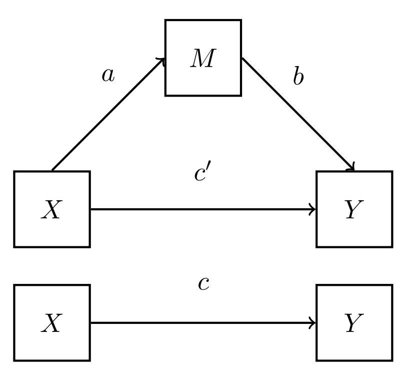
PSY 5939: Longitudinal Data Analysis

Contents

1 Review

1.1 Mediation

1.1.1 Mediation



1.1.2 Why mediation?

Design and evaluation of multi-component interventions

- Can identify which specific components are changed by the intervention and in turn affect the outcome

Better understanding of causal ordering of the variables in time

- Three time points are better than two

Mediation is a model of **process** and processes unfold **over time**, so mediation is inherently a **longitudinal** model

1.1.3 Mediation in time

Several different ways that X, M, and Y can exist in time

- Cross-sectional: $X_1 \to M_1 \to Y_1$
- Semi-longitudinal:

$X_1 \to M_1 \to Y_2$

$X_1 \to M_2 \to Y_2$

$X_1 \to M_1 \to Y_3$

$X_1 \to M_3 \to Y_3$

- Longitudinal*: $X_1 \to M_2 \to Y_3$

1.1.4 Cross-sectional estimates of longitudinal effects

Maxwell & Cole (2007) and Maxwell, Cole, & Mitchell (2011)

Cross-sectional mediation *almost always* produces **biased** estimates of longitudinal mediation effects

Cross-sectional mediation effects may be **higher** OR **lower** than the longitudinal effects

The reason for this should be fairly obvious

- How two variables are related at one time doesn't have much to do with how they are related across time

Take-home message: If you want to know about longitudinal effects, don't use cross-sectional data - use longitudinal data!

1.1.5 Mediated effect as product

The mediated effect is the effect of X on Y via M

In SEM, such a path is described as the **product** of the regression coefficients that go into it

The a coefficient reflects the X $\to$ M path

The b coefficient reflects the M $\to$ Y path

The mediated effect is a $\times$ b

1.1.6 Modern methods for mediation

MacKinnon et al. (2002), MacKinnon et al. (2004)

Joint significance: best balance of type I error and statistical power across conditions (sample size, effect size)

Product of coefficients: pretty good but difficult to actually use until PRODCLIN

Bootstrap: better confidence intervals than most other methods, requires programming skill or use of additional program, very flexible for more complex designs

Monte Carlo: better confidence intervals than most other methods, requires programming skill or use of additional program, very flexible for more complex designs

1.1.7 Effect size for mediation

Proportion mediated $= \frac{ab}{c} = 1 - \frac{c'}{c} = \frac{ab}{c'+ab}$

Ratio of mediated to direct effect $= \frac{ab}{c'}$

Standardized mediated effect $= \frac{ab}{SD_Y}$

See Miočević et al. (2018) for comparisons

- Proportion and ratio measures require **very large samples** and show **bias** in many situations
- Proportion can be **negative** if direct and indirect effects are opposite sign
- Standardized effect behaves well across sample sizes, effects

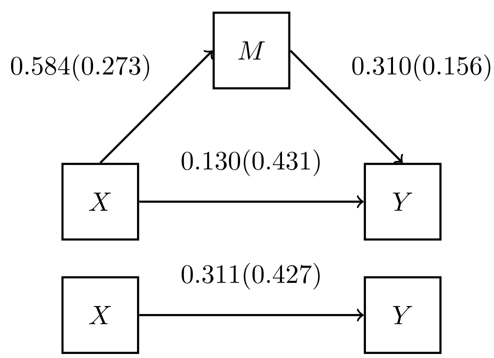
Also some R^2 measures: see Lachowicz et al. (2018)

1.1.8 References for mediation effect size

- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. Psychological methods, 16(2), 93.
- Miočević, M., O'Rourke, H. P., MacKinnon, D. P., & Brown, H. C. (2018). Statistical properties of four effect-size measures for mediation models. Behavior research methods, 50(1), 285-301.
- Lachowicz, M. J., Preacher, K. J., & Kelley, K. (2018). A novel measure of effect size for mediation analysis. Psychological Methods, 23(2), 244.

1.2 Example

1.2.1 Example data



1.2.2 Indirect effect = 0.181

$a \times b = 0.584 \times 0.310 = 0.181$

$c - c' = 0.311 - 0.130 = 0.181$

Joint significance: a is significant, b is significant

PRODCLIN via web: 95% CI = [-0.013, 0.489]

Bootstrap (lavaan): 95% CI = [-0.019, 0.488]

Monte Carlo program: 95% CI = [-0.019, 0.492]

1.2.3 Effect sizes for example

Proportion mediated $= \frac{ab}{c} = \frac{0.181}{0.311} = 0.582$

Ratio of mediated to direct effect $= \frac{ab}{c'} = \frac{0.181}{0.130} = 1.392$

Standardized mediated effect $= \frac{ab}{SD_Y} = \frac{0.181}{4.729} = 0.038$

2 Assumptions

2.1 Assumptions

2.1.1 General assumptions of mediation

1. Temporal precedence
2. Timing of change is accurately measured
3. M and Y are normally distributed*
4. No additional influences have been omitted: confounders
5. The relationships are causal

2.1.2 Temporal precedence

- X comes before M
- M comes before Y

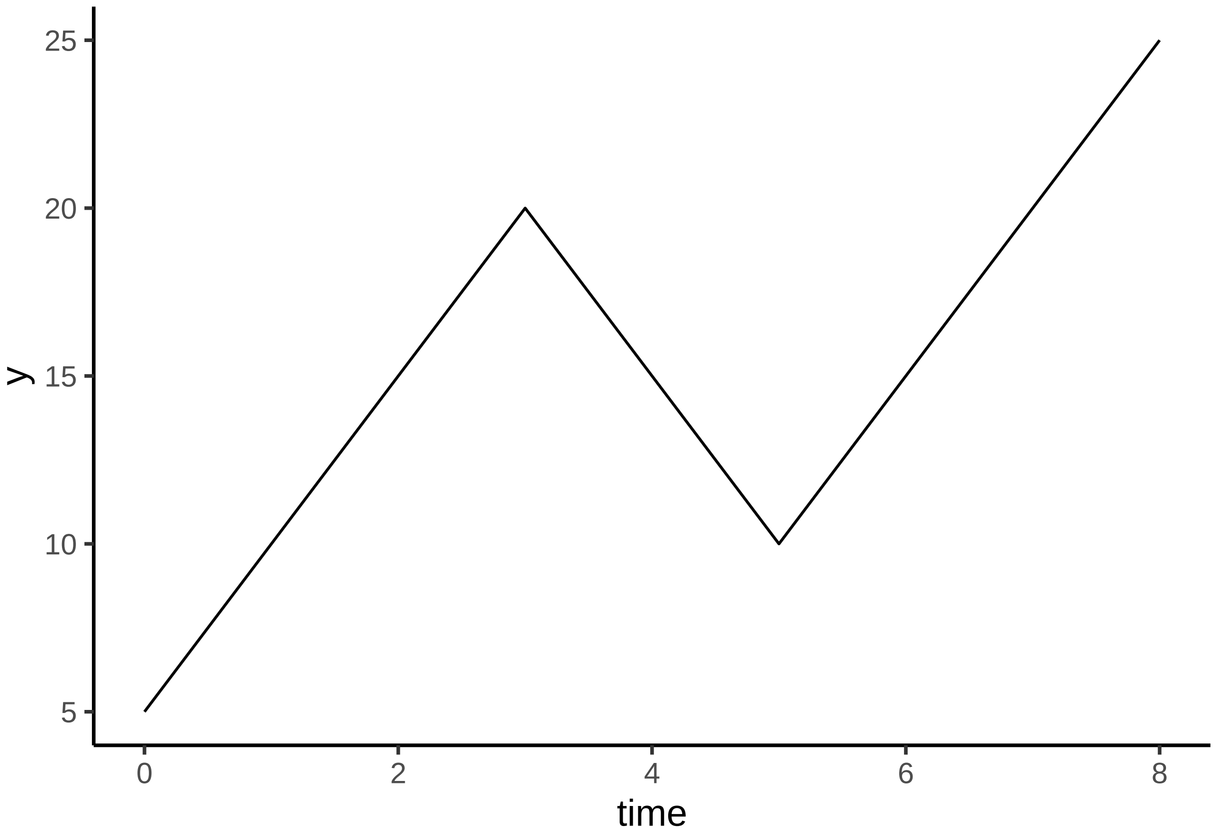
Mediation is a causal chain, so we expect that the links occur in order

2.1.3 Timing of change is accurately measured

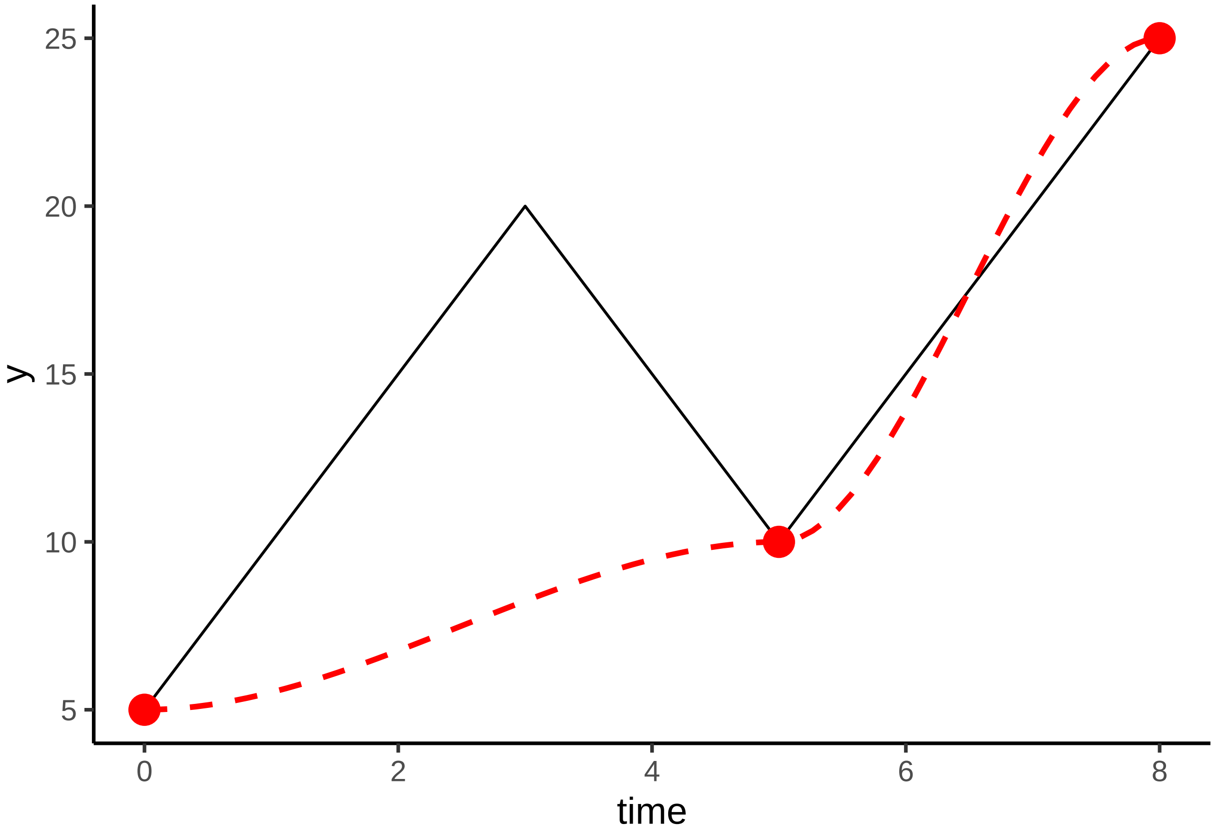
Timing of measurement is important

- Frequency
  - More rapid / fluctuating change requires more measurements
  - Nyquist criteria in EEG: sample at double the frequency to avoid aliasing
- Timing
  - Have you allowed time for change to occur?
  - But measure before change fades
- Timmons, A. C., & Preacher, K. J. (2015). The importance of temporal design: How do measurement intervals affect the accuracy and efficiency of parameter estimates in longitudinal research?. Multivariate behavioral research, 50(1), 41-55.

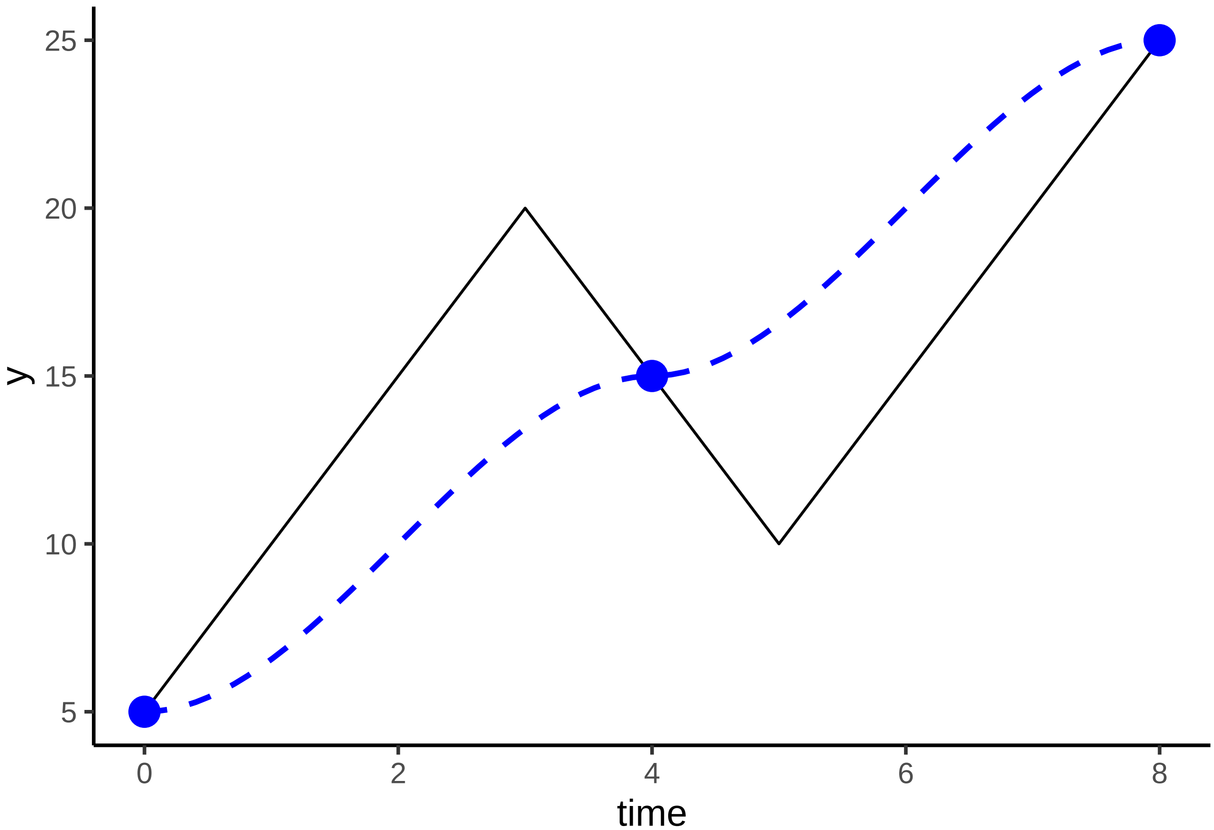
2.1.4 Timing of change is accurately measured



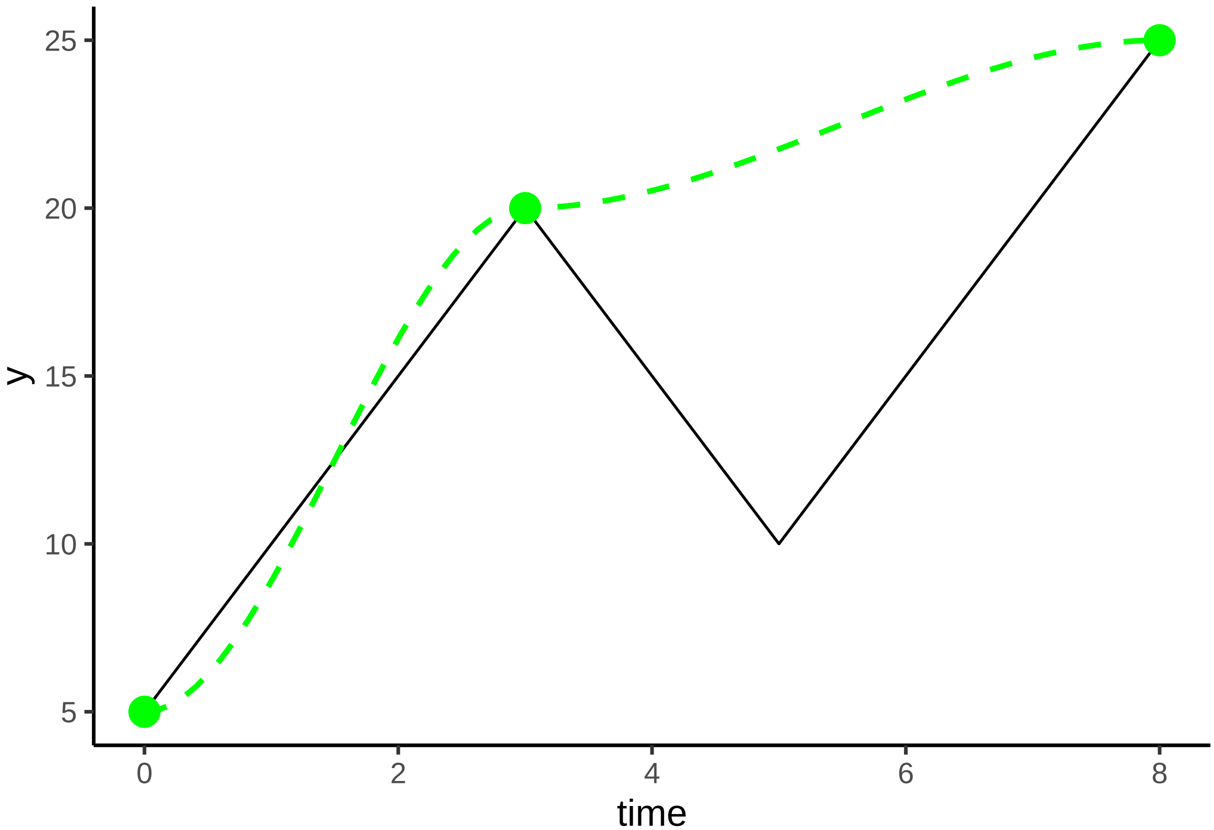
2.1.5 Timing of change is accurately measured



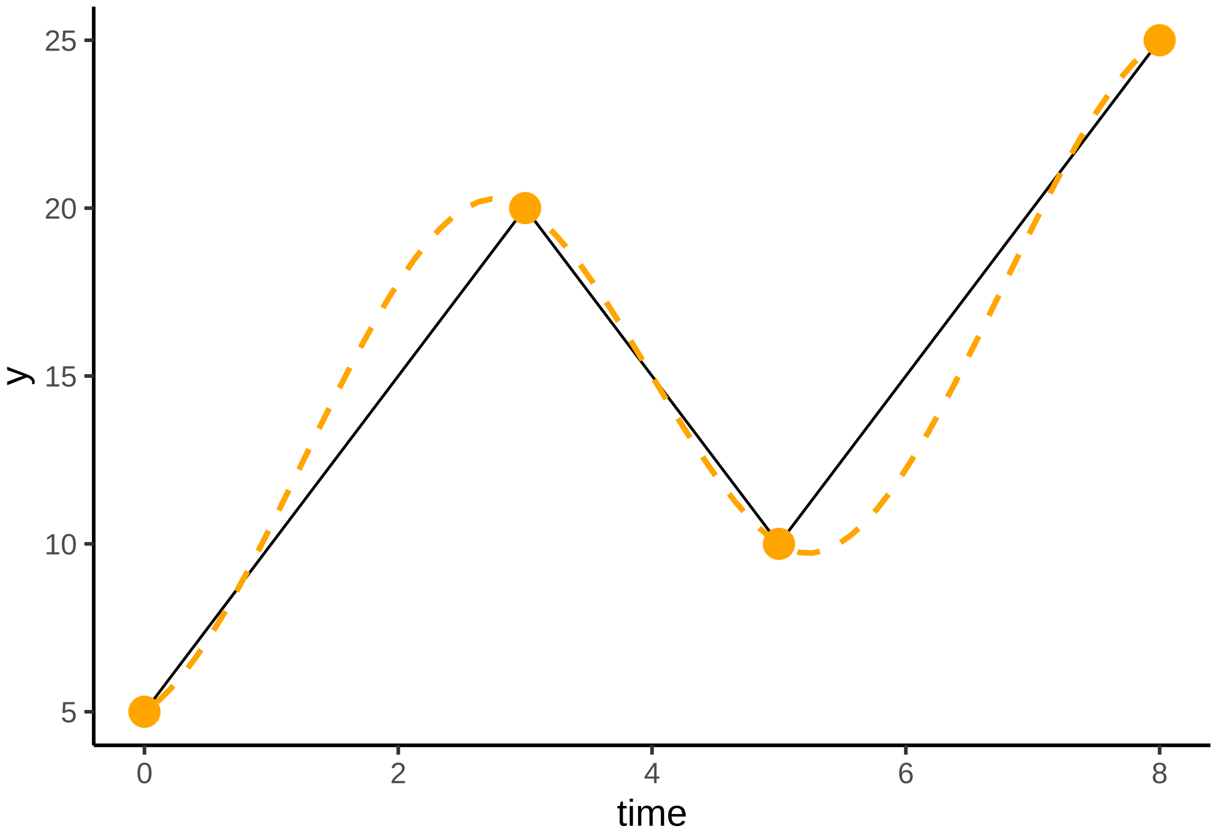
2.1.6 Timing of change is accurately measured



2.1.7 Timing of change is accurately measured



2.1.8 Timing of change is accurately measured



2.1.9 M and Y are normally distributed

This is for standard methods of mediation analysis

For binary / count outcomes, see Geldhof et al. (2018)

Causal mediation models are also more flexible with this assumption

2.1.10 No confounders have been omitted and The relationships are causal

Last two assumptions combine into "two part sequential ignorability"

- Assume that there are no additional influences affecting the relationships (i.e., no confounders of X, M, or Y), which is easy to do for **randomized** predictors
- X **can** be randomized: no confounders if it is
- M is **never** randomized, essentially turning the right half of the model into an observational study (even if X were randomized)

Can you make **causal** statements about the **mediated effect**?

- Common criticism of mediation

2.2 Causality

2.2.1 Causality and randomization

Randomization is the gold standard for establishing causality

When people are randomized to condition, there should be no differences between the groups on any measured or unmeasured variables

- Any differences between groups must be due to the manipulation
  - i.e., the manipulation **causes** the differences

But there are many situations where randomization is not *feasible* or *ethical*

- What to do then?

Two approaches: methodological and statistical

2.2.2 Causal inference in the absence of randomization

Logic of determining causation: Hill (1965)

1. Strength: stronger vs weaker relationship
2. Consistency: consistency by multiple people in multiple samples
3. Specificity: specific findings (i.e, specific disease vs general health)
4. Temporality: "cause" occurs prior to "effect"
5. Biological gradient: larger effect with larger exposure to "cause"
6. Plausibility: plausible and sensible mechanism
7. Coherence (agreement): agreement between laboratory and observational studies
8. Experiment: experimental evidence
9. Analogy: similar "causes" result in similar "effects"

2.2.3 Causal mediation

Potential outcomes framework

- Counter-factual:
  - You were assigned to condition X = 0
  - What would your Y value be if you were in condition X = 1 instead?
- Also accounts for
  - Confounders of X, M, and Y
  - XM interaction on Y: Does the b path vary depending on X?

2.2.4 Causal mediation references

- MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. Prevention Science, 21(2), 147-157.
- Rijnhart, J. J., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W., & Heymans, M. W. (2021). Mediation analysis methods used in observational research: a scoping review and recommendations. BMC medical research methodology, 21(1), 1-17.
- Valente, M. J., Rijnhart, J. J., Smyth, H. L., Muniz, F. B., & MacKinnon, D. P. (2020). Causal mediation programs in R, M plus, SAS, SPSS, and Stata. Structural equation modeling: a multidisciplinary journal, 27(6), 975-984.

Older, more technical sources

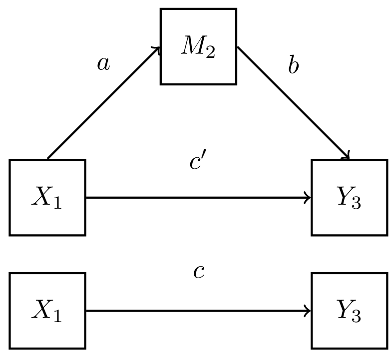
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological methods, 15(4), 309.

Judea Pearl, Tyler Vanderwheele

3 Longitudinal mediation

3.1 Longitudinal mediation

3.1.1 Mediation across 3 time points



3.1.2 From prospective to longitudinal

The last slide showed a **prospective** model

It seems like it's longitudinal because there are 3 times points

- But it's NOT because there's no **change** in any variable
- X is only at time 1, M is only at time 2, Y is only at time 3

The prospective model does not convey actual **change** in a variable over time

We can make some modifications to the model to include change

3.1.3 Prospective mediation

a path: Relationship between X1 and M2
(No change in either X or M)

b path: Relationship between M2 and Y3
(No change in either M or Y)

3.1.4 Longitudinal mediation

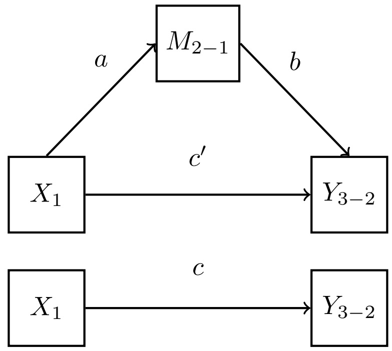
There are two simple ways to turn a prospective model into a longitudinal model

1. Difference scores
2. ANCOVA / control for earlier waves

There are also some more complex ways and methods to incorporate mediation into growth models

3.2 Difference scores

3.2.1 Mediation with difference scores



3.2.2 Still 3 equations

a path:

$$\hat{M}_{2-1} = i_{MX} + aX_1$$

b and c' paths:

$$\hat{Y}_{3-2} = i_{YXM} + bM_{2-1} + c'X_1$$

c path:

$$\hat{Y}_{3-2} = i_{YX} + cX_1$$

3.2.3 Mediation with difference scores

a path: Relationship between X1 and the **absolute change** in M from time 1 to time 2

b path: Relationship between the **absolute change** in M from time 1 to time 2 and **absolute change** in Y from time 2 to time 3

3.2.4 Mediation with difference scores

The mediated effect reflects the effect of X on the change in Y from time 2 to 3, *via the change in M from time 1 to time 2*

Similar strengths & weaknesses to the 2-wave difference score models

- Relatively easy to run and interpret

- Focus on the absolute amount of change

- Focus on individual change

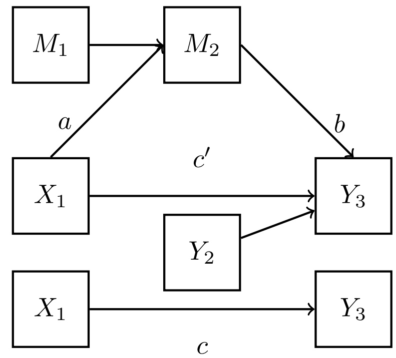
- Implies a perfect relationship at 2 times

Difference scores work well if there are few pre-test differences

- But now “pre-test” is time 1 for M and time 2 for Y

3.3 ANCOVA / lagged regression

3.3.1 Mediation with control variables



3.3.2 Still 3 equations

a path:

$$\hat{M}_2 = i_{MX} + aX_1 + dM1$$

b and c' paths:

$$\hat{Y}_3 = i_{YXM} + bM_2 + c'X_1 + eY_2$$

c path:

$$\hat{Y}_3 = i_{YX} + cX_1 + fY_2$$

3.3.3 Mediation with control variables

a path: Relationship between X1 and the **average change** in M from time 1 to time 2

b path: Relationship between the **average change** in M from time 1 to time 2 and **average change** in Y from time 2 to time 3

3.3.4 Mediation with control variables

The mediated effect reflects the effect of X on Y3 (controlling for Y2) via M2 (controlling for M1)

Similar strengths & weaknesses to the 2-wave ANCOVA

- Relatively easy to run and interpret
- Focus on average change
- Control for initial differences
- Estimates the relationship between a variable at 2 time points

3.3.5 Equating and causality

As with 2 wave models, ANCOVA works well if you have pre-test differences that make sense to equate across groups

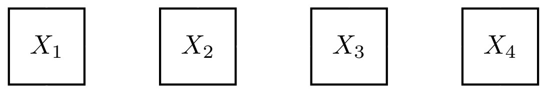
Concern:

- X can be randomized at time 1, so may make sense to equate
- M is NOT randomized, so does it make sense to equate?

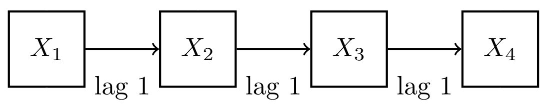
- Sequential ignorability
- Confounders

3.4 Auto-regression mediation

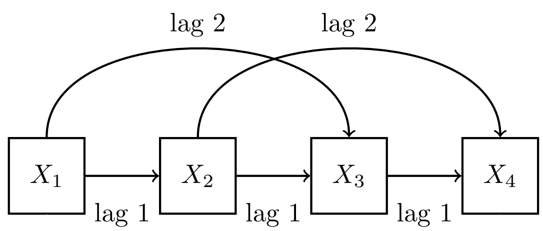
3.4.1 Auto-regression



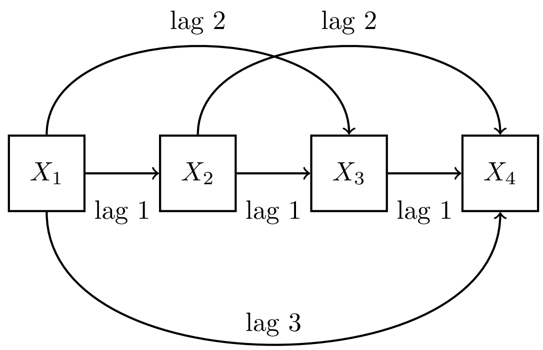
3.4.2 Auto-regression



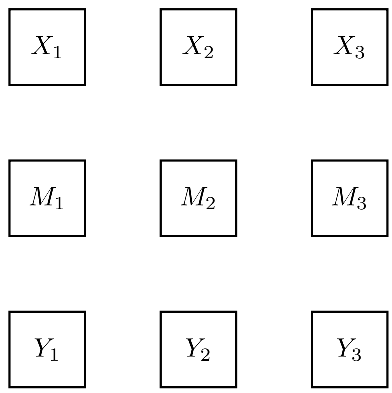
3.4.3 Auto-regression



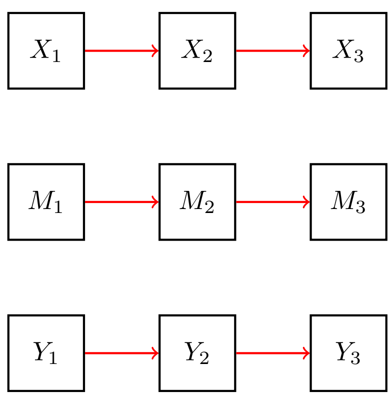
3.4.4 Auto-regression



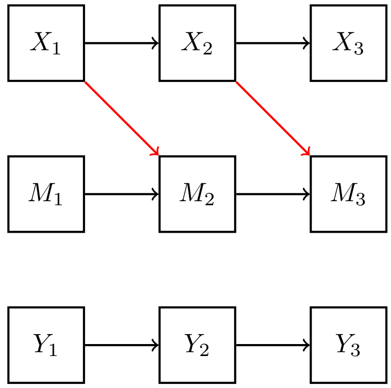
3.4.5 Auto-regression mediation: variables



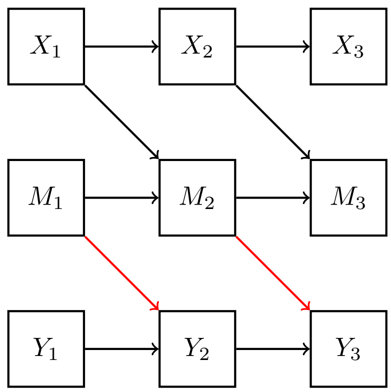
3.4.6 Auto-regression mediation: auto-regression paths



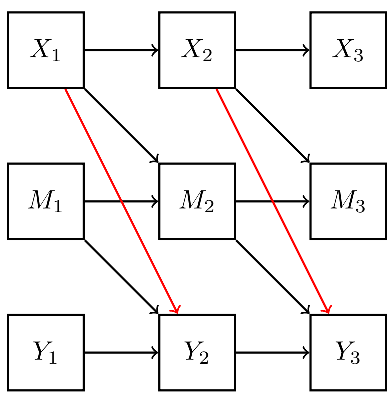
3.4.7 Auto-regression mediation: X to M



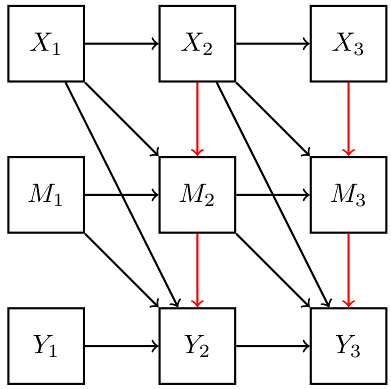
3.4.8 Auto-regression mediation: M to Y



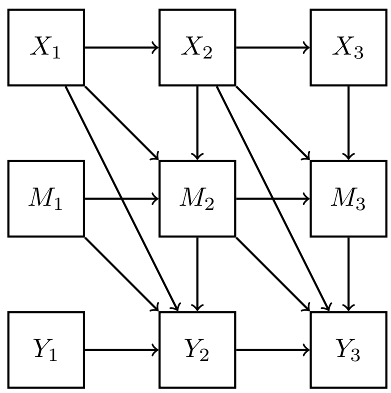
3.4.9 Auto-regression mediation: X to Y



3.4.10 Auto-regression mediation: contemporaneous



3.4.11 Auto-regression mediation



3.4.12 Auto-regression mediation

These models are very complicated, require SEM software to run

Time is treated as discrete and equally spaced

Auto-regressive models tend to focus more on the "stability" aspect of the model, while many of our research questions focus on the "change" aspect

The cross-lag relations among variables are often inaccurate

- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. Psychological methods, 20(1), 102.

3.5 Growth models and mediation

3.5.1 Growth model mediation

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. Structural Equation Modeling, 10(2), 238-262.

Basic extension of what we've alredy talked about with parallel process models with regression paths

But now we can talk about the indirect effects too

3.5.2 Growth model mediation

X = parent substance use

- Ranges from 1 to 22 (higher value = more use)

M = growth model of cigarette use (centered at age 16)

Y = growth model of alcohol use (centered at age 19)

Thinking about mediation:

- X is temporally before M – measured at same time but X is about past behavior
- M is temporally before Y – at least, the intercepts are

3.5.3 Growth model mediation

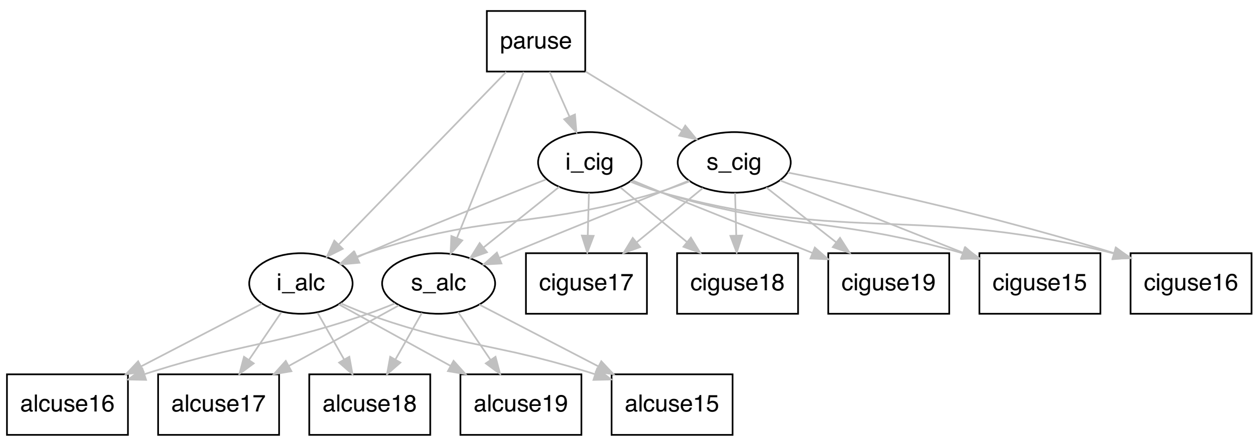
```
## lavaan 0.6-10 ended normally after 83 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        27
##   Number of equality constraints                     8
##
##   Number of observations                           749
##   Number of missing patterns                         6
##
## Model Test User Model:
##
##   Test statistic                                96.087
##   Degrees of freedom                                56
##   P-value (Chi-square)                           0.001
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Observed
##   Observed information based on                Hessian
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   i_alc =~
##     alcuse15          1.000
##     alcuse16          1.000
##     alcuse17          1.000
##     alcuse18          1.000
##     alcuse19          1.000
##   s_alc =~
##     alcuse15         -4.000
##     alcuse16         -3.000
##     alcuse17         -2.000
##     alcuse18         -1.000
##     alcuse19          0.000
##   i_cig =~
##     ciguse15          1.000
##     ciguse16          1.000
##     ciguse17          1.000
##     ciguse18          1.000
##     ciguse19          1.000
##   s_cig =~
```

```
##     ciguse15         -1.000
##     ciguse16          0.000
##     ciguse17          1.000
##     ciguse18          2.000
##     ciguse19          3.000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   i_cig ~
##     paruse    (a1)    0.080    0.014    5.711    0.000
##   s_cig ~
##     paruse    (a2)    0.009    0.007    1.227    0.220
##   i_alc ~
##     i_cig     (b1)    0.187    0.068    2.753    0.006
##     s_cig     (b2)    1.543    0.452    3.412    0.001
##   s_alc ~
##     i_cig     (b3)   -0.028    0.020   -1.412    0.158
##     s_cig     (b4)    0.508    0.144    3.515    0.000
##   i_alc ~
##     paruse   (cp1)    0.020    0.023    0.863    0.388
##   s_alc ~
##     paruse   (cp2)   -0.001    0.007   -0.092    0.927
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##  .i_alc ~~
##    .s_alc             0.047    0.039    1.206    0.228
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .alcuse15          0.000
##    .alcuse16          0.000
##    .alcuse17          0.000
##    .alcuse18          0.000
##    .alcuse19          0.000
##    .ciguse15          0.000
##    .ciguse16          0.000
##    .ciguse17          0.000
##    .ciguse18          0.000
##    .ciguse19          0.000
##    .i_alc             4.828    0.466   10.368    0.000
##    .s_alc             0.296    0.138    2.148    0.032
##    .i_cig             5.507    0.173   31.760    0.000
##    .s_cig             0.017    0.093    0.178    0.858
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .alcuse15  (r1)    0.392    0.023   17.275    0.000
##    .alcuse16  (r1)    0.392    0.023   17.275    0.000
##    .alcuse17  (r1)    0.392    0.023   17.275    0.000
##    .alcuse18  (r1)    0.392    0.023   17.275    0.000
##    .alcuse19  (r1)    0.392    0.023   17.275    0.000
##    .ciguse15  (r2)    0.402    0.025   16.267    0.000
##    .ciguse16  (r2)    0.402    0.025   16.267    0.000
```

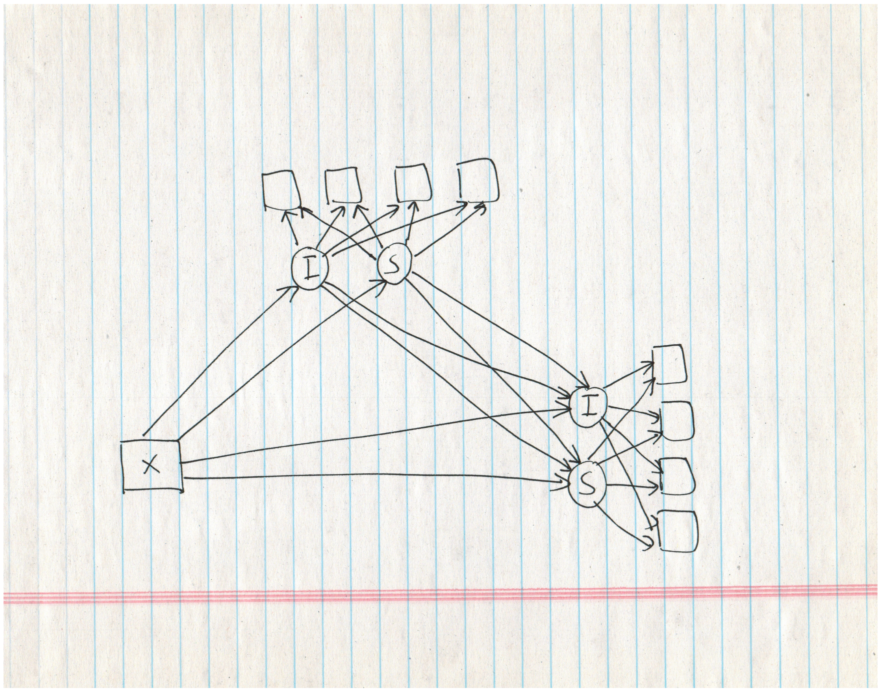
```
##    .ciguse17  (r2)    0.402    0.025   16.267    0.000
##    .ciguse18  (r2)    0.402    0.025   16.267    0.000
##    .ciguse19  (r2)    0.402    0.025   16.267    0.000
##    .i_alc             0.377    0.130    2.899    0.004
##    .s_alc             0.024    0.014    1.756    0.079
##    .i_cig             1.176    0.078   15.005    0.000
##    .s_cig             0.079    0.019    4.050    0.000
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     a1b1              0.015    0.006    2.481    0.013
##     a1b3             -0.002    0.002   -1.371    0.170
##     a2b2              0.014    0.012    1.153    0.249
##     a2b4              0.005    0.004    1.151    0.250
```

3.5.4 Growth model mediation figure

```
lavaanPlot(model = fit1,
           node_options = list(shape = "box", fontname = "Helvetica"),
           edge_options = list(color = "grey"),
           coefs = FALSE)
```



3.5.5 Growth model mediation figure



3.5.6 Growth model mediation results: direct effects

```
AICEPT   ON
   CICEPT              0.187      0.066      2.826      0.005
   CLINEAR             1.543      0.560      2.755      0.006

ALINEAR  ON
   CICEPT             -0.028      0.021     -1.358      0.174
   CLINEAR             0.508      0.176      2.884      0.004

AICEPT   ON
   PARUSE              0.020      0.025      0.773      0.439

ALINEAR  ON
   PARUSE             -0.001      0.008     -0.081      0.935

CICEPT   ON
   PARUSE              0.080      0.015      5.273      0.000

CLINEAR  ON
   PARUSE              0.009      0.008      1.110      0.267
```

3.5.7 Growth model mediation results: direct effects

Parent use to cigarette intercept = 0.080, $p<.001$

Parent use to cigarette slope = 0.009, NS

Cigarette intercept to alcohol intercept = 0.187, $p<.01$

Cigarette intercept to alcohol slope = -0.028, NS

Cigarette slope to alcohol intercept = 1.543, p<.01

Cigarette slope to alcohol slope = 0.508, p<.01

3.5.8 Growth model mediation results: indirect effects

```
Effects from PARUSE to AICEPT

  Total                 0.048      0.025      1.967      0.049
  Total indirect        0.029      0.016      1.801      0.072

  Specific indirect 1
    AICEPT
    CICEPT
    PARUSE              0.015      0.006      2.427      0.015

  Specific indirect 2
    AICEPT
    CLINEAR
    PARUSE              0.014      0.014      0.964      0.335

  Direct
    AICEPT
    PARUSE              0.020      0.025      0.773      0.439


Effects from PARUSE to ALINEAR

  Total                 0.002      0.007      0.237      0.813
  Total indirect        0.002      0.005      0.482      0.630

  Specific indirect 1
    ALINEAR
    CICEPT
    PARUSE             -0.002      0.002     -1.348      0.178

  Specific indirect 2
    ALINEAR
    CLINEAR
    PARUSE              0.005      0.005      0.971      0.332

  Direct
    ALINEAR
    PARUSE             -0.001      0.008     -0.081      0.935
```

3.5.9 Growth model mediation results: indirect effects

Parent use to cigarette intercept to alcohol intercept = 0.015, p<.05

Parent use to cigarette intercept to alcohol slope = -0.002, NS

Parent use to cigarette slope to alcohol intercept = 0.014, NS

Parent use to cigarette slope to alcohol slope = 0.005, NS

- **Summary**: parent substance use predicts alcohol use at age 19 via cigarette use at age 16

- Nothing to do with change in anything
- Slightly different ages, but similar pattern to the prospective model from last week

3.5.10 Longitudinal mediation

The standard mediation model includes X, M, and Y

But X, M, and Y can be

- observed variables
- latent variables
- growth models
- something else?

Same rules apply though:

- Try to establish temporal precedence
- Think about potential confounders
- Randomize X if you can
- Causal approaches