STATSU webinar:
Path Analysis using Mplus lab session

RME—Research, Measurement, and Evaluation

June 4, 2020
Welcome! Mplus lab session: path analysis

“If a model is consistent with reality, then the data should be consistent with the model. But if the data are consistent with the model, this does not imply the model corresponds to reality”

-K. A. Bollen
Outline

1. Resources
2. Path analysis
3. Mplus basics
4. Importing data
5. DATA & VARIABLE
6. MODEL
7. Examples
Resources

University of Miami virtual lab for Mplus:
https://vlabs.it.miami.edu/

Text editor
https://www.sublimetext.com/

Mplus website:
https://www.statmodel.com/

Model identification:
http://davidakenny.net/cm/identify_formal.htm
Resources

Code provided for SEM and MLM framework: Mplus, R, SAS
Theory

A influences, or causes B

$A \rightarrow B$

X and Y covary

$X \leftrightarrow Y$

Sample size:

5 to 20 observations per estimated parameter is needed, on average.

All models require justified theory and remember...

this method is disconfirmatory!

Exogenous/gives causal input = A

Endogenous/receives causal input = B
### Notation

<table>
<thead>
<tr>
<th><strong>Diagram Symbol</strong></th>
<th><strong>Notation</strong></th>
<th><strong>Mplus</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifest variable:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>directly observed, and measured</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Exogenous variable = X</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Endogenous variable = Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Latent variable: | xi | F |
| directly observed, and measured | |
| (not used in path analysis) | eta |

| Structural effects coefficients: | beta | MODEL RESULTS values |
| Slope (factor loadings in SEM) | lambda |
| Numeric value or $\beta_{11}$ or $\lambda_{11} \ldots \lambda_{22}$ | |
| first subscript = DV | |
| second subscript = IV | |

| Manifest variable error (also called residual): | epsilon | E |
| Error term for observed variables | |
| $\varepsilon$, $\varepsilon_1$, $\varepsilon_2$, $\varepsilon_3$,... | |

| Latent variable error (also called residual or disturbance): | delta | D |
| Error term for observed variables | |
| $\delta$ | |

| | zeta | |
| | |
Path analysis example

Calorie intake (kcal) \rightarrow Z = .20 \rightarrow Exercise (time)

X = .55 \rightarrow Exercise (time) \rightarrow Y = .60 \rightarrow Weight (lbs.)

<table>
<thead>
<tr>
<th>Path variables</th>
<th>Direct effect</th>
<th>Indirect effect</th>
<th>Spurious associations</th>
<th>Total Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV1 IV2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV1 DV</td>
<td>X</td>
<td></td>
<td>ZY</td>
<td>X + ZY</td>
</tr>
<tr>
<td>IV2 DV</td>
<td>Y</td>
<td></td>
<td>ZX</td>
<td>Y + ZX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Path variables</th>
<th>Direct effect</th>
<th>Indirect effect</th>
<th>Spurious associations</th>
<th>Total Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV1 DV</td>
<td>.55</td>
<td></td>
<td>.67</td>
<td>.67</td>
</tr>
<tr>
<td>IV2 DV</td>
<td>.60</td>
<td></td>
<td>.71</td>
<td>.71</td>
</tr>
<tr>
<td>IV1 IV2</td>
<td>.60</td>
<td></td>
<td>.20</td>
<td>.71</td>
</tr>
</tbody>
</table>
Model fit

Joint criteria is recommended
e.g., CFI \geq .96 & SRMR \leq .10

1 Loglikelihood: compares nested models (deviance)
   \( H_0 \) Value: null model \( p < .05 \)

2 Information Criteria: non-nested models
   AIC, BIC, SBIC: \textit{smallest value}

3 RMSEA and SRMR
   value \(< .05\)

4 CFI/TLI: (comparative fit index \& Tucker-Lewis Index\(^1\))
   value \(> .95\)
   literature can support values above .90 as acceptable

5 \(\chi^2\) test for baseline
   Baseline is a null model \( p < .05 \)

\(^1\)NNFI is another name for TLI
1. Text editor (Sublime Text/Vim/notepad...) SPSS or R/SAS/Julia
   *Link*: R package: MPlusAutomation

2. .dat or .csv text files only

3. Comment Mplus code: ! <code-not-run> ;

4. Mplus will not read past 80 characters
   *Link*: good v. bad coding

5. Primary commands are color coded in **BLUE**, lines end with a ;

6. Code will run for the entire file

7. Missing data: accepts empty cells. -9999 is defined as missing

8. Mplus files are .inp for the syntax and .out for results
SPSS convert to Mplus-friendly .csv file:

File > Save as>
select *.csv file

save as type box

uncheck: “write variable names to spreadsheet” or
DATA: ! required;
VARIABLE: ! required;
MODEL:
OUTPUT:

! OTHER commands;

TITLE:
DEFINE:
SAVEDATA:
ANALYSIS:
1. Viewing data and data tidying in alternative text editor/program.

2. Labels and names restricted to 8 characters

3. Every variable in data set must be initialized

4. **TITLE**: Print text at the beginning of your .out file
Data type to be processed: tab-, space-, comma-delimited text

.dat files with:

white space delimiter: 100 200
comma-delimiter: 100,200

Windows: Shift key then right click on any file and select "copy as path name"

Mac OS: CTL then right click on any file, press OPT: "copy as path name"
VARIABLE NAMES ARE var1 var2 var3;
VARIABLE NAMES ARE var1-var3;

USEVARIABLES = <list what will be used in the model command>;

VARIABLE NAMES ARE var1-var9;
USEVARIABLES = var2-5 var8;
VARIABLE
CATEGORICAL = <ranked/ordinal>

NOMINAL = <no ranks>

CENSORED =
right-censoring (from values above):
food scale at market is 20lbs. – placing an item weighing 30 lbs.
will read 20 lbs. Only know the maximum value.

left-censoring (from values below):
medical tests have set sensitivities, antibodies below the threshold
are false-negative results. Only floor values known.
Validate import!

**OUTPUT:** SAMPSTAT
Means, variances/covariances for continuous variables

Frequency table for categorical data.
DEFINE:

```plaintext
new = <existing_var_name> + 100;

new = IF <existing_var_name> EQ 100
    THEN <new_var_name> = 1;
```
! ON defines change/influence/slope;

MODEL:   Y ON X1; ! X1 impacts the value of Y;

! endogenous/dep. var. = exogenous/indep. var.;

! WITH defines covariance;

MODEL:   X1 WITH X2; ! X1 and X2 are associated;
Mplus Example: 3.11
UM virtual lab: https://vlabs.it.miami.edu/

Mplus examples website for dataset:
https://www.statmodel.com/usersguide/chapter3.shtml

1. Right click data and save data: “ex3.11.dat”
2. Write syntax (or download “ex3.11.inp”) in same location with ex3.11.dat; then click run to produce output.

```
TITLE: Example of a path analysis with continuous dependent variables
DATA: FILE IS ex3.11.dat;
VARIABLE: NAMES ARE y1-y3 x1-x3;
MODEL: y1 y2 ON x1 x2 x3;
y3 ON y1 y2 x2;
```

Can also click Diagram from menu to produce your structural figure.
Questions?

Thank you!