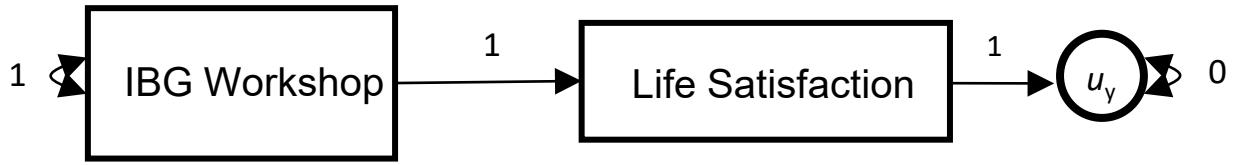
University of Colorado **Boulder**

Institute for Behavioral Genetics

Short Primer on Structural Equation Modeling (SEM) in Lavaan

Andrew Grotzinger

IBG Workshop 2021



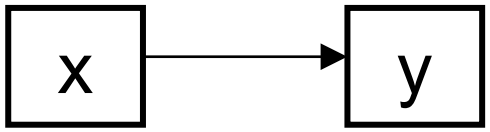
How to specify a model In Lavaan

Regression:

y ~ x

-or-

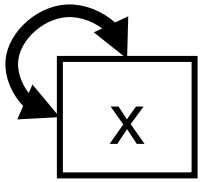
A ~ B ; Dependent ~ Independent; Outcome ~ Predictor



How to specify a model In Lavaan

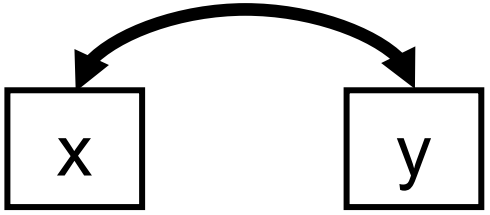
Variance:

x ~~ x



Covariance:

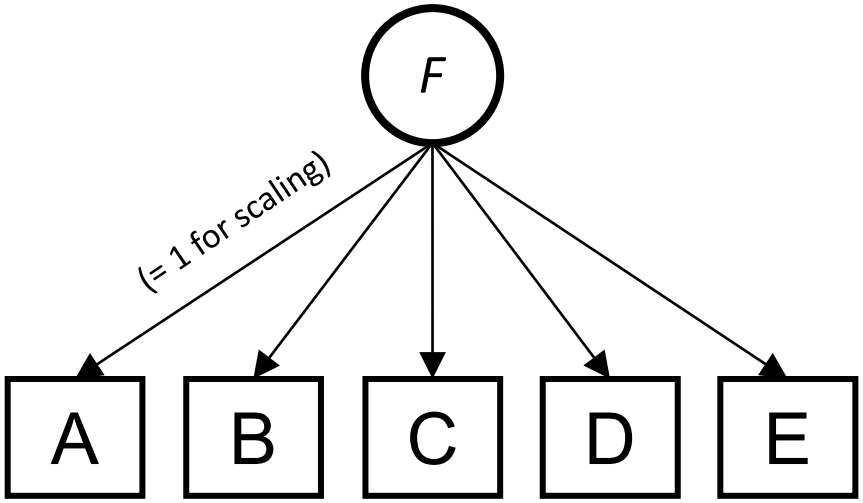
x ~~ y



How to specify a model In Lavaan

Factor Loadings:

F =~ A + B + C + D + E



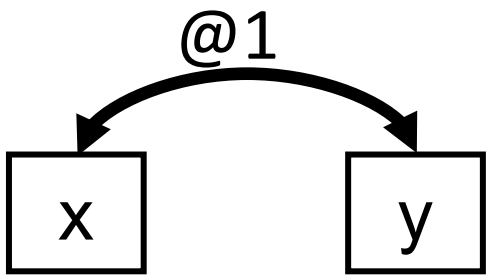
Squares are used to depict *observed* variables. These are variables that are present in your dataset, such as A –E

Circles are used to depict unobserved, or latent, variables. This includes latent factors, like F

How to specify a model In Lavaan

Fix a parameter:

x ~~ 1*y (the covariance between x and y is 1)



How to specify a model In Lavaan

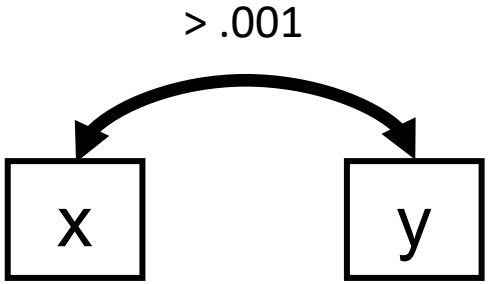
Name a parameter:

x ~~ a*y

(the covariance between x and y = parameter label a)

Allows you to use model constraints for this parameter:

a > .001



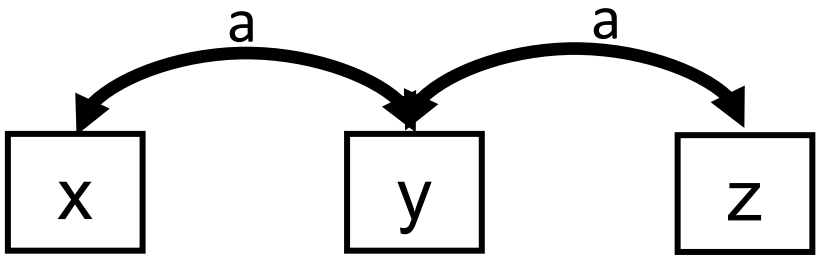
How to specify a model In Lavaan

Name a parameter (a cautionary note):

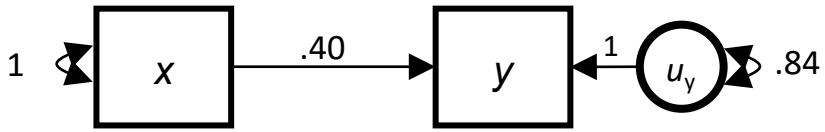
x ~~ a*y

y ~~ a*z

(the covariance between x and y and the covariance between y and z are the same)

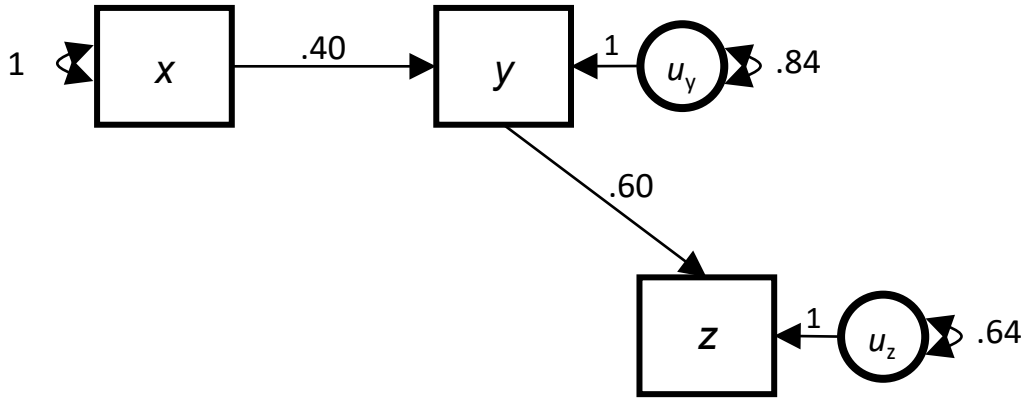


Imagine we knew the generating causal process



$y = .40\,x + u_y$

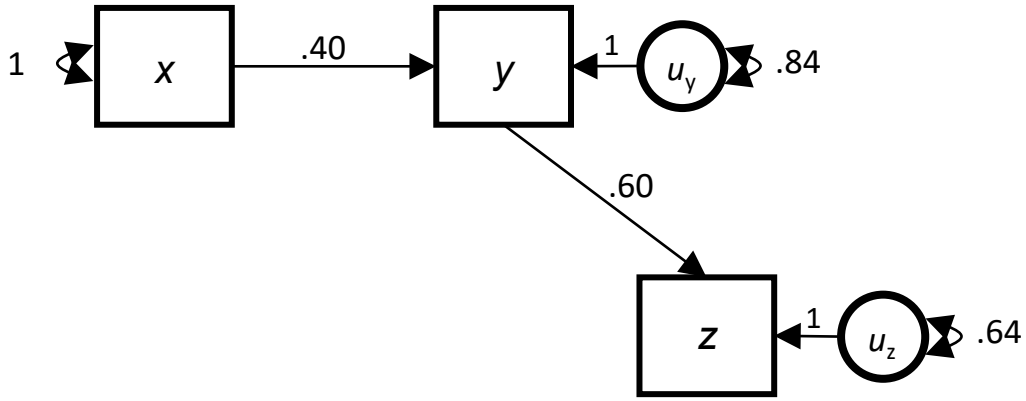
Imagine we knew the generating causal process



$y = .40\,x + u_y$ $\quad\quad$ $x \sim (0,1)$, $u_y \sim (0,.84)$

$z = .60\,y + u_z$ $\quad\quad$ $u_z \sim (0,.64)$

Imagine we knew the generating causal process



Implied covariance matrix *in the population*

$cov(x,y,z)_{pop}$ =

1.00		
.40	1.00	
.24	.60	1.00

$y = .40\,x + u_y$ $\quad\quad$ $x \sim (0,1)$, $u_y \sim (0,.84)$

$z = .60\,y + u_z$ $\quad\quad$ $u_z \sim (0,.64)$

In practice, we only observe the sample data, and we propose a model

observed covariance matrix
in a sample

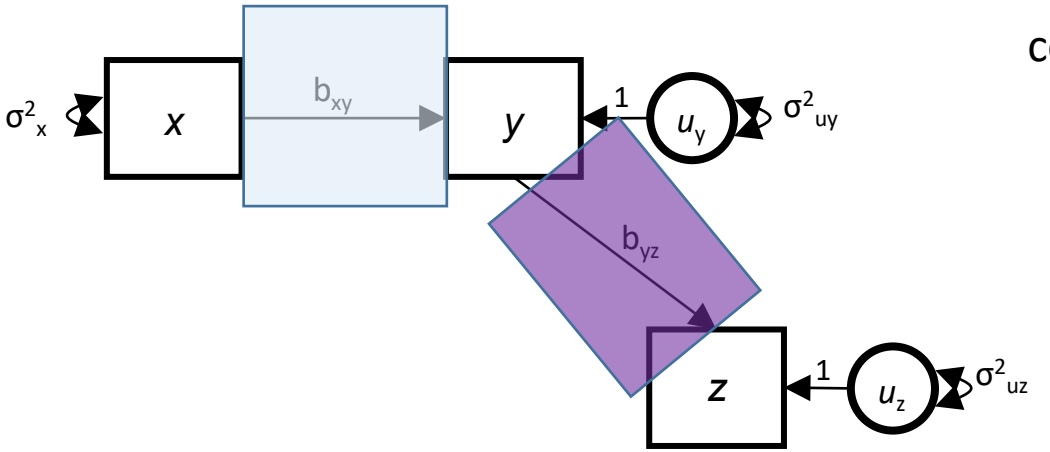
.94		
.33	1.02	
.27	.62	1.02

≈

covariance matrix
in population

1.00		
.40	1.00	
.24	.60	1.00

For the proposed model, estimate parameters from the data, and evaluate model fit to the data



$cov(x,y,z)_{sample}$ =

.94		
.33	1.02	
.27	.62	1.02

In lavaan syntax:

y~x

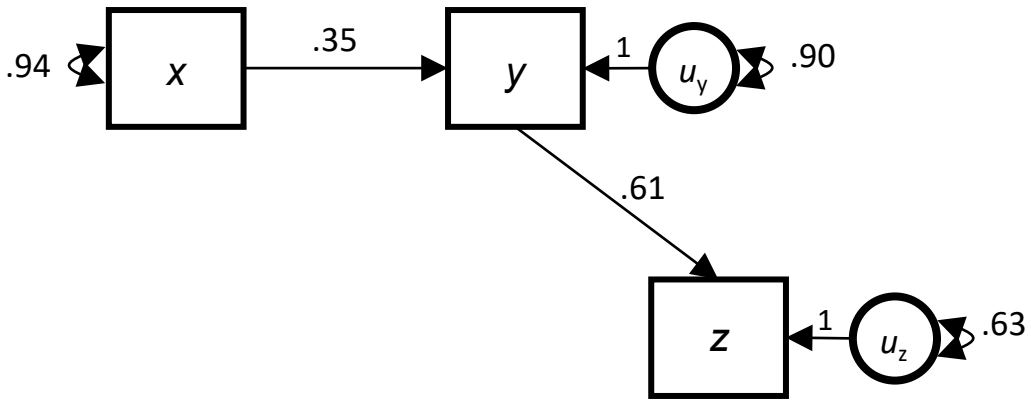
z~y

6 unique elements in the covariance matrix being modeled

5 free model parameters

1 degree of freedom (df)

For the proposed model, estimate parameters from the data, and evaluate model fit to the data



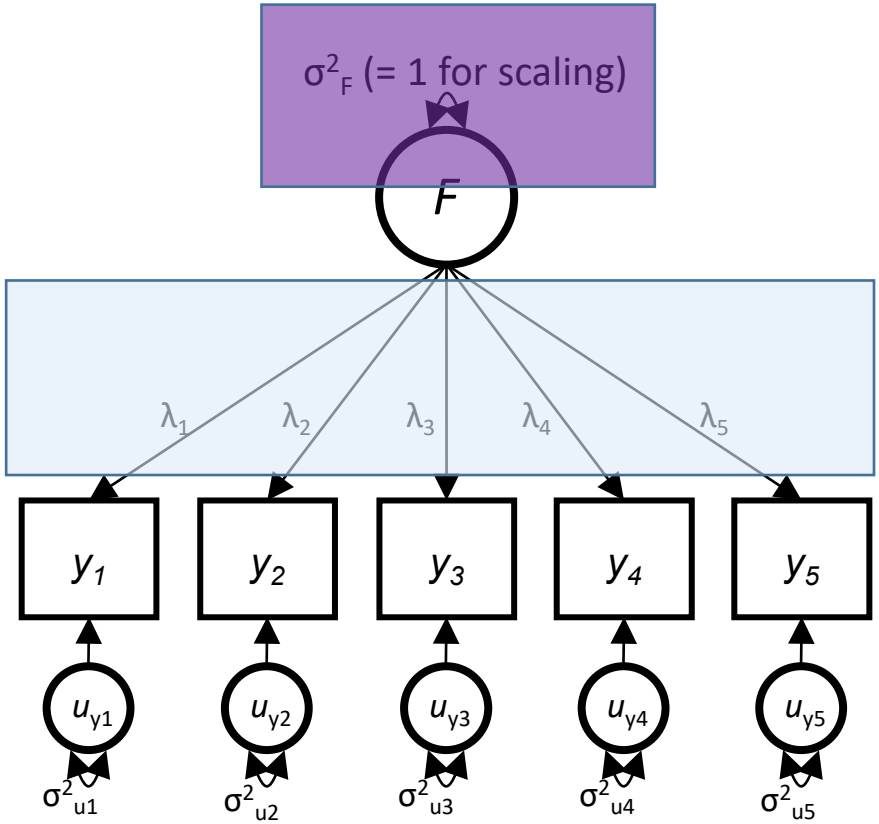
$\text{cov}(x,y,z)_{\text{sample}}$ =

.94		
.33	1.02	
.27	.62	1.02

$\text{cov}(x,y,z)_{\text{implied}}$ =

.94		
.33	1.03	
.20	.63	1.00

The model that we fit may include some variables for which we do not observe data

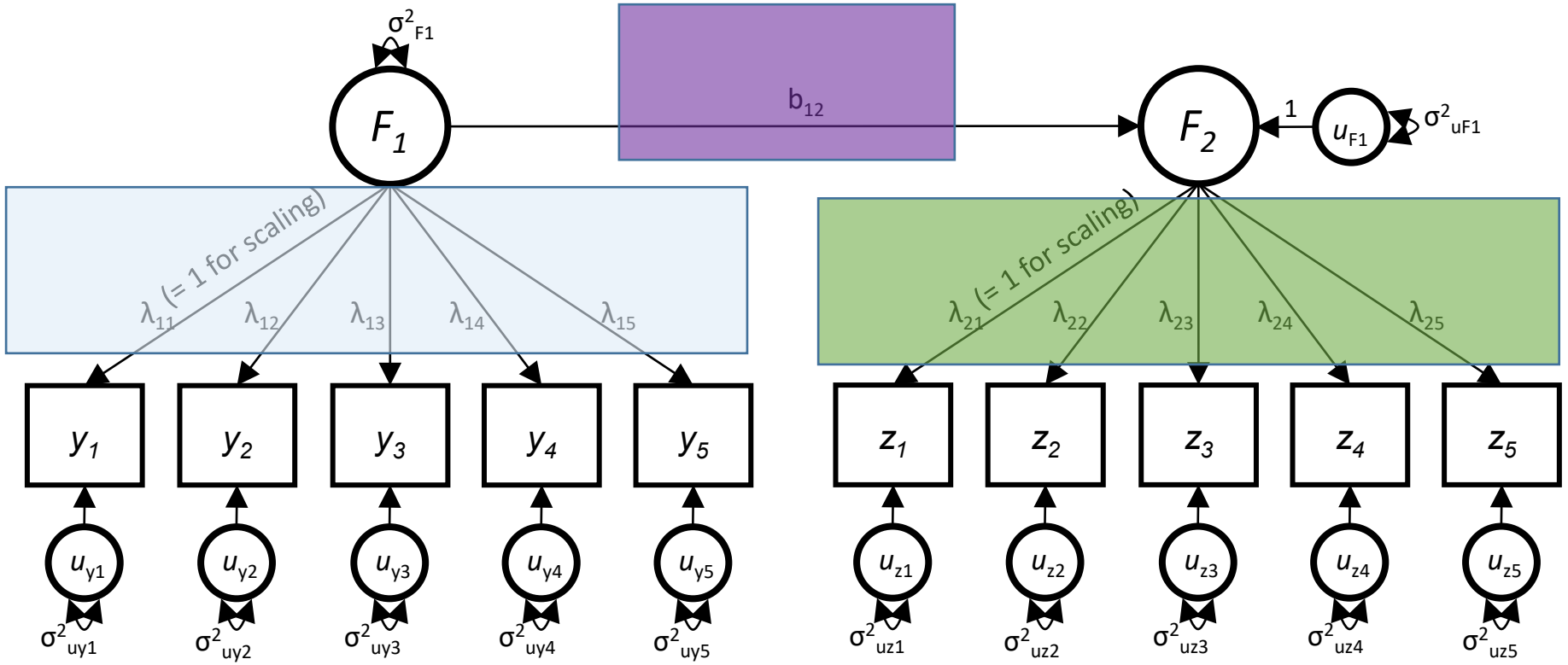


F is unobserved.
Parameters are estimated from,
and fit is evaluated relative to,
the sample covariance matrix for y_1-y_k.

In lavaan syntax:

```
F=~NA*y1+y2+y3+y4+y5
F~~1*F
```

The model that we fit may include some variables for which we do not observe data



In lavaan syntax:

F1=~y1+y2+y3+y4+y5

F2=~z1+z2+z3+z4+z5

F2~F1

Genomic SEM uses these principles to fit structural equation models to genetic covariance matrices derived from GWAS summary statistics using 2 Stage Estimation

- Stage 1: Estimate Genetic Covariance Matrix and associated matrix of standard errors and their co-dependencies
  - We use LD Score Regression, but any method for estimating this matrix (e.g. GREML) and its sampling distribution can be used
- Stage 2: Fit a Structural Equation Model to the Matrices from Stage 1