)
- Calculate “interaction term”: X * Z
- Add interaction term to your data, e.g.

```
data$int <- data$X * data$Z
```

- Include int, X and Z as predictors of your outcome

- A nice trick:

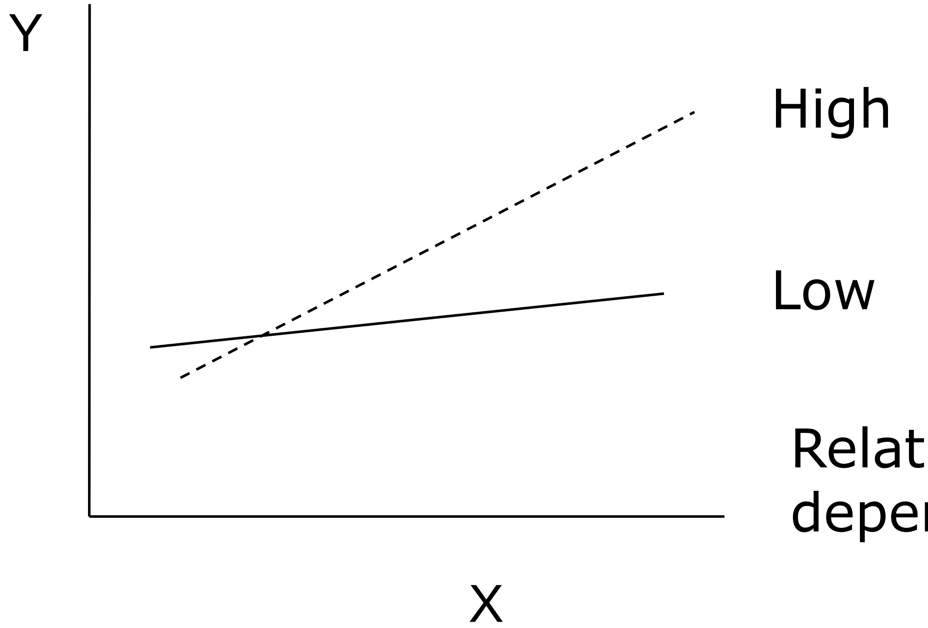
```
sem_data <- model.matrix(~X*Z, data)[, -1]
head(sem data)
```

x	z	x:z
2	5	10
10	7	70
1	9	9
1	10	10
6	4	24
5	2	10

Interaction with binary moderator

Moderation

- Is the relationship between X and Y different for low vs high narcissists?
- Does narcissism <u>moderate</u> the relations specified in the model?
- Is there an <u>interaction</u> effect?



Relation between X and Y depends on narcissism.

In regression

- Regress “social media use” on rejection
  - Include dummy for narcissism
  - Include interaction (narcissism * rejection)
  - Does the interaction add significantly to explained variance? (hierarchical models)

In regression

```
>head(df)
```

reject	narcissist	SM
3.470118	High	4.728864
2.48158	Low	4.97856
3.446392	Low	6.604417
4.705579	Low	6.735068
2.625999	Low	2.213947
4.320125	High	6.348611

Using hierarchical regression

```
> res_main <- lm(SM ~ reject + narcissist,
                 data = df)
> res_int  <- lm(SM ~ reject*narcissist,
                 data = df)

> anova(res_main, res_int)

Analysis of Variance Table

Model 1: SM ~ reject + narcissist
Model 2: SM ~ reject * narcissist
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     47 231.17
2     46 223.37  1    7.8047 1.6073 0.2113
```

Inspect results

```
> summary(res_main)

Call:
lm(formula = SM ~ reject + narcissist, data = df)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.3834     0.9977   1.387  0.17211
reject           0.8237     0.2413   3.414  0.00133 **
narcissistHigh  -0.6192     0.7167  -0.864  0.39201
---
Residual standard error: 2.218 on 47 degrees of freedom
Multiple R-squared:  0.2095,  Adjusted R-squared:  0.1759
F-statistic: 6.228 on 2 and 47 DF,  p-value: 0.003987
```

Inspect results

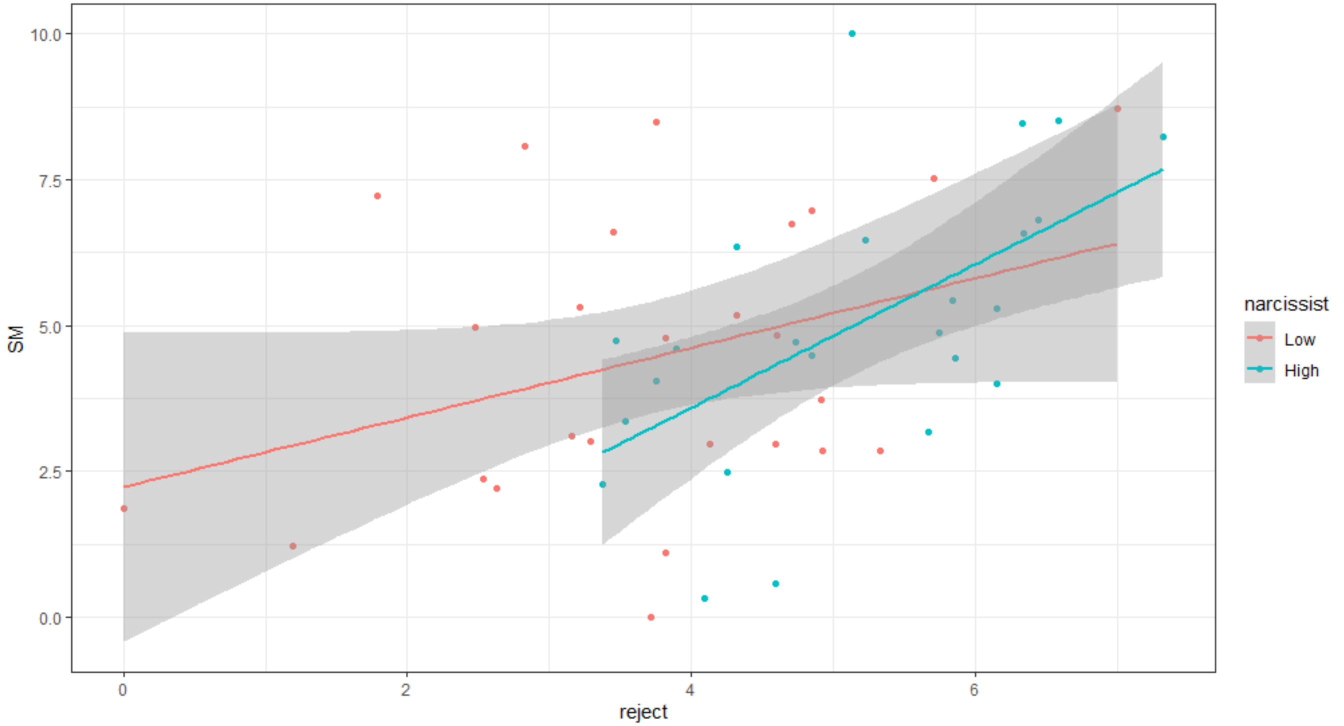
```
> summary(res_int)

Call:
lm(formula = SM ~ reject * narcissist, data = df)

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             2.2280     1.1944   1.865   0.0685 .
reject                  0.5968     0.2992   1.995   0.0520 .
narcissistHigh         -3.5618     2.4278  -1.467   0.1492
reject:narcissistHigh   0.6339     0.5000   1.268   0.2113
---
Residual standard error: 2.204 on 46 degrees of freedom
Multiple R-squared:  0.2362,  Adjusted R-squared:  0.1864
F-statistic: 4.741 on 3 and 46 DF,  p-value: 0.005786
```

Inspect interaction visually

```
> library(ggplot2)
> ggplot(df,
         aes(x = reject, y = SM, colour = narcissist)) +
         geom_point() +
         geom_smooth(method = "lm") +
         theme_bw()
```



Why not use regression?

- 1. Difficult for complicated models
    - E.g. Moderated Mediation

- 2. Cannot correct for measurement error

- 3. Cannot test fit of entire model to the data

Reading Q 4:

- The framework that Baron and Kenny present on page 1179 is one way to study complicated models (mediating moderation, or moderating mediation). They do not discuss alternatives to this framework, but (surprise!) Structural Equation Modeling is one of them. What would you think be the main advantage of SEM over the framework of Baron and Kenny.

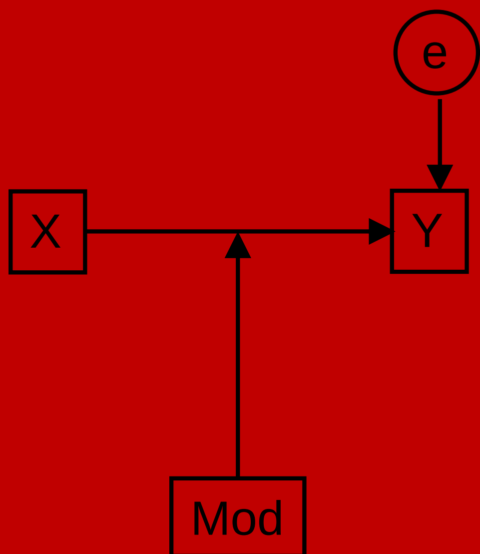
It is much more simple (practically), and all model parameters are estimated at once, so that complex interdependencies can be modeled in a better and more sophisticated way

OUTLINE

- Moderation in regression
- Moderation in SEM
  - With a little mediation to spice it up
  - Comparing nested models in lavaan
- Modifying models
  - Modification indices

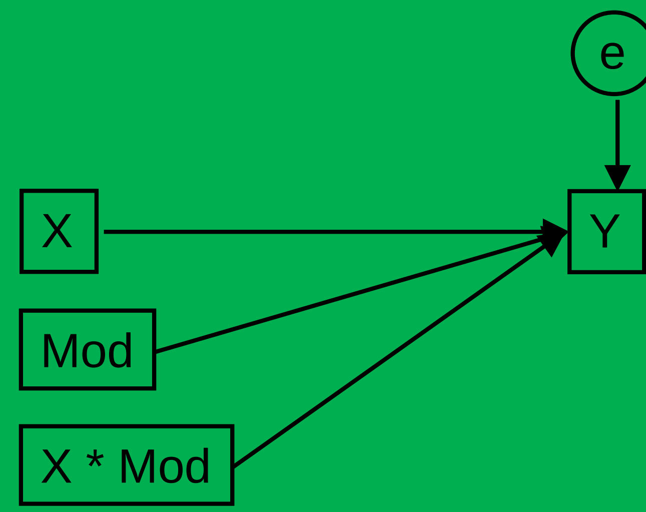
Visualization

Common shorthand



Note: this drawing convention does not accurately represent the SEM model

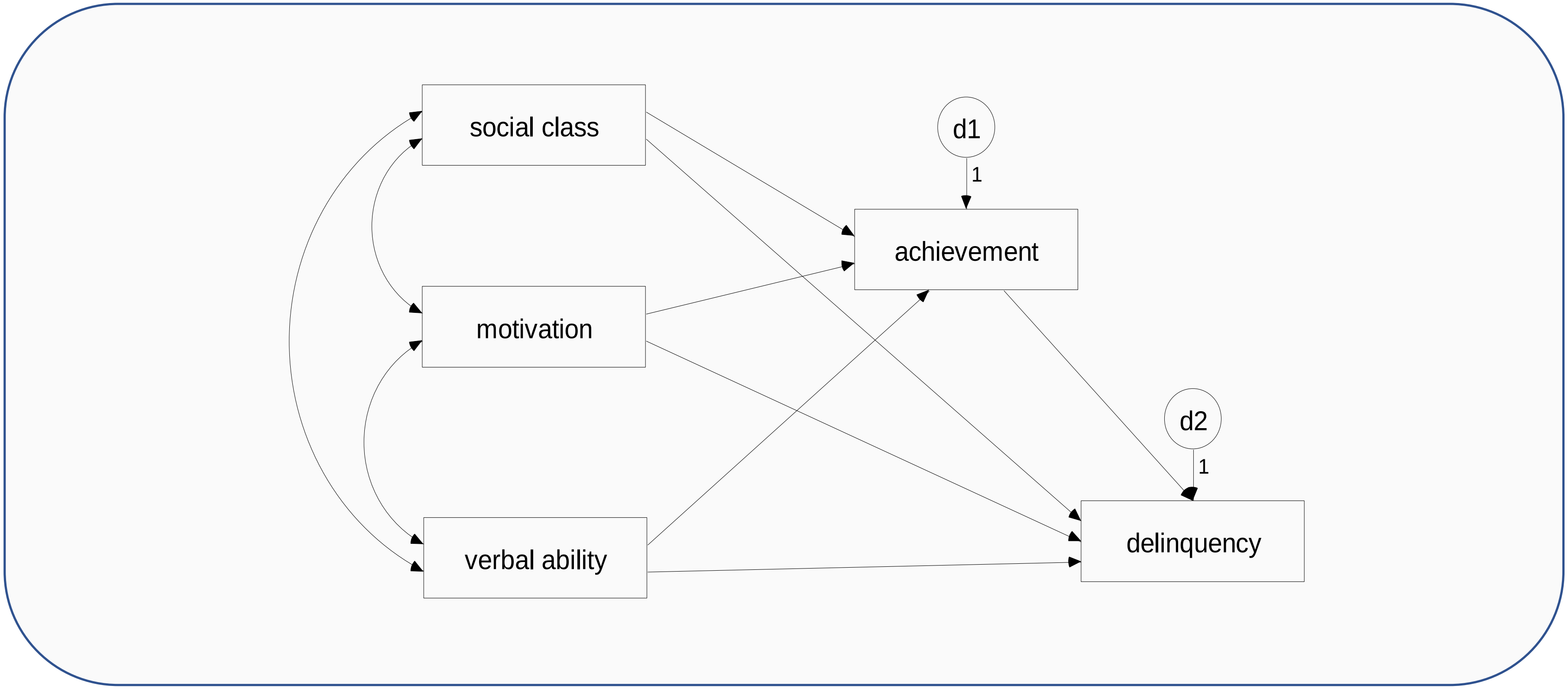
Officially



(more advanced) Example for today

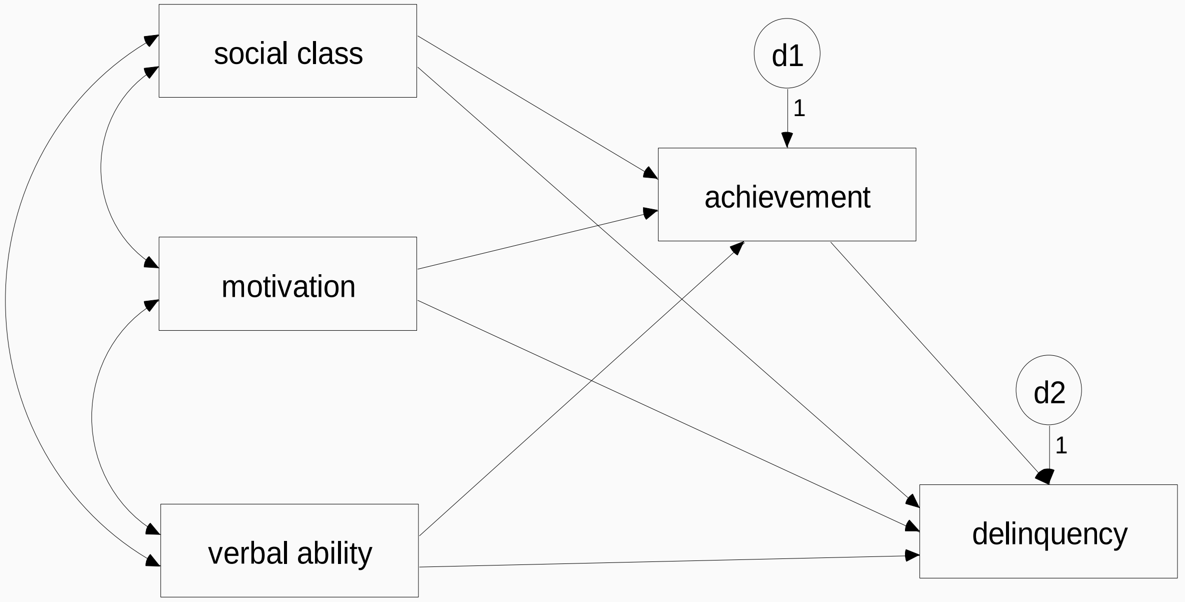
- Is adolescents' deviant behavior a predictor for later criminal behavior?
- Two theories: specificity versus generalization
- Moderation in combination with mediation

Modeling



Mediation model

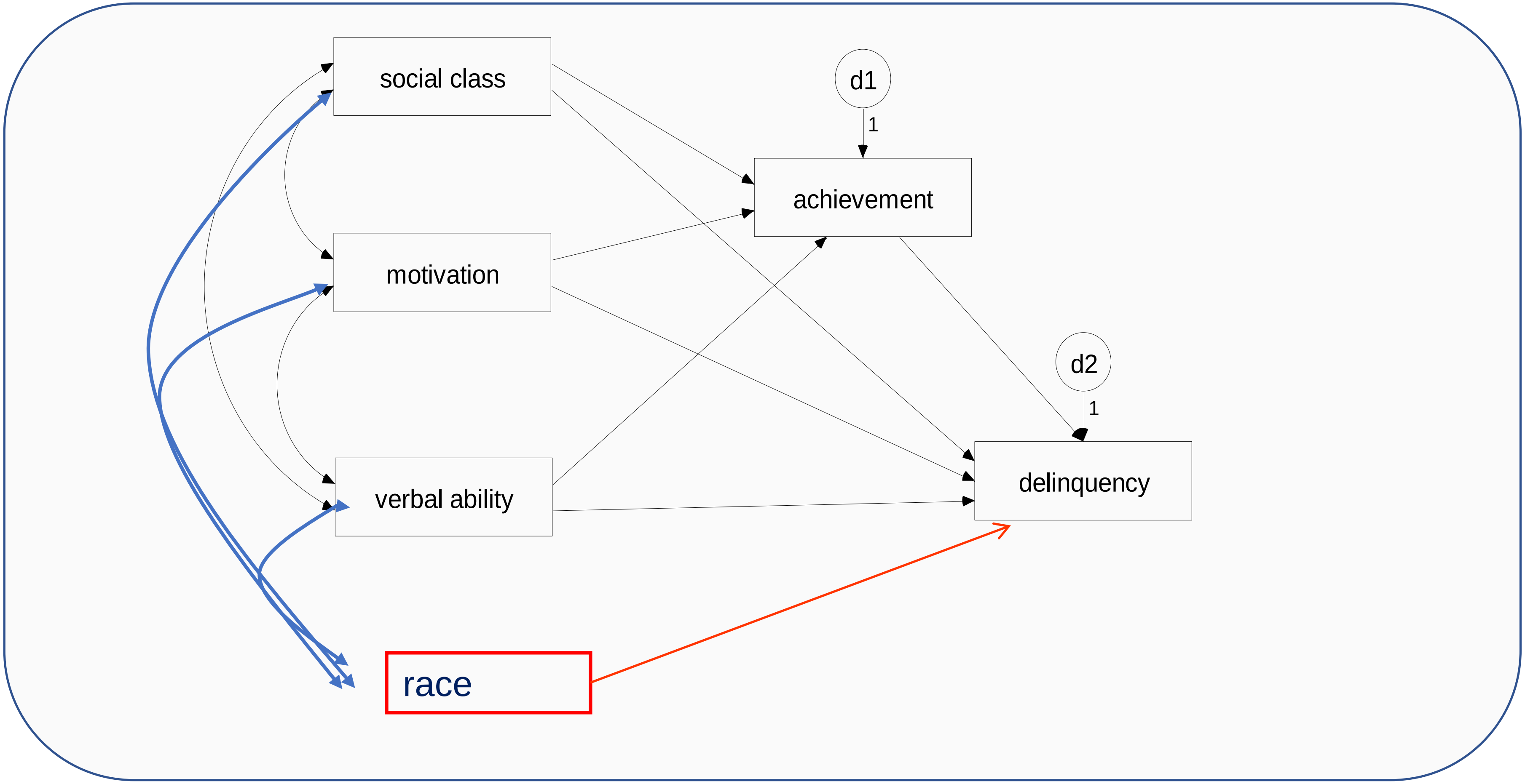
Modeling



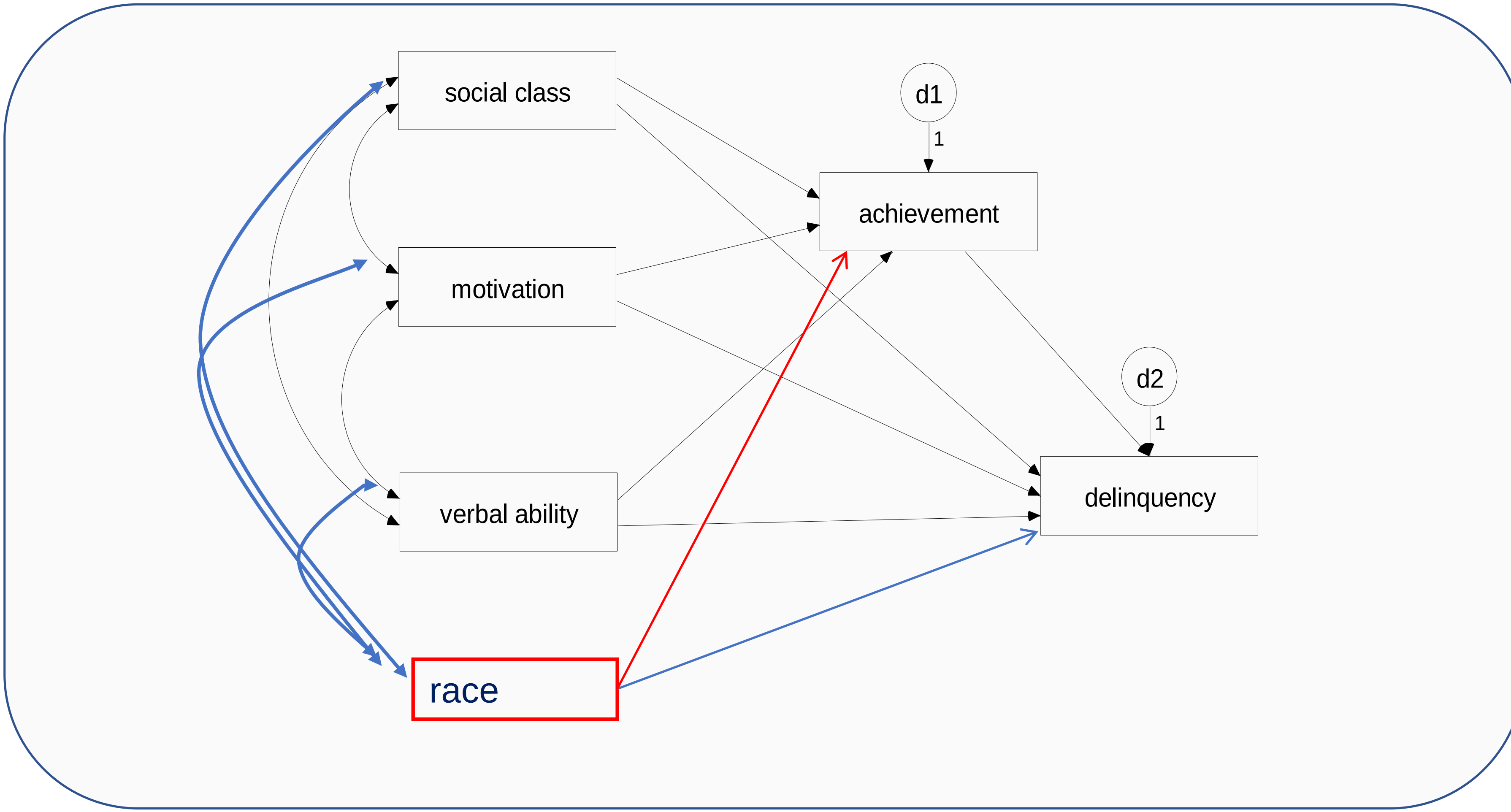
Race?

Depends on the research question!

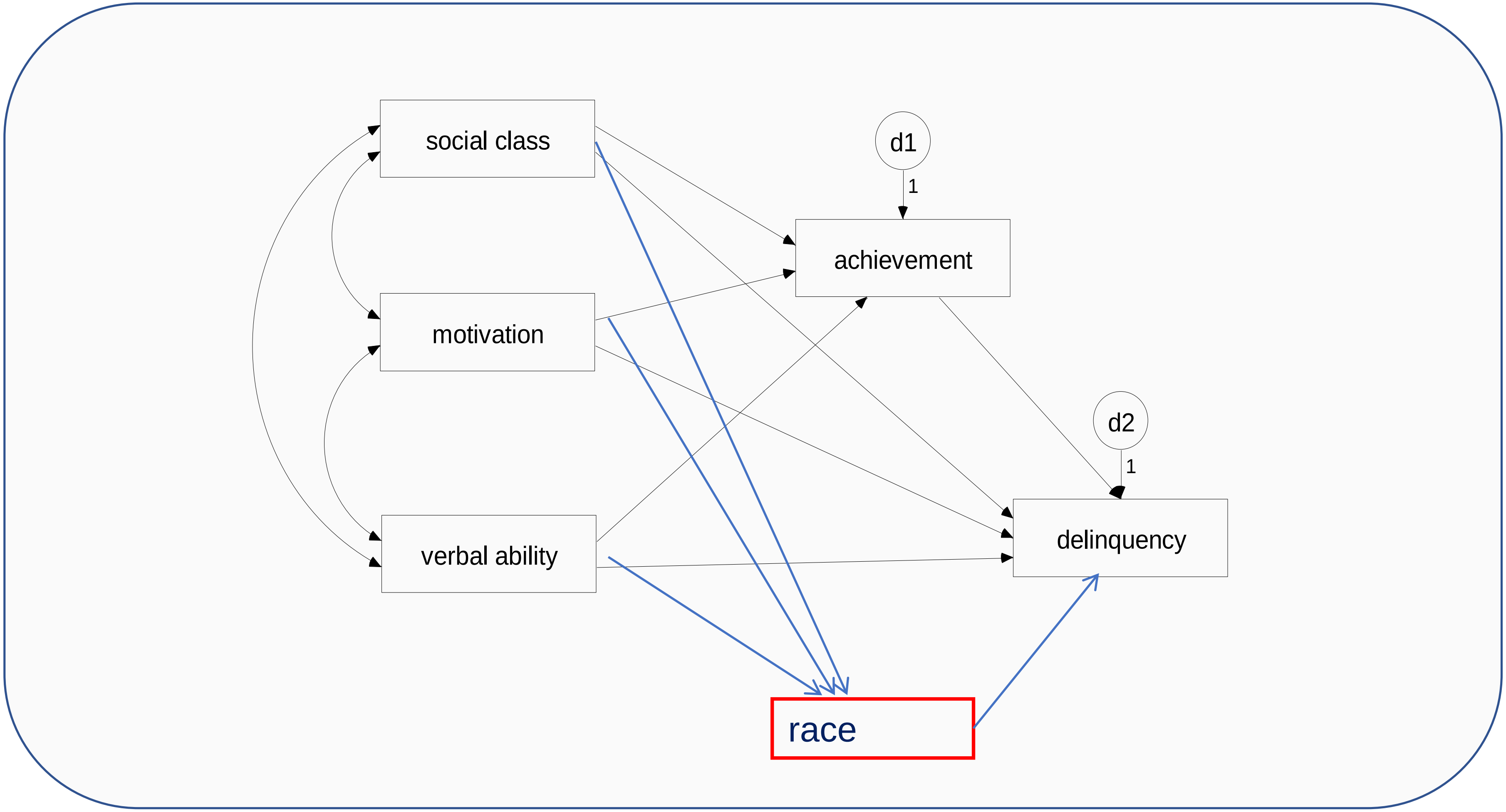
1. Predictor


social class
d1
1
achievement
motivation
d2
1
delinquency
verbal ability
race

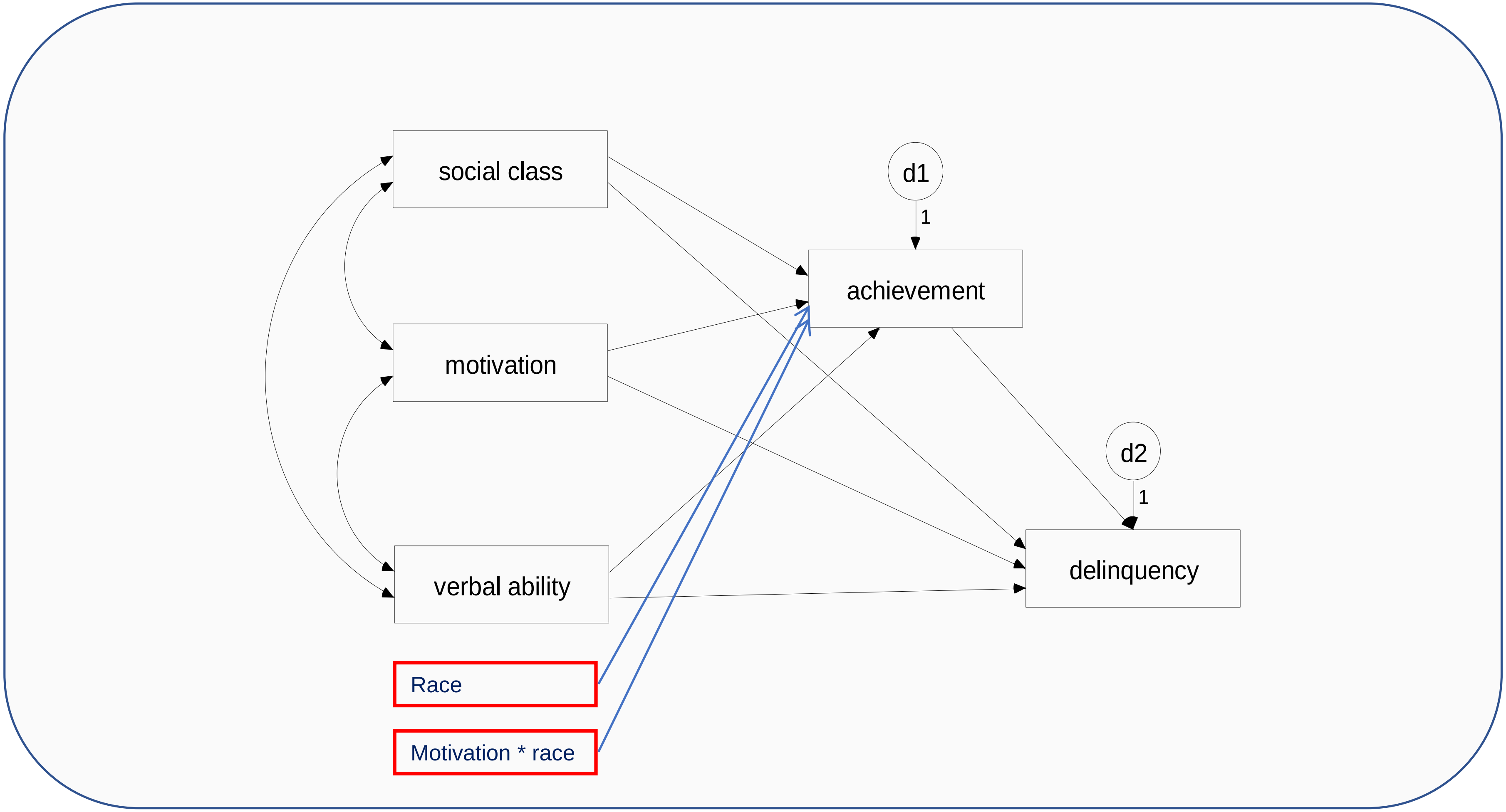
1. Predictor with mediation



2. As mediator?



3. Classic moderation (like regression)



Moderation as multiple group

In SEM: Multiple group analysis

- Main question in multiple-sample SEM is whether the values of the model parameters vary across groups.

- Examples of groups:
  - sex
  - Different nations
  - Rural – urban
  - Ethnic groups

OUTLINE

- Moderation in regression
- Moderation in SEM
  - With a little mediation to spice it up
  - Comparing nested models in lavaan
- Modifying models
  - Modification indices

Moderation as a multi-group model

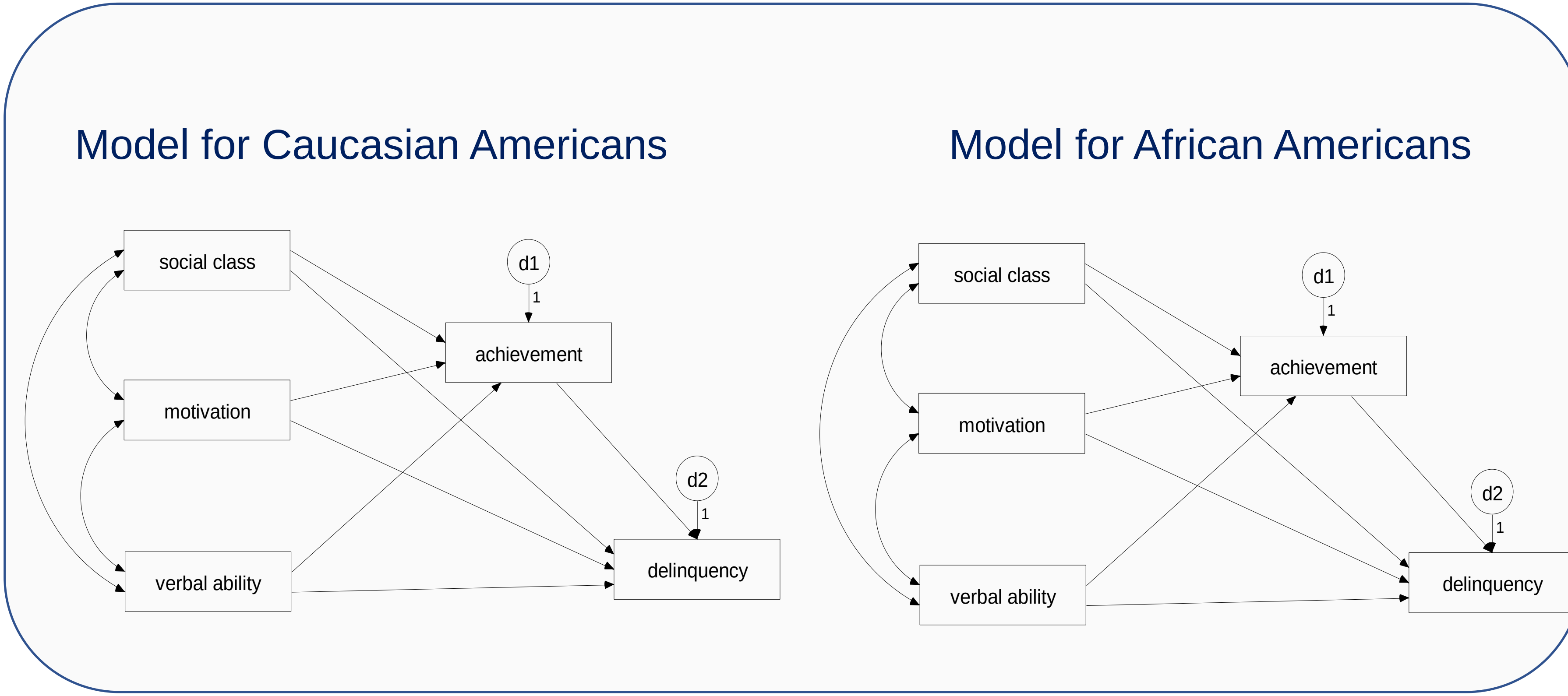
- Advantages
  - Easy to specify
    ```
    sem(model, data, group = "moderator")
    ```
  - Any parameter can be constrained or freed across groups
    ```
    model <- "X ~ c(c1, c1) * Y"
    model <- "X ~ c(f1, f2) * Y"
    ```

- Limitations:
  - Only for categorical moderators
  - Quickly becomes complicated with more than 2 groups

Example of using labels in lavaan: ANOVA with free variances

- If you want an ANOVA without assumption of equal variances:
- `model <-    "Y ~ group_variable"`
- `model <-    "Y ~ group_variable`
  `Y ~~ c(v1, v2) * Y"`

4. Multiple group moderation



- Now, not interested in effect of race on delinquency
- Rather, the paths are hypothesized to be different for African versus Caucasian Americans

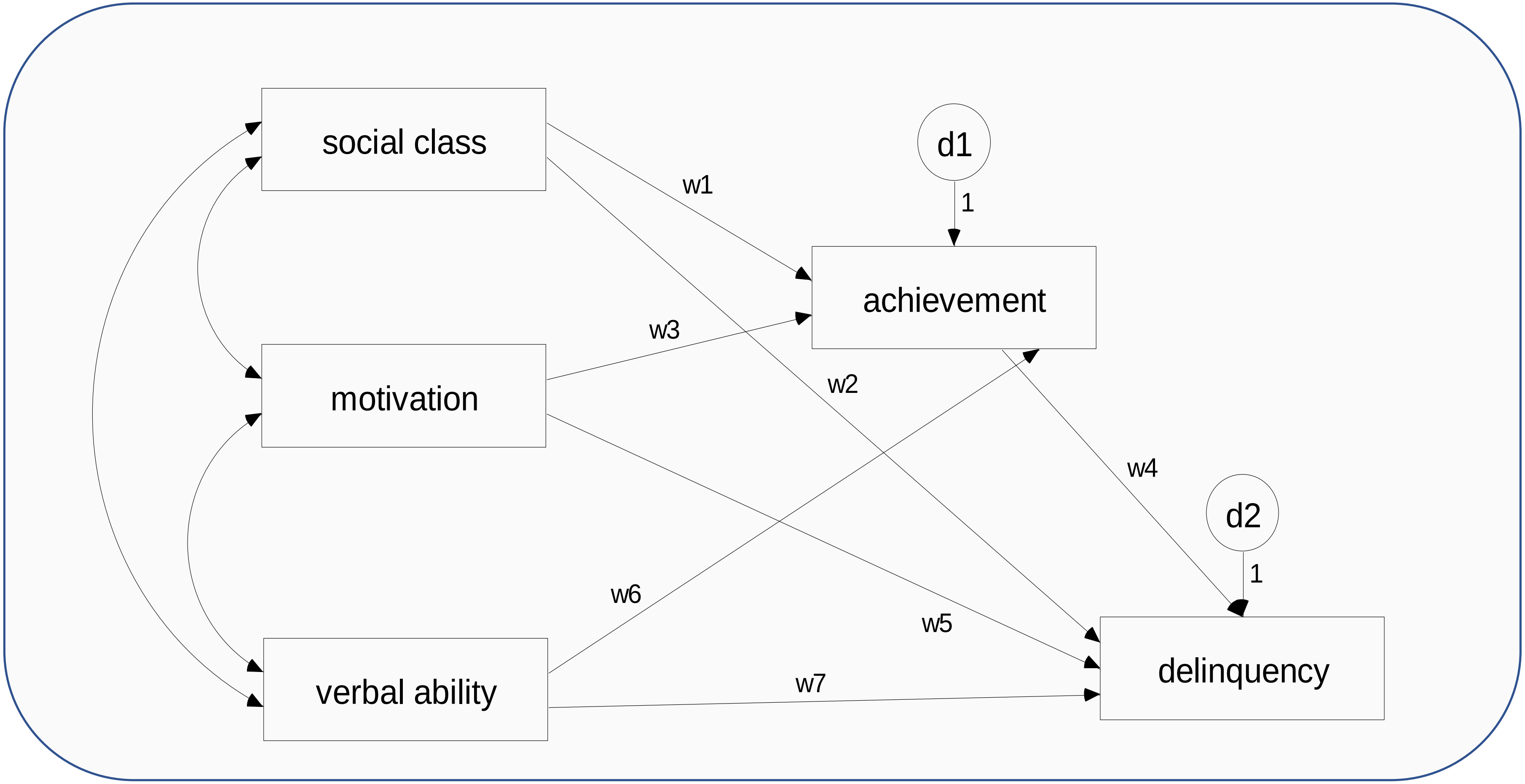
Testing in steps

- Often, model is not exactly the same

- But it can be largely the same
  - Constrain paths one-by-one
  - Manually

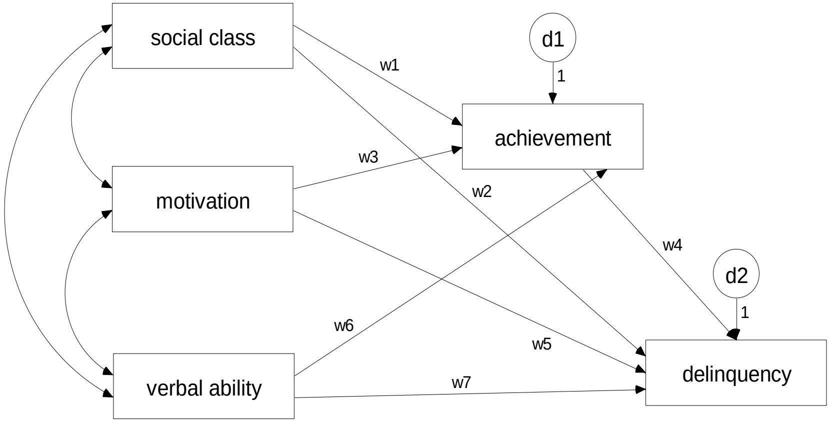
- Compare nested models against each other

lavaan setup: use labels



The letter w is used for whites. Use the letter a for African Americans. Different numbers are used for different paths.

lavaan setup: use labels

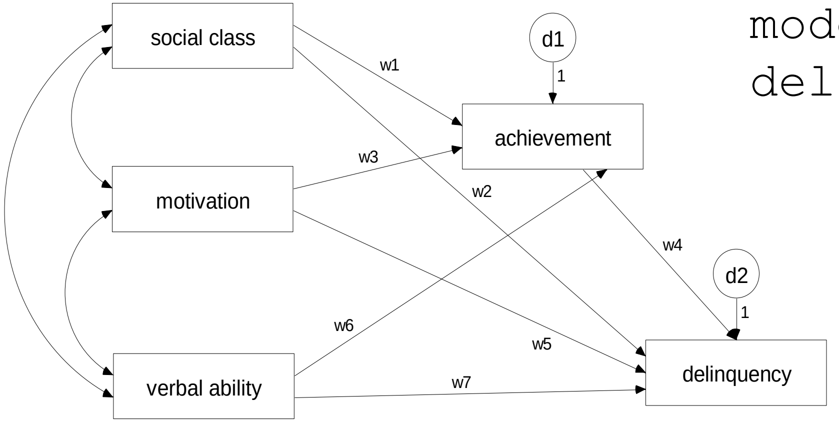



```
model <- “
delinquency ~ achievement +
                 class +
                 motivation +
                 verbal

Achievement ~ class +
                 motivation +
                 verbal
”
```

The letter w is used for whites. Use the letter a for African Americans. Different numbers are used for different paths.

lavaan setup: use labels




```
model <- "
delinquency ~ c(w4, b4) *  achievement +
                           c(w2, b2) * class +
                           c(w5, b5) * motivation +
                           c(w7, b7) * verbal

achievement ~ c(w1, b1) * class +
                           c(w3, b3) * motivation +
                           c(w6, b6) * verbal
"
```

The letter w is used for whites. Use the letter a for African Americans. Different numbers are used for different paths.

Constraining groups of parameters

- Default: All parameters freely estimated in each group
- You can constrain a group of similar parameters all at once

```
HS.model <- '  visual =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 '
fit <- cfa(HS.model,
           data = HolzingerSwineford1939,
           group = "school",
           group.equal = c("loadings"))
```

Constraining groups of parameters

- The following groups are available:

- `intercepts:` observed variable intercepts
- `means:` latent variable intercepts/means
- `residuals:` observed variable residual variances
- `residual.covariances:` observed variable residual covariances
- `lv.variances:` latent variable (residual) variances
- `lv.covariances:` latent variable (residual) covariances
- `regressions:` all regression coefficients

More options

- E.g., automatic measurement invariance testing between groups (is the measurement model identical across groups):
- http://lavaan.ugent.be/tutorial/groups.html

Bachelor thesis
Methods and Statistics

Bachelor thesis MS

- Groups of one or two students
- Passed all mandatory MS courses;
    - minor not necessary
- Usually half-time in blocks 3 and 4, other blocks/full time also possible
- Supervisors’ proposals available from your own bachelor program coordinator

Bachelor thesis MS

- Bachelor research contributes to ongoing research of MS department

- Methodological or statistical research questions
- Conduct analyses and report results

- Final report: thesis or paper, written in English

- Oral final presentation

Coordinator bachelor research MS

Mirjam Moerbeek

m.moerbeek@uu.nl

030-253 1450

Room C107 Sjoerd Groenman Building

Research master

METHODOLOGY AND STATISTICS FOR THE BEHAVIOURAL,

BIOMEDICAL AND SOCIAL SCIENCES

More information at uu.nl/master/ms

Master orientation in February

Or ask coordinator Irene Klugkist:
I.Klugkist@uu.nl

Universiteit Utrecht

University Medical Center Utrecht

UNIVERSITY OF TWENTE.

BREAK

Comparing nested models

Compare two models

- Are the weights are equal across groups?

- Compare two <u>nested</u> models

- *When is model nested?*

- Models can be distinguished by giving them different names
  - Model ‘unconstrained’: no constraints

  - Model ‘constrained’: some weights equal across groups.

Nested models in SEM

Model 2 has 1 parameter less

Does this simplification make the “distance” significantly larger?

Distance between observed (sample) covariance matrix and model-implied matrix.

If distance is significantly larger, model 2 fits worse

Nested models: constraining paths

Is the distance larger?

↓

larger Chi-square

↓

Significant chi-square difference:
Model 2 is significantly worse, retain Model 1

Non-significant chi-square difference:
Model 2 is simpler, and not much worse. Choose 2

lavaan setup: enter constraints

```
model <- “delinquency ~ achievement + class +
                        motivation + verbal
          achievement ~ class + motivation + verbal”

m_free <- sem(model, data, group = "race")
m_fix <- sem(model, data, group = "race", group.equal =
"regressions")

anova(m_free, m_fix
```

<u>OR</u>

```
library(semTools)
compareFit(free = m_free,
           fix = m_fix)
```

Which parameters to be constrained?

- In path analysis we usually only compare regression coefficients across groups

- We can constrain any other parameters to be equal across groups
  - E.g., variances and covariances

- Some of these constraints may be implausible from a theoretical point of view.
- Other constraints might represent reasonable theoretical assumptions
  - E.g., measurement model equal across groups

Constraining, how to do it

1. Start with ‘unconstrained’ model (everything free between groups)
2. Constrain paths you theoretically expect to be equal across groups
3. Compare model fits; is the constrained model significantly worse?

- What is your conclusion if the fit is significantly worse?

Constraining, how to do it

1. Start with ‘constrained’ model (**<u>everything</u>** fixed between groups)
2. Free paths you theoretically expect to be different between groups
3. Compare model fits; is the constrained model significantly worse?

- What is your conclusion if the fit is significantly worse?

Risks of step-wise approach

- Why not constrain paths one-by-one until the fit is super good?

- You are running many repeated tests

- You might end up overfitting noise:
  - Free parameter, even though differences between groups are due to chance

- Solutions:
  - Make theory-driven decisions
  - Use fit indices with a penalty for number of parameters

OUTLINE

- Moderation in regression
- Moderation in SEM
  - With a little mediation to spice it up
  - Comparing nested models in lavaan
- Modifying models
  - Modification indices

No fit!

- What to do when your initial model does not fit?
- The good and bad sides of Modification Indices

Modifying the model

- Add extra parameters
  - factor loading, path, covariance
- Which new parameters?
- Suggested by modification index (MI)
  - Sometimes called Lagrange multiplier test

Remember? Model fit

modification indices based on difference

Sample Covariances (Girls)

	wordmean	sentence	paragrap	lozenges	cubes	visperc
wordmean	68,260					
sentence	28,845	25,197				
paragrap	21,718	12,864	12,516			
lozenges	23,947	13,228	9,056	61,726		
cubes	6,840	4,036	3,356	17,416	20,265	
visperc	13,037	12,645	8,335	26,531	14,931	47,175

Implied Covariances (Girls - Default model)

	wordmean	sentence	paragrap	lozenges	cubes	visperc
wordmean	68,260					
sentence	28,859	25,197				
paragrap	21,633	12,916	12,516			
lozenges	19,583	11,692	8,765	61,726		
cubes	9,966	5,950	4,461	17,024	20,265	
visperc	16,344	9,759	7,315	27,919	14,209	47,175

or

Sample Correlations (Girls)

	wordmean	sentence	paragrap	lozenges	cubes	visperc
wordmean	1,000					
sentence	,696	1,000				
paragrap	,743	,724	1,000			
lozenges	,369	,335	,326	1,000		
cubes	,184	,179	,211	,492	1,000	
visperc	,230	,367	,343	,492	,483	1,000

	wordmean	sentence	paragrap	lozenges	cubes	visperc
wordmean	1,000					
sentence	,696	1,000				
paragrap	,740	,727	1,000			
lozenges	,302	,296	,315	1,000		
cubes	,268	,263	,280	,481	1,000	
visperc	,288	,283	,301	,517	,460	1,000

Modification index

- For each constrained parameter
  - including omitted paths…

- Modification index = estimated χ^2 decrease
  if constraint released
  - release 1 constraint (= add 1 path)
  - χ^2 decrease at least 3.84

Modification indices in lavaan

- For each nonfree parameter
- `summary(…,` **`modindices = TRUE`**`)`
- `modindices(your_analysis)`

```
 lhs op rhs          mi    epc sepc.lv sepc.all sepc
25  visual =~  x4   1.211  0.077   0.069    0.059
26  visual =~  x5   7.441 -0.210  -0.189   -0.147
27  visual =~  x6   2.843  0.111   0.100    0.092
28  visual =~  x7  18.631 -0.422  -0.380   -0.349
```

- Cutoff: 3.84 (chi square with df = 1 and alpha = .05)
- In practice:
  - Focus on biggest MIs first
  - Did you make a mistake?
  - Balance theory and pragmatism

Be careful!

- Never follow modification indices automatically!
- Make theory-driven decisions
  - e.g., no directional paths against time flow

- After enough modifications, *something* will fit
  - but possibly absurd model
  - and likely to replicate badly

  - This is the result of **overfitting** noise in the data!!!