An Introduction to Structural Equation Modeling and OpenMx

Michael D. Hunter

Human Development and Family Studies
The Pennsylvania State University

March 4, 2024

Plan

- What is an SEM?
- Picture book time: Making and understanding diagrams
- What OpenMx's deal?
- More on matrices and paths
- Time to actually do something with this
- An OpenMx cheat sheet of functions
- Fitting more realistic models is Hermine's job

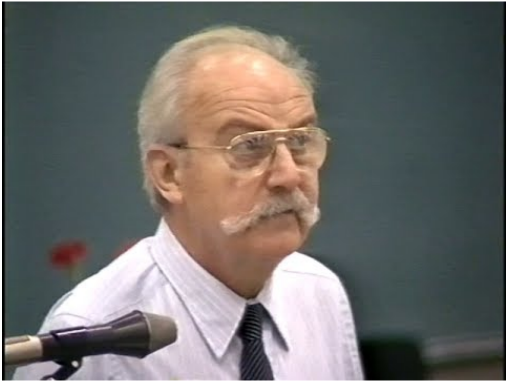
PennState

What is a structural equation model?

What is a ~~structural equation~~ model?

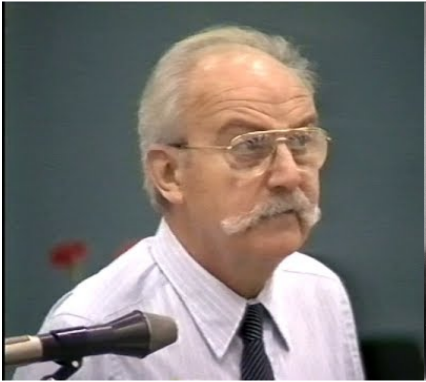
Models

All models are wrong, but some are useful.
–George E.P. Box



Models

All models are wrong, but some are useful.
–George E.P. Box



Models


The Map is Not the Territory


1
Vaughan Metropolitan Centre
Highway 407
Pioneer Village
York University
Finch West
Downsview Park
Sheppard West
Wilson
Yorkdale
Lawrence West
Glencairn
Eglinton West
Pearson Airport
900 Airport Express
St Clair West
Dupont
Spadina
St George
2 Kipling
Islington
Royal York
Old Mill
Jane
Runnymede
High Park
Keele
Dundas West
Lansdowne
Dufferin
Ossington
Christie
Bathurst
Museum
Queen's Park
St Patrick
Osgoode
St Andrew
Union
Bay
Bloor-Yonge
Sherbourne
Wellesley
College
Dundas
Queen
King
Summerhill
Rosedale
St Clair
Davisville
Eglinton
Lawrence
York Mills
4 Sheppard-Yonge
North York Centre
Finch
Bayview
Bessarion
Leslie
Don Mills 4
Castle Frank
Broadview
Chester
Pape
Donlands
Greenwood
Coxwell
Woodbine
N

Models

Picasso: “Art is a lie that tells the truth.”
Hunter: “~~Art~~ A Model is a lie that tells the truth.”

SEM by the names and programs

- LInear Structural RELations (LISREL)
- Structural Equation Modeling (SEM)
- COvariance Structure ANalysis (COSAN)
- Latent variable Analysis with dichotomous, ordered Categorical, and Continuous Indicators (LACCI)
- LInear Structural COMPonents (LISCOMP, M*plus*)
- linear EQuationS (EQS)
- Reticular Action Model (RAM)
- LAtent VAriable ANalysis (lavaan)
- Classic Mx (1991–2008)
- OpenMx (2008–)

SEM by the students 1

SEM is . . .

- A network of linear regressions with both observed and latent variables

SEM by the students 1

SEM is . . .

- A network of linear regressions with both observed and latent variables
- A model that represents the relationship between multiple variables, both latent and observed, through related regressions

SEM by the students 1

SEM is . . .

- A network of linear regressions with both observed and latent variables
- A model that represents the relationship between multiple variables, both latent and observed, through related regressions
- A network of regressions among latent and observed variables

SEM by the students 1

SEM is . . .

- A network of linear regressions with both observed and latent variables
- A model that represents the relationship between multiple variables, both latent and observed, through related regressions
- A network of regressions among latent and observed variables
- A model that explains the relationship between variables – whether latent or observed – using a covariance matrix

SEM by the students 1

SEM is . . .

- A network of linear regressions with both observed and latent variables
- A model that represents the relationship between multiple variables, both latent and observed, through related regressions
- A network of regressions among latent and observed variables
- A model that explains the relationship between variables – whether latent or observed – using a covariance matrix
- A regression model for latent and observed variables

SEM by the students 1

SEM is ...

- A network of linear regressions with both observed and latent variables
- A model that represents the relationship between multiple variables, both latent and observed, through related regressions
- A network of regressions among latent and observed variables
- A model that explains the relationship between variables – whether latent or observed – using a covariance matrix
- A regression model for latent and observed variables
- A method of modeling the means, variances, and covariances of latent and observed variables

SEM by the students 2

SEM is . . .

- A modeling framework for variances, covariances, and means

SEM by the students 2

SEM is . . .

- A modeling framework for variances, covariances, and means
- A set of equations explaining the path diagram between latent and observed variables including means, variances, and covariances

SEM by the students 2

SEM is . . .

- A modeling framework for variances, covariances, and means
- A set of equations explaining the path diagram between latent and observed variables including means, variances, and covariances
- The use of latent and observed variables to simplify the world and its patterns of data

SEM by the students 2

SEM is . . .

- A modeling framework for variances, covariances, and means
- A set of equations explaining the path diagram between latent and observed variables including means, variances, and covariances
- The use of latent and observed variables to simplify the world and its patterns of data
- A compact way of doing many related regression models simultaneously

SEM by the students 2

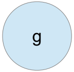
SEM is ...

- A modeling framework for variances, covariances, and means
- A set of equations explaining the path diagram between latent and observed variables including means, variances, and covariances
- The use of latent and observed variables to simplify the world and its patterns of data
- A compact way of doing many related regression models simultaneously
- A model with structured relations between latent and observed variables based on reparameterizing the covariance matrix

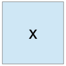
My definition

SEM is a systematic way of representing the joint distribution of a set of (possibly latent) variables in terms of structurally related free parameters that correspond to a set of graphing/diagramming rules.

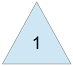
Diagrams



A latent variable g

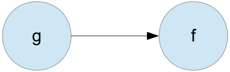


A manifest variable x

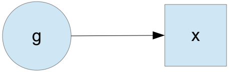


A constant (i.e., means)

Regression from x to y

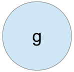


Regression from g to f

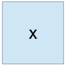


Regression from g to x

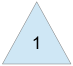
Diagrams



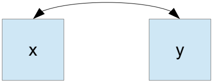
A latent variable g



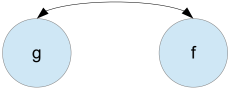
A manifest variable x



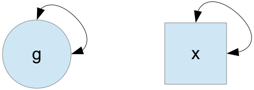
A constant (i.e., means)



Covariance between x and y

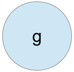


Covariance between g and f

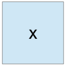


Variance of g; Variance of x

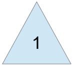
Diagrams



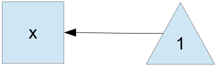
A latent variable g



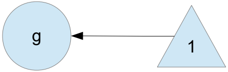
A manifest variable x



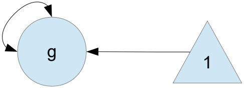
A constant (i.e., means)



Mean of x

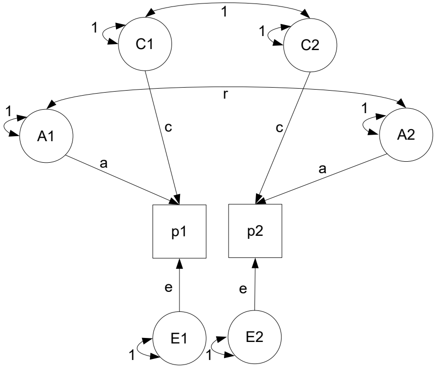


Mean of g

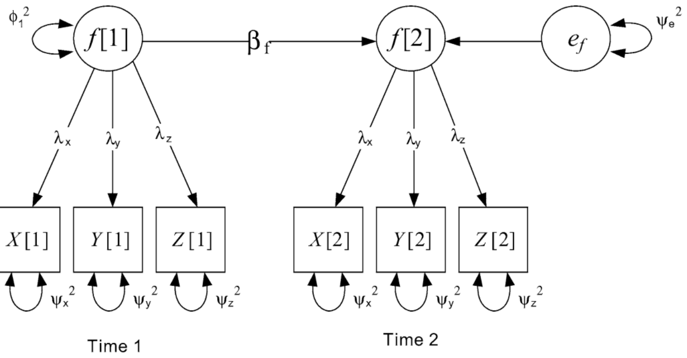


Mean and variance of g

Examples

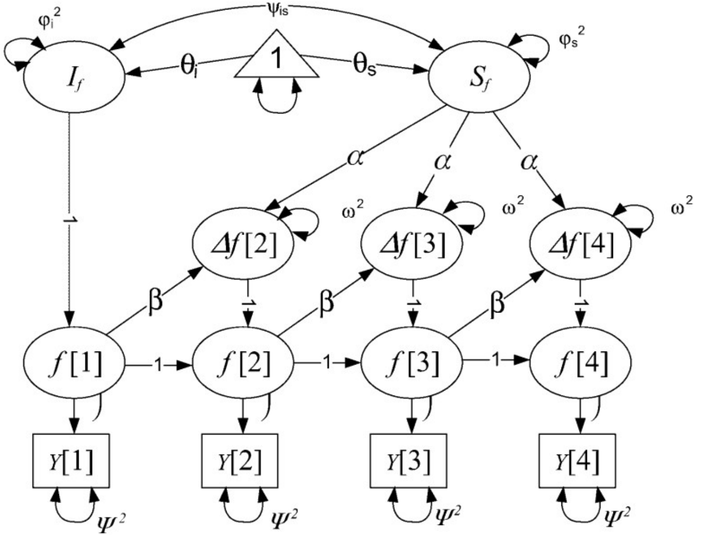


Examples



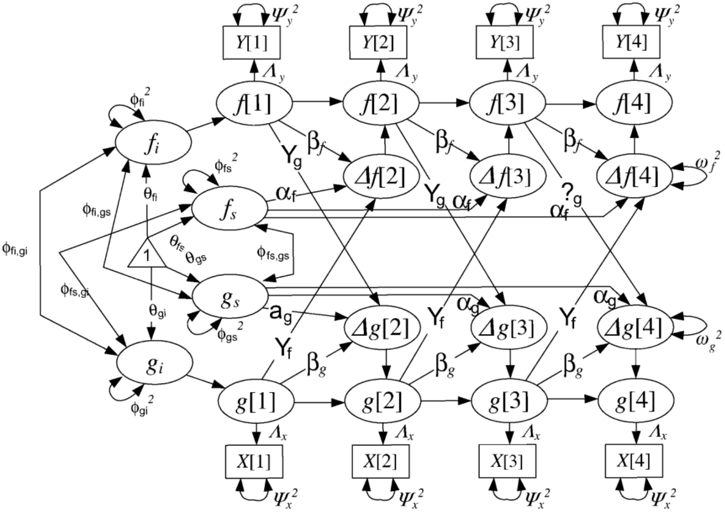
ϕ_1^2
$f[1]$
β_f
$f[2]$
e_f
ψ_e^2
λ_x
λ_y
λ_z
λ_x
λ_y
λ_z
$X[1]$
$Y[1]$
$Z[1]$
$X[2]$
$Y[2]$
$Z[2]$
ψ_x^2
ψ_y^2
ψ_z^2
ψ_x^2
ψ_y^2
ψ_z^2
Time 1
Time 2

Examples



ψis
φi2
φs2
1
θi
θs
If
Sf
α
α
α
ω2
ω2
ω2
Δf[2]
Δf[3]
Δf[4]
β
β
β
1
1
1
1
1
1
f[1]
f[2]
f[3]
f[4]
Y[1]
Y[2]
Y[3]
Y[4]
Ψ2
Ψ2
Ψ2
Ψ2

Examples


Y[1]
Y[2]
Y[3]
Y[4]
f[1]
f[2]
f[3]
f[4]
Δf[2]
Δf[3]
Δf[4]
fi
fs
gs
gi
1
Δg[2]
Δg[3]
Δg[4]
g[1]
g[2]
g[3]
g[4]
X[1]
X[2]
X[3]
X[4]

OpenMx

The SEM program you develop when you are serious about

- R
- Open-Source
- Object oriented programming
- Modular software development
- Model scriptability
- Extensibility
- User freedom
- What you say is what it does (WYSIWID)

OpenMx

Huh! What is it good for?

- Specifying the exact model YOU want
- Forcing you to make your model explicit
- Fitting models the developers had not thought of
- Making multivariate models with complicated relations
- Fitting a series of nested models in a single script

In short, fitting SEMs like it's 2024, not 1985


WHAT YEAR IS IT?!

What does a path do?

Creates a relation between variables

```
mxPath(from, to, connect, arrows, free,
       values, labels)
```

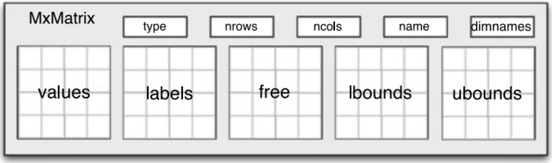
Paths create elements of matrices

What does a matrix do?

Map from a column to a row

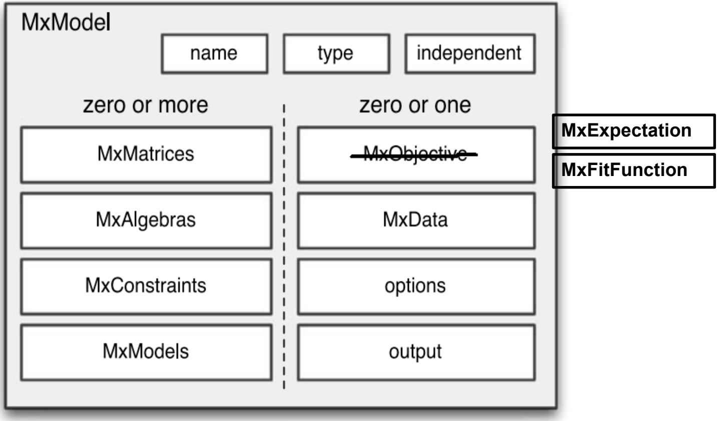
Matrices are the set of all linear functions

```
mxMatrix(type, nrow, ncol, free, values, labels,
         lbound, ubound)
```



What does a model do?

A model is a container



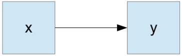
Let's get practical, practical

Let's get into practical

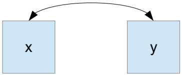
Mapping

lavaan	M*plus*	OpenMx
`g =~ x`	g BY x	`mxPath(from='g', to='x', arrows=1)`
`y ~ x`	y ON x	`mxPath(from='x', to='y', arrows=1)`
`x ~~ y`	x WITH y	`mxPath(from='x', to='y', arrows=2)`
`x ~ 1`	[x]	`mxPath(from='one', to='x', arrows=1`

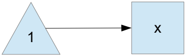
Regression from g to x



Regression from x to y



Covariance between x and y

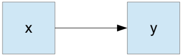


Mean of x

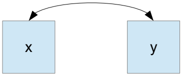
Mapping

lavaan	M*plus*	OpenMx
`g =~ x`	g BY x	`mxPath('g', 'x')`
`y ~ x`	y ON x	`mxPath('x', 'y')`
`x ~~ y`	x WITH y	`mxPath('x', 'y', arrows=2)`
`x ~ 1`	[x]	`mxPath('one', 'x')`

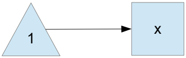
Regression from g to x



Regression from x to y



Covariance between x and y



Mean of x

Programmatic

OpenMx

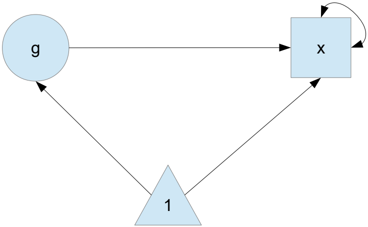
```
xs <- paste0('x', 1:9)
mxPath('g', xs)
```

lavaan

```
g =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
```

Exercise 1

Write paths that go with this diagram

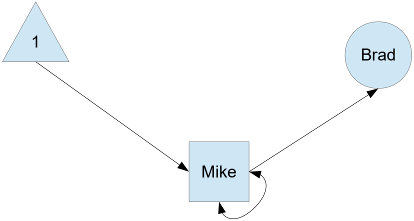


Exercise 2

Draw a diagram of this model `require(OpenMx)`

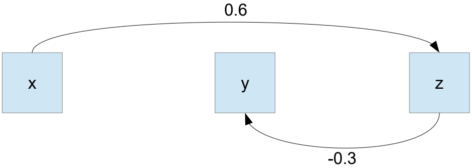
```
amod <- mxModel('Exercise 2',
   type='RAM',
   manifestVars='Mike',
   latentVars='Brad',
   mxPath(from='one', to='Mike', arrows=1,
      values=1.4, labels='m'),
   mxPath(from='Mike', arrows=2, values=.5,
      labels='v'),
   mxPath(from='Mike', to='Brad', arrows=1,
      values=2, labels='r')
)
```

Exercise 2



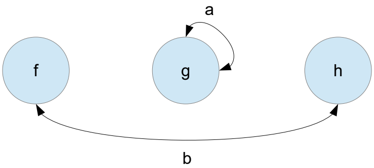
Exercise 3

Write a matrix that represents this diagram



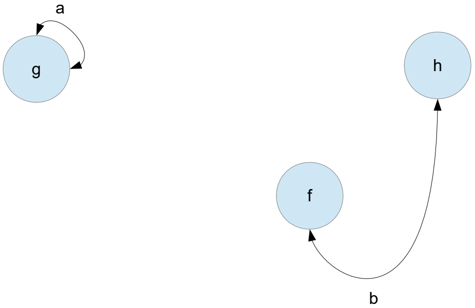
Exercise 4

Write a matrix that represents this diagram



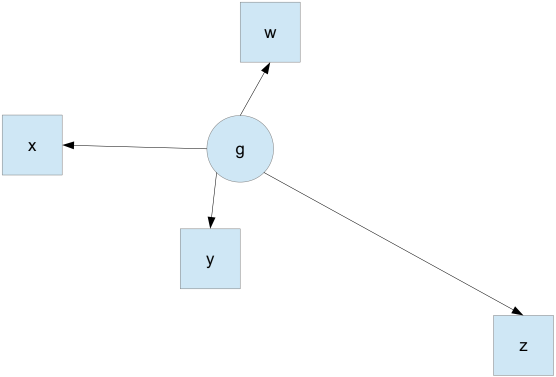
Exercise 5

Write a matrix that represents this diagram



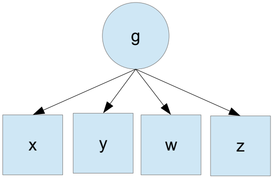
Exercise 6

Write a model that represents this diagram



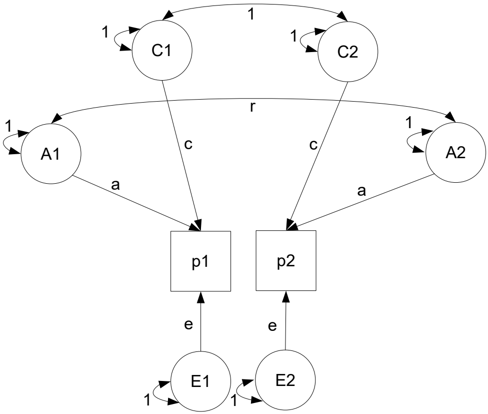
Exercise 7

Write a model that represents this diagram



Exercise 8

Write a matrix for the covariances of A1 & A2.
Write a matrix for the covariances of C1 & C2.



Useful OpenMx Functions

Tier 1

Stuff you'll use all the time	
mxModel()	Create a model
mxPath()	Create a path between variables
mxMatrix()	Create a matrix
mxRun()	Run a model to estimate parameters
mxCompare()	Compare nested models
mxFitFunction*()	Set of functions for model fits
mxExpectation*()	Set of functions for model types
summary()	Summarize model results
coef()	Get the estimated parameters
anova()	Likelihood ratio testing

Useful OpenMx Functions

Tier 2

Stuff you'll use often	
mxTryHard()	Refit a model with jittered starting values
mxCI()	Make profile likelihood confidence intervals
mxAlgebra()	Algebraically combine matrices/algebras
omxRunCI()	Get CIs without refitting the model
omxSetParameters()	Change parameter values

Useful OpenMx Functions

Tier 3

Some useful others	
mxSave()	Save your model results
mxRestore()	Restore your saved model results
mxEval()	Evaluate a matrix or algebra
mxGetExpected()	Obtain expected means, covariances, etc.
mxSE()	Get standard errors
mxCheckIdentification()	Is your model identified?
mxGenerateData()	Simulate data according to your model
logLik()	Get the log likelihood from a model
AIC()	Get the AIC from a model
confint()	Get Wald-type confidence intervals

PennState

What have we learned?

- SEM is . . .
- Diagramming conventions for SEM
- OpenMx design philosophy
- What matrices and paths do
- Translating between matrices, paths, diagrams, and code
- OpenMx can do some cool things

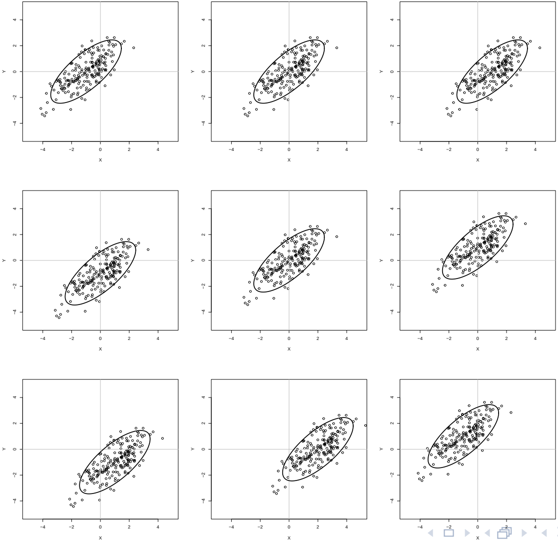
What have we not learned

- How to fit more realistic and interesting models

Thank You

mdh282@psu.edu

Means: X=Anxiety; Y=BMI



Variance and Covariance: X=Anxiety; Y=BMI