Statistical Analysis Using Structural Equation Models

EPsy 8266

Christopher David Desjardins

Research Methodology Consulting Center

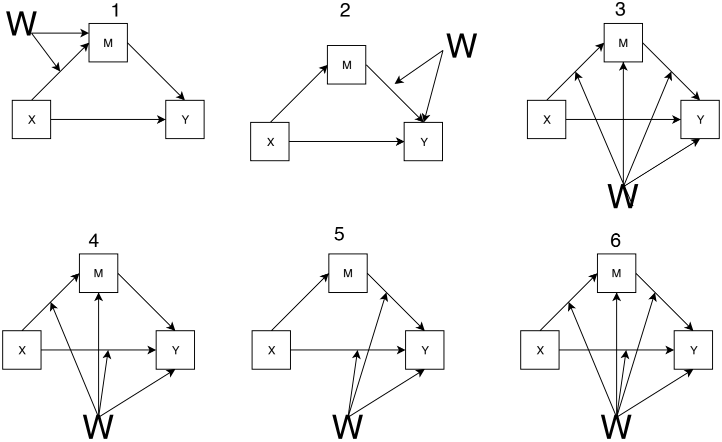
5/2/19

Conditional process modeling (vocab)

Boundary conditions of direct or indirect effects or the circumstances where causal effects occur.

- **Mediated moderation** - When the moderated effect of W is transmitted at least in part through a mediator M on Y (see #1 below)
- **Moderated mediation (conditional indirect effects)** - When the effect of X on Y via M depends on the levels of moderator W.
    1. **First-stage moderation** - When X on M depends on W
    2. **Second-stage moderation** - When M on Y depends on W
    3. **First- and second-stage moderation** - When X on M and M on Y depends on W (W could be W_1 and W_2)
    4. **Direct effect and first-stage moderation** - When X on Y and X on M depend on W.
    5. **Direct effect and second-stage moderation** - When M and X on Y depend on W.
    6. **Total effect moderation** - When everything depends on W.

Graphical depictions of moderated mediation



Latent Variable Interactions

- There is no single agreed upon approach.
- A multiple-groups approach
- Product indicator approach
- QML/LMS

Multiple-groups approach

- When we've been examining measurement invariance, we've been doing multiple-groups SEM.
- For moderation, we could fit a multiple-groups model and fix all the parameters except the structural parameter of interest.
  - Compare the model where this path is freed across the groups to one where it is fixed.
- Works with **observed predictor-observed moderator** and **latent predictor-observed moderator**.
- For dichotomous moderator, multiple-group model is natural.
- For continuous moderator, must discretize.
  - Where?

Example in lavaan with HolzingerSwineford

HZ consists of mental ability test scores of 7th and 8th grade student. Does sex is moderate the relationship of visual perception (visual) on paragraph comprehension (paragrap).

$$
\begin{aligned}
&\text{Paragrap}|\text{Visual} \sim N(\beta_0 + \beta_{1_m}\text{Visual}, \sigma^2) \\
&\text{Paragrap}|\text{Visual} \sim N(\beta_0 + \beta_{1_w}\text{Visual}, \sigma^2) \\
&H_0 : \beta_{1_m} = \beta_{1_w} \\
&H_1 : \beta_{1_m} \neq \beta_{1_w}
\end{aligned}
$$

```
library(lavaan)
hz <- read.csv("https://bit.ly/2PEDFre")
mod <- "
paragrap ~ 1 + visual
"
fit.free <- sem(mod, hz, group = "gender",
           group.equal = c("intercepts", "residuals", "means", "lv.variances"))
fit.fixed <- sem(mod, hz, group = "gender",
           group.equal = c("intercepts", "residuals", "regressions"))
anova(fit.fixed, fit.free)
```

```
## Chi Square Difference Test
##
##           Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit.free   2 1560.1 1574.9 0.1969
## fit.fixed  3 1567.0 1578.2 9.1408     8.9439       1   0.002784 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare to MR

```
mod <- lm(paragrap ~ 1 + visual:gender, data = hz)
coef(mod)
```

```
##         (Intercept) visual:genderFemale   visual:genderMale
##           3.5463489           0.2083482           0.1719075
```

```
parameterEstimates(fit.free)
```

```
##         lhs op      rhs block group label    est    se      z pvalue ci.lower ci.upper
## 1  paragrap ~1               1     1  .p1.  3.546 0.801  4.429      0    1.977    5.116
## 2  paragrap  ~   visual     1     1        0.172 0.027  6.444      0    0.120    0.224
## 7  paragrap  ~   visual     2     2        0.208 0.027  7.627      0    0.155    0.262
## 8  paragrap ~~ paragrap     2     2  .p3. 10.161 0.828 12.268      0    8.538   11.784
```

With an observed moderator, latent predictor, and a latent DV

Let's say we want to see if sex moderates the relationship between the visual factor on the textual factor. This is essentially the same null hypotheses but we have to add some more variables constraints.

```
lv.mod <- '
  vis  =~ visual + cubes + paper
  text =~ paragrap + sentence + wordm
  text ~ vis
'
fit.free.lv <- sem(lv.mod, hz, group = "gender",
           group.equal = c("loadings","intercepts", "residuals"))
fit.fixed.lv <- sem(lv.mod, hz, group = "gender",
           group.equal = c("loadings","intercepts", "residuals", "regressions"))
anova(fit.free.lv, fit.fixed.lv)
```

```
## Chi Square Difference Test
##
##              Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit.free.lv  30 10226 10315 41.068
## fit.fixed.lv 31 10224 10309 41.069  0.0010295       1     0.9744
```

Bonus Q: What did I just assume about the visual and textual factors?

With an observed moderator, latent predictor, and a latent DV

Let's say we want to see if sex moderates the relationship between the visual factor on the textual factor controlling for age.

```
lv.mod.cov <- '
  vis  =~ visual + cubes + paper
  text =~ paragrap + sentence + wordm
  text ~ vis + c(b1, b1)*agey
'
fit.free.lv.cov <- sem(lv.mod.cov, hz, group = "gender",
          group.equal = c("loadings","intercepts", "residuals"))
fit.fixed.lv.cov <- sem(lv.mod.cov, hz, group = "gender",
          group.equal = c("loadings","intercepts", "residuals", "regressions"))
anova(fit.free.lv.cov, fit.fixed.lv.cov)
```

```
## Chi Square Difference Test
##
##                  Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit.free.lv.cov  41 10214 10306 54.449
## fit.fixed.lv.cov 42 10212 10301 54.458  0.0092393       1     0.9234
```

Product indicator approach

- The **indicant product approach** in SEM involves creating product terms that are specified as indicators of a latent product variable(s).
- We'll assume the indicators are continuous and we have two factors, A and B, that each have two indicators.

$$
\begin{aligned}
X_1 &= A + E_{X_1} \\
X_2 &= \lambda_2 A + E_{x_2} \\
W_1 &= B + E_{W_1} \\
W_2 &= \lambda_4 B + E_{W_2}
\end{aligned}
$$

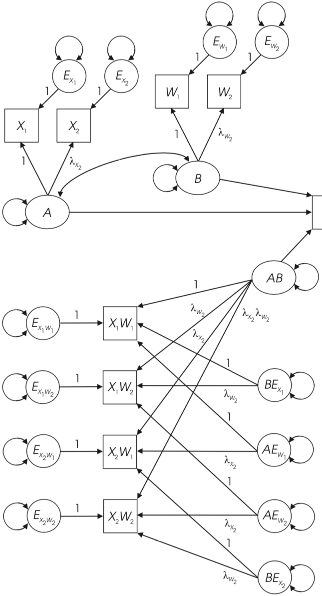
- There are two free unstandardized pattern coefficients, 4 residual variances, 2 LV variances, and 1 LV covariance

Latent product variable

- The latent product variable, AB, represents the interactive effect of A and B.
- The indicators for AB are the four product indicators X_1W_1, X_1W_2, X_2W_1, and X_2W_2.

$$
\begin{aligned}
X_1W_1 &= AB + BE_{X_1} + AE_{W_1} + E_{X_1}E_{W_1} \\
X_1W_2 &= \lambda_4 AB + \lambda_4 BE_{X_1} + AE_{X_1} + E_{X_1}E_{X_2} \\
X_2W_1 &= \lambda_2 AB + BE_{X_2} + \lambda_2 AE_{W_1} + E_{X_2}E_{W_1} \\
X_2W_2 &= \lambda_2\lambda_4 AB + \lambda_4 BE_{X_2} + \lambda_2 AE_{W_2} + E_{X_2}E_{W_2}
\end{aligned}
$$

- This requires an additional 8 more latent product variables.
- All pattern coefficients are either 1 or functions of the coefficients for X_2 and W_2 and the other parameters are the variances/covariances of the implied LVs.
- Assuming normality of the nonproduct LVs and that the nonproduct indicators are centered 1) covariance amoung the latent product variables and nonproduct LVs are 0 and 2) variances of the latent product variables can be expressed as functions of the variance of the nonproduct LVs.



E_{X_1}
E_{X_2}
E_{W_1}
E_{W_2}
X_1
X_2
W_1
W_2
1
λ_{X_2}
λ_{W_2}
A
B
D_Y
Y
AB
$\lambda_{X_2}\lambda_{W_2}$
$E_{X_1W_1}$
$E_{X_1W_2}$
$E_{X_2W_1}$
$E_{X_2W_2}$
X_1W_1
X_1W_2
X_2W_1
X_2W_2
BE_{X_1}
AE_{W_1}
AE_{W_2}
BE_{X_2}

Kenny-Judd approach (traditional constrained approach)

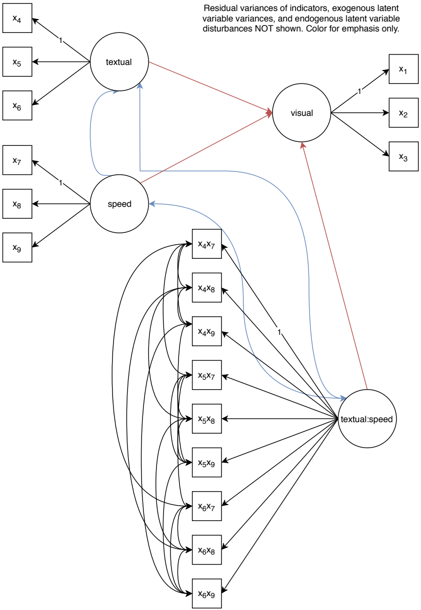
- This is the Kenny-Judd approach.
- It can only be done in software that allows nonlinear constraints (lavaan doesn't).
- Kline's book website has Mplus code for this approach.

Product of indicators with lavaan (unconstrained approach)

- There are two broads options for using a product of indicators approach with lavaan that don't require nonlinear constraints.
- Both approaches involves creating product of indicators that are **double-mean centered** (Lin, Wen, Marsh, & Lin, 2010)
  - $X_1X_2 = (X_1 - \bar{X_1})(X_2 - \bar{X_2}) - \overline{(X_1 - \bar{X_1})(X_2 - \bar{X_2})}$
  - Use the `indProd` in **semTools** package to make it easier.

1. Use all the product indicators (unconstrained approach) .
   - Should correlate the residuals of the constituent parts of the product indicators.
   - Disadvantage: Need a larger sample size cause lots of parameters are getting estimated!
2. Use the matched-pairs approach (Marsh, Wen, & Hau, 2006)
   - Create pairs such that no two indicators are used more than once.
   - Don't need correlated residuals

Example with the HZ dataset

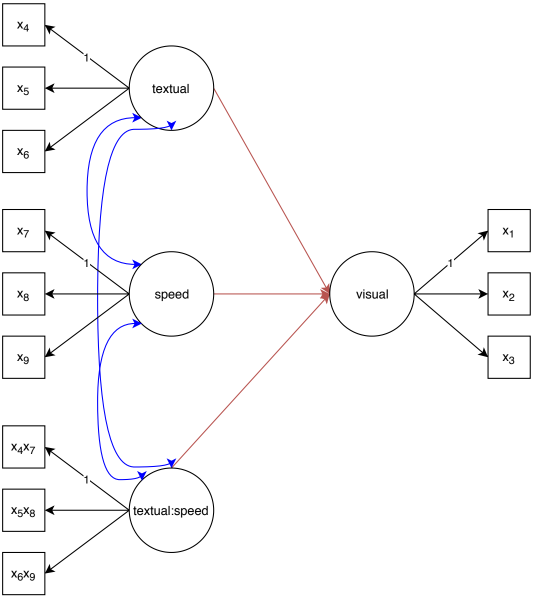


All product indicators in lavaan

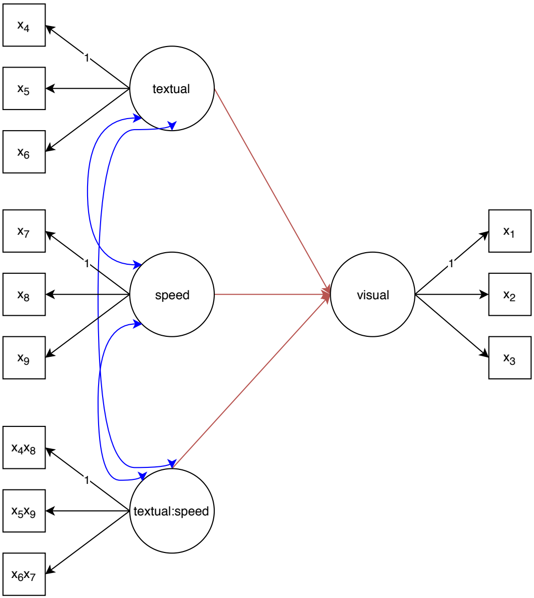
In our model, we want to predict performance on verbal ability given textual and speed abilities and their interaction. This is maybe an uninteresting example, but it illustrates how to do this.

```
library(semTools)
hz <- subset(HolzingerSwineford1939, select = x1:x9)
hz <- indProd(hz, var1 = 4:6, var2 = 7:9, match = FALSE)

all.mod <- "
visual  =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed   =~ x7 + x8 + x9
textspeed =~ x4.x7 + x4.x8 + x4.x9 + x5.x7 + x5.x8 + x5.x9 + x6.x7 + x6.x8 + x6.x9

x4.x7 ~~ x4.x8 + x4.x9 + x5.x7 + x6.x7
x4.x8 ~~ x4.x9 + x5.x8 + x6.x8
x4.x9 ~~ x5.x9 + x6.x9

x5.x7 ~~ x5.x8 + x5.x9 + x6.x7
x5.x8 ~~ x5.x9 + x6.x8
x5.x9 ~~ x6.x9

x6.x7 ~~ x6.x8 + x6.x9
x6.x8 ~~ x6.x9

visual ~ speed + textual + int*textspeed
"
fit.all <- sem(all.mod, hz)
```

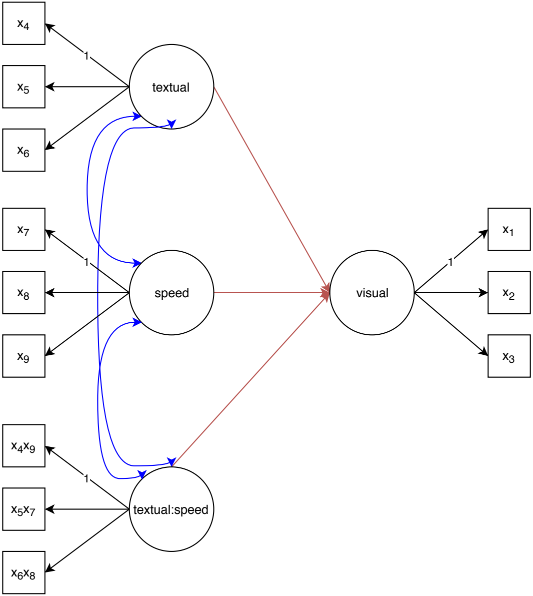
Matched-Pairs #1



Matched-Pairs #2



Matched-Pairs #3



Matched-pairs approach

3 (possible) matched-pairs

```
matched1.mod <- "
visual  =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed   =~ x7 + x8 + x9
textspeed =~ x4.x7 + x5.x8 + x6.x9
visual ~ speed + textual + int*textspeed
"
matched2.mod <- "
visual  =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed   =~ x7 + x8 + x9
textspeed =~ x4.x8 + x5.x9 + x6.x7
visual ~ speed + textual + int*textspeed
"
matched3.mod <- "
visual  =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed   =~ x7 + x8 + x9
textspeed =~ x4.x9 + x5.x7 + x6.x8
visual ~ speed + textual + int*textspeed
"
fit.matched1 <- sem(matched1.mod, hz)
fit.matched2 <- sem(matched2.mod, hz)
fit.matched3 <- sem(matched3.mod, hz)
```

LMS and QML

- Alternative approaches involve using latent moderated structural equations (LMS) (Klein & Moosbrugger, 2000).
- LMS doesn't create product indicators and requires only normality for your nonproduct LVs and product LVs can depart from normality.
- A computational simpler version is known as QML (Klein & Muthen (2007) estimation and it approximates LMS.
- Relies on numerical integration and is available in **nlsem**.
  - CAUTION: **nlsem** has not been updated in two years!

Example with nlsem

```r
library(nlsem)
lav.mod <- "
visual =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed =~ x7 + x8 + x9

visual ~ textual + speed + textual:speed
"
model.int <- lav2nlsem(lav.mod)
model.int
```

```
## Model of class singleClass
##
## Number of latent endogenous variables: 1 (with 3 indicators)
## Number of latent exogenous variables: 2 (with 6 indicators)
##
## Structural model:
## ------------------------------------------------------------
## eta1 = xi1 + xi2
## eta1 = xi1:xi2
## ------------------------------------------------------------
##
## Measurement model:
## ------------------------------------------------------------
## xi1 = x1 + x2 + x3
## xi2 = x4 + x5 + x6
## eta1 = y1 + y2 + y3
## ------------------------------------------------------------
```

```r
set.seed(5135)
hz <- HolzingerSwineford1939
hz <- subset(hz, select = c(x4:x9, x1:x3))
start <- runif(count_free_parameters(model.int))
nsem.int <- qml(model = model.int, data = hz, start = start)
```

```r
summary(nsem.int)
```

QML in MPlus

```
TITLE: Test moderation
DATA: File is holz.dat;
VARIABLE: Names are id sex ageyr agemo school grade x1 x2 x3
          x4 x5 x6 x7 x8 x9;
USEVARIABLES ARE x1 x2 x3 x4 x5 x6 x7 x8 x9;

ANALYSIS: TYPE = RANDOM;
ALGORITHM = INTEGRATION;

MODEL:
textual BY x4 x5 x6;
speed BY x7 x8 x9;
visual BY x1 x2 x3;

textxspeed | textual XWITH speed;
visual ON textual speed textxspeed;

OUTPUT:
cinterval;
```

Summary of results

Method	Speed	Textual	Speed:Textual
All indicators (lavaan)	0.529*	0.313*	0.107
Matched-pairs 1 (lavaan)	0.533*	0.308*	0.104
Matched-pairs 2 (lavaan)	0.528*	0.321*	0.048
Matched-pairs 3 (lavaan)	0.532*	0.308*	0.141
QML (nlsem)	0.287	-0.053	0.087
QML (Mplus)	0.533*	0.311*	0.094

I would recommend either lavaan or Mplus and DO NOT recommend nlsem.