

# Introduction to OpenMx and Structural Equation Modeling

Elizabeth Prom-Wormley, Lucia Colodro Conde,  
Sarah Medland

# Session Objectives

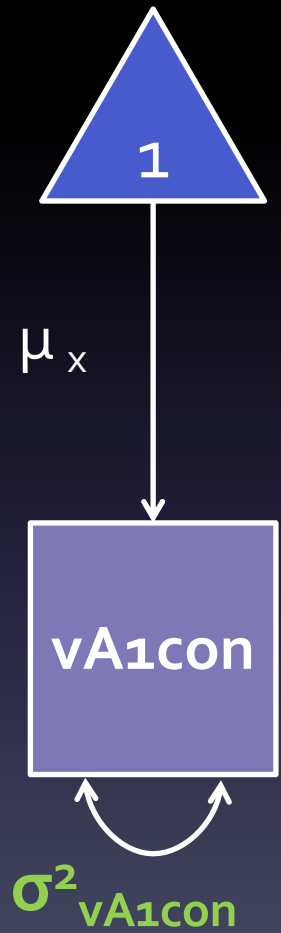
By the end of this session, students will be able to

- Recognize major steps in the development of an OpenMx model
- Translate the implementation of a linear regression between a statistical equation, a structural equation model, and an OpenMx model
- Identify and explain the matrix algebra operations used for implementation of structural equation models in OpenMx

# The Challenge – Part 1

- You have been handed a dataset and you have been asked to consider the degree to which genetic and environmental effects influence the outcome ( $VA_{1con1}$ ).
- Before you begin, you want to get a sense of the outcome, so you calculate the mean and variance
  - Mean =
  - Variance =

# Mean and Variance as a Path Diagram



**Triangle**: a constant variable, usually a vector of ones (here, we use it to reflect a deviation from the mean (none for this picture))

**Single-headed arrows**: linear relationship between two variables. Starts from an independent variable and ends on a dependent variable

**Squares or rectangular boxes**: observed or manifest variables (MEASURED)

**Double-headed arrows**: variance of a variable or covariance between two variables

# The Challenge – Part 2

- You have been handed a dataset and you have been asked to consider the degree to which genetic and environmental effects influence the outcome (VA1con1).
- Before you begin, you want to get a sense of whether you need to account for the effect of AGE on the outcome (VA1con1)
  - Correlation
  - Regression

# Traditional Linear Regression

$$vAcon1_i = \beta_0 + \beta_1 * AGE_i + \varepsilon_i$$

$$y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1$$

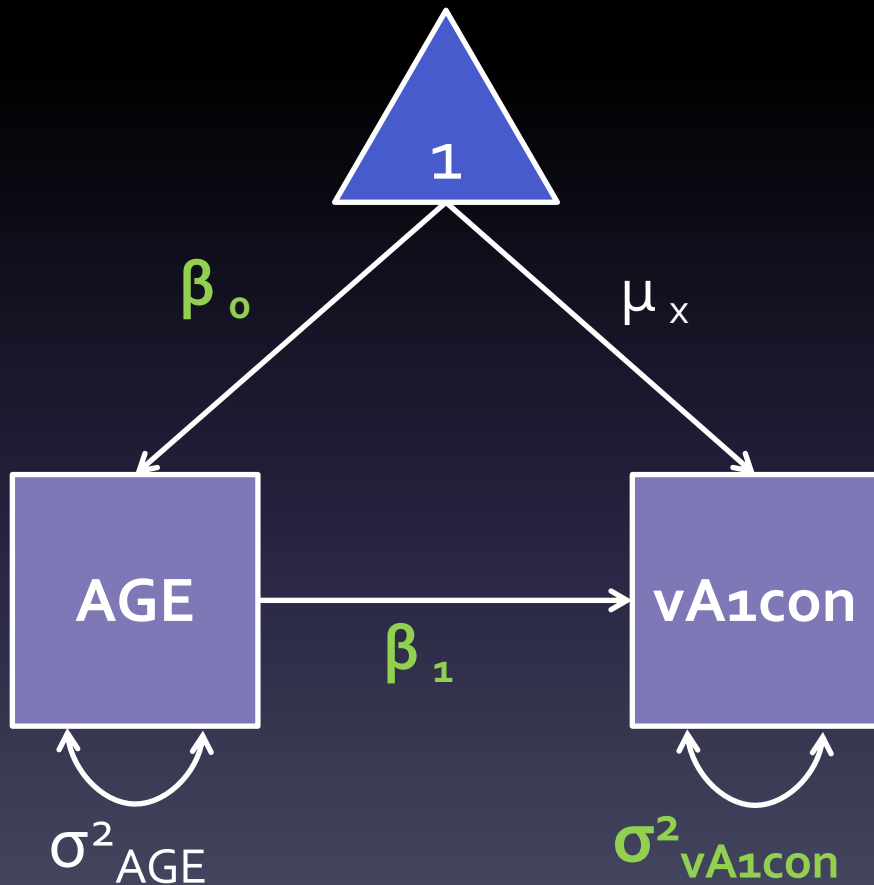
$$y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2$$

$$y_3 = \beta_0 + \beta_1 x_3 + \varepsilon_3$$

⋮

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

# Linear Regression as a Path Diagram



Squares or rectangular boxes: observed or manifest variables

Single-headed arrows: linear relationship between two variables. Starts from an independent variable and ends on a dependent variable

Double-headed arrows: variance of a variable or covariance between two variables

Triangle: a constant variable, usually a vector of ones

Circles or ovals: errors, factors, latent variables

$$y = \beta X + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_i \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \end{bmatrix}$$

- 4 matrices
- 3 estimated parameters
  - vAcon1-Free observed variable
    - $\beta_0$  estimated
    - $\sigma^2_{vA1con}$  estimated
  - AGE- Fixed observed variable
  - Regression of vAcon1 on AGE
    - $\beta_1$  is estimated



# Matrix: A rectangular array of elements arranged in rows and columns

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{array}{l} \swarrow \\ \leftarrow \\ \swarrow \end{array} \text{ROWS}$$

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{array}{l} \uparrow \\ \uparrow \end{array} \text{COLUMNS}$$

Order or dimension of a matrix is defined by the number of ROWS then COLUMNS in the matrix.

Matrix A is a 3 x 2 matrix.

Each element in the matrix is referred to by its placement in a row and column, where  $a_{ij}$  is the element in Matrix A in the  $i_{\text{th}}$  row and  $j_{\text{th}}$  column.

In Matrix A, the number 4 is element  $a(1,2)$

$$Y = \beta X + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ 1 & x_i \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \end{bmatrix}$$

- $Y = i \times 1$  column vector
- $X = i \times 2$  matrix
- $\beta = 2 \times 1$  column vector
- $\varepsilon = n \times 1$  column vector
- $\beta X$  is calculated using matrix multiplication
- $\beta X + \varepsilon$  is calculated using matrix addition

# A Quick Detour into Matrix Algebra

Matrices are a core building block of regression and SEM modeling as well as OpenMx

# Matrix Operations

## Matrix Addition and Subtraction:

- Matrices must be the same size

$$\begin{pmatrix} 2 & 1 \\ 3 & 5 \\ 6 & 2 \end{pmatrix} + \begin{pmatrix} 4 & 8 \\ 7 & 2 \\ 9 & 6 \end{pmatrix} = \begin{pmatrix} 2+4 & 1+8 \\ 3+7 & 5+2 \\ 6+9 & 2+6 \end{pmatrix} = \begin{pmatrix} 6 & 9 \\ 10 & 7 \\ 15 & 8 \end{pmatrix}$$

If the matrices are of different orders, it is impossible to add them

$$\begin{pmatrix} 2 & 1 \\ 3 & 5 \\ 6 & 2 \end{pmatrix} + \begin{pmatrix} 4 & 7 & 9 \\ 8 & 2 & 6 \end{pmatrix} = \text{Undefined}$$

# Dot Product

Also known as the element-wise product

**OpenMx symbol \***

$$A \circ B = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} & A_{12}B_{12} & A_{13}B_{13} \\ A_{21}B_{21} & A_{22}B_{22} & A_{23}B_{23} \\ A_{31}B_{31} & A_{32}B_{32} & A_{33}B_{33} \end{pmatrix}$$

## Matrix Multiplication (Star product)

Number of columns of the first matrix must equal the number of rows of the second matrix.

Product will have as many rows as the first matrix and as many columns as the second matrix.

**OpenMx symbol %\*%**

$$C = A \times B$$

$$C = \begin{pmatrix} 3 & 4 & 7 \\ 5 & 6 & 1 \end{pmatrix} \times \begin{pmatrix} 2 & 1 \\ 3 & 5 \\ 6 & 2 \end{pmatrix} = \begin{pmatrix} 3*2 + 4*3 + 7*6 & 3*1 + 4*5 + 7*2 \\ 5*2 + 6*3 + 1*6 & 5*1 + 6*5 + 1*2 \end{pmatrix} = \begin{pmatrix} 60 & 37 \\ 34 & 37 \end{pmatrix}$$

# Kroneker Product

**OpenMx symbol %x%**

$$A \otimes B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$$

# Quadratic Product

- The quadratic product is extremely useful in statistical analysis (particularly in Structural Equation Modeling)
- OpenMx symbol %&%

$$A\%&\%B = ABA^T =$$
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \times \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$



# Detour Complete

Now, back to regression

Go To

`/home/elizabeth/2020/IntroToOpenMx`

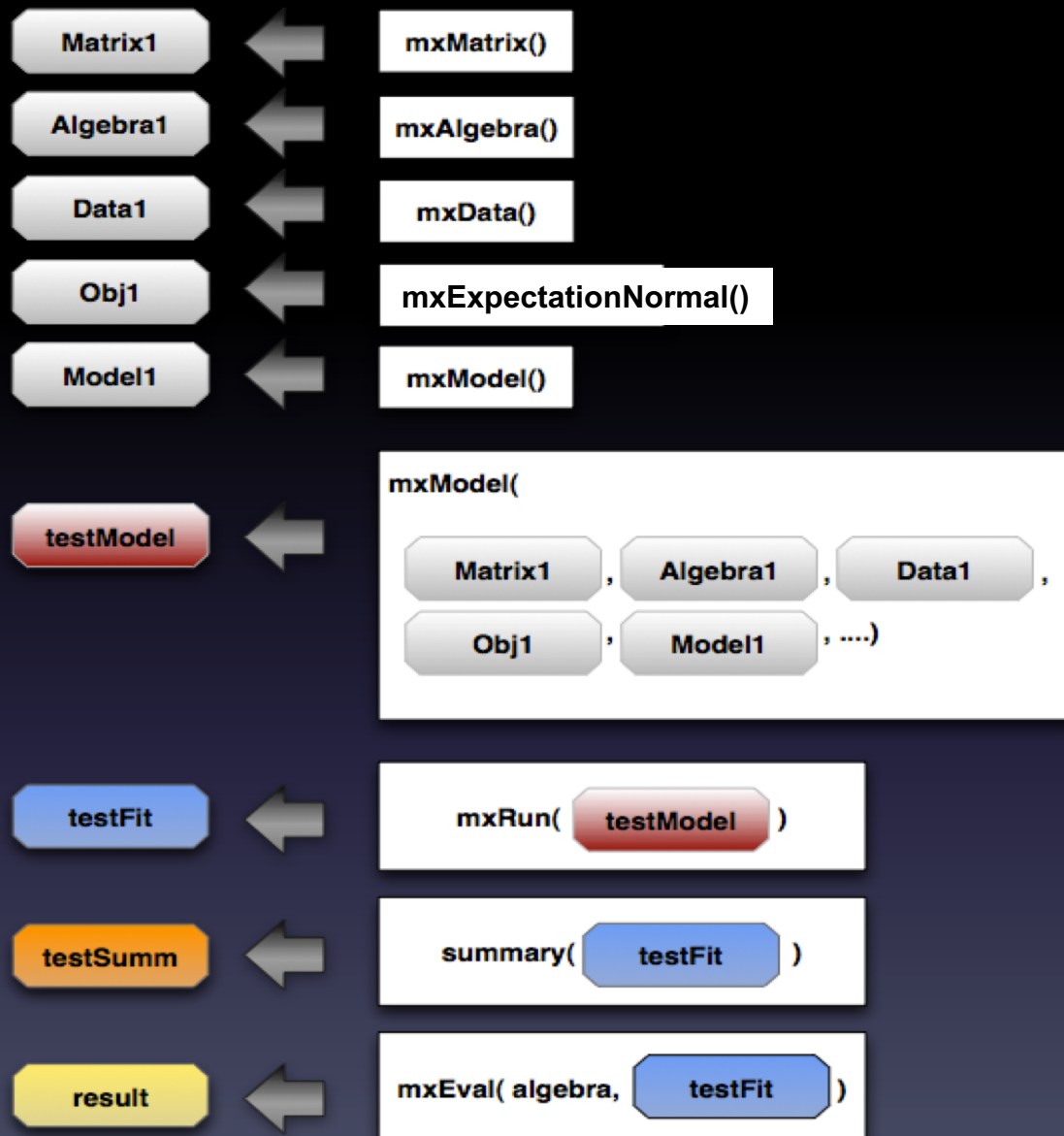
`IntroToOpenMx322020.R`

# What is OpenMx?

- Free, Open-source software package for use in R
- Estimation of advanced multivariate statistical models, particularly structural equation modelling
- Runs on Windows, Mac OSX, and Linux/GNU
- Two main approaches to writing OpenMx models
  - Path or Matrix Specification

# How Does OpenMx Work?

- OpenMx uses functions to build objects
- Arguments to the function have an order
- Order can be changed by naming arguments



**Make matrices** *Variance,  $b_0$ ,  $b_1$ ,  $x$*

**Do Matrix Algebra w Matrices** *expMean*

**Call Data for Use in the Model** *regData*

**Build Model from Matrices/Algebras**  
*exp via mxExpectationNormal*

**Build/Compile Overall Model from Matrices/Algebras and identify fit function**  
*regModel / funML*

**Run Overall Model** *regFit*

**Get Summary Information from Overall Model**  
*regSum*

**Generate Parameter Estimates from Overall Model**

*Resid,  $Beta_0$ ,  $Beta_1$*



# OpenMx Matrices

<code>mxMatrix( type="Diag", nrow=3, ncol=3, free=TRUE, name="a" )</code>	<code>? 0 0</code> <code>0 ? 0</code> <code>0 0 ?</code>
<code>mxMatrix( type="Sdiag", nrow=3, ncol=3, free=TRUE, name="a" )</code>	<code>0 0 0</code> <code>? 0 0</code> <code>? ? 0</code>
<code>mxMatrix( type="Stand", nrow=3, ncol=3, free=TRUE, name="a" )</code>	<code>1 ? ?</code> <code>? 1 ?</code> <code>? ? 1</code>
<code>mxMatrix( type="Symm", nrow=3, ncol=3, free=TRUE, name="a" )</code>	<code>? ? ?</code> <code>? ? ?</code> <code>? ? ?</code>
<code>mxMatrix( type="Lower", nrow=3, ncol=3, free=TRUE, name="a" )</code>	<code>? 0 0</code> <code>? ? 0</code> <code>? ? ?</code>
<code>mxMatrix( type="Full", nrow=2, ncol=4, free=TRUE, name="a" )</code>	<code>? ? ? ?</code> <code>? ? ? ?</code>

# Let's Play!

Open

Open IntroToOpenMx.R

# Linear Regression in R Using lm()

- Intercept =
- Beta<sub>1</sub> =
- Variance =



# Linear Regression in OpenMx

- Intercept =
- Beta<sub>1</sub> =
- Variance =

# Regression Across All Twin 1 Members of a Twin Pair

```
require (OpenMx)
```

```
depVar <- 'vA1con_1'
```

```
# Variance/Covariance matrix
```

```
Variance <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE,  
                      values=10, labels='resid', name="residualVar" )
```

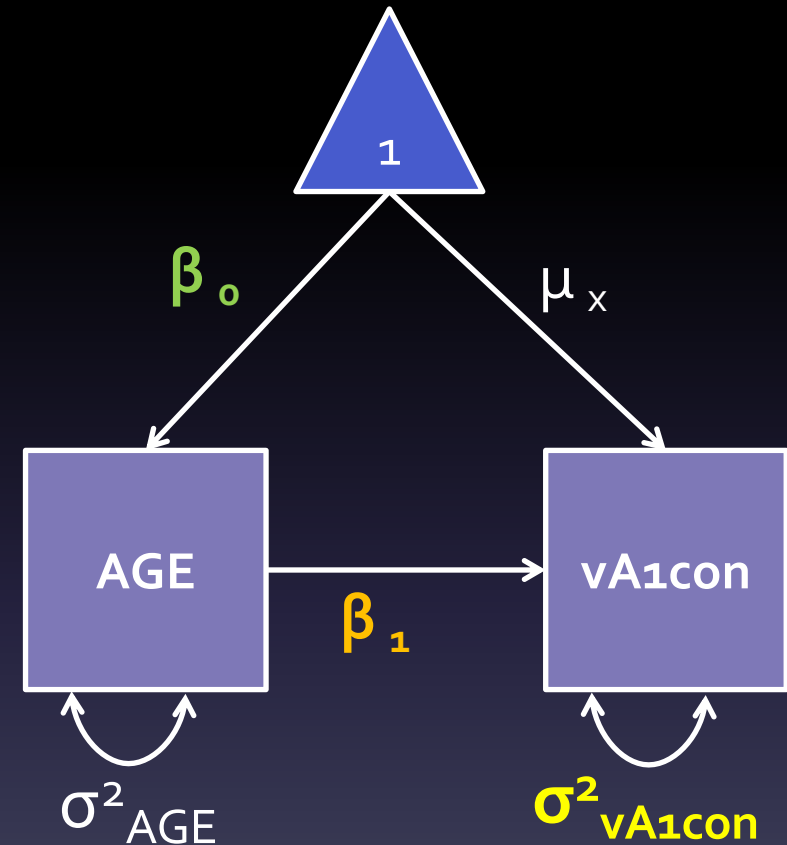
```
# Regression betas
```

```
b0 <- mxMatrix(type="Full", nrow=1, ncol=1, free=T, values=30,  
              labels="beta0", name="Intercept" )
```

```
b1 <- mxMatrix(type="Full", nrow=1, ncol=1, free=T, values=0,  
              labels="beta1", name="bAge" )
```

```
# Independent variable
```

```
x <- mxMatrix(type="Full", nrow=1, ncol=1, free=F,  
             labels="data.AGE_1", name="Age" )
```



```
# Building the model ()
```

```
expMean    <- mxAlgebra(intercept + bAge%x%Age, name="regress")
```

```
# Specify the data
```

```
regData    <- mxData( observed=OZbmi2, type="raw" )
```

```
inclusions <- list(Variance, b0, b1, bAge, expMean)
```

```
exp        <- mxExpectationNormal( covariance="residualVar",  
                                   means="regress", dimnames=depVar )
```

```
# Run the model & summarize output
```

```
funML      <- mxFitFunctionML()
```

```
regModel2  <- mxModel( "Regression201", inclusions, regData, exp, funML )
```

```
regFit2    <- mxRun( regModel2, intervals=FALSE )
```

```
regSum2    <- summary( regFit2 )
```

# The Challenge – Part 3

- You have been handed a dataset and you have been asked to consider the degree to which genetic and environmental effects influence the outcome ( $VA_{1con1}$ ).
- Before you begin, you want to get a sense of how AGE influences both members of a twin pairs on the outcome ( $VA_{1con1}$ ).
  - Do the means differ in both twins for  $VA_{1con1}$  or are they the same?
  - Does the influence of age on  $VA_{1con1}$  differ in both twins or are they the same?

# Let's Play Again!

Open

Open IntroToOpenMx322020.R

# Regression with Two Members of a Twin Pair

	T1	T2
T1	varA	cMZ21
T2	covAB	varB

cov = 2 x 2 matrix

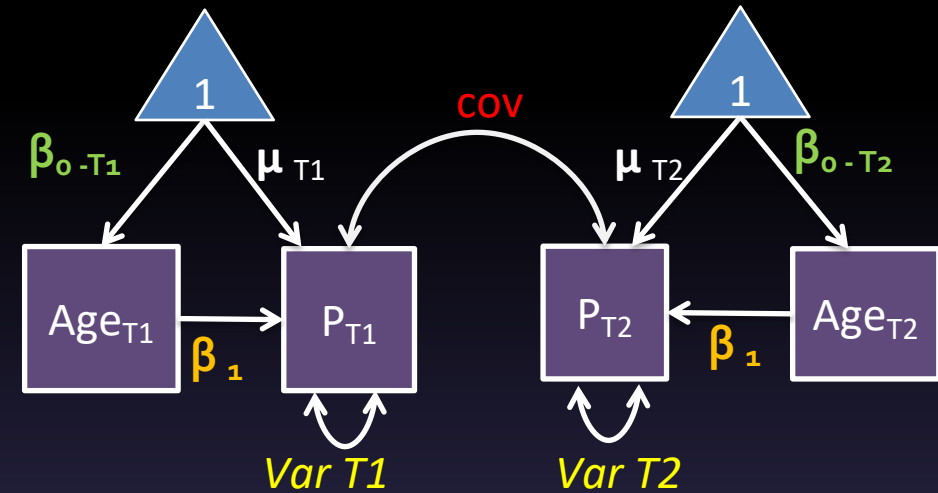
```
Variance <- mxMatrix( type="Symm", nrow=2,
ncol=2, free=TRUE, values=c(15,10,15),
labels=c('varA','covAB','varB'),
name="residualVar" )
```

T1	T2
betaoA	betaoB

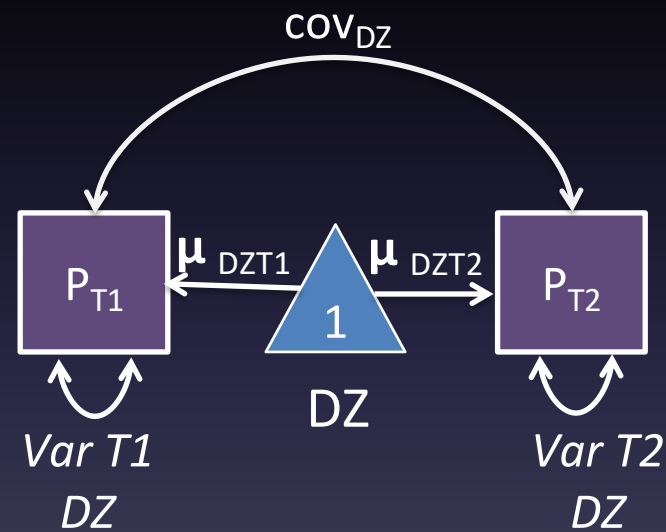
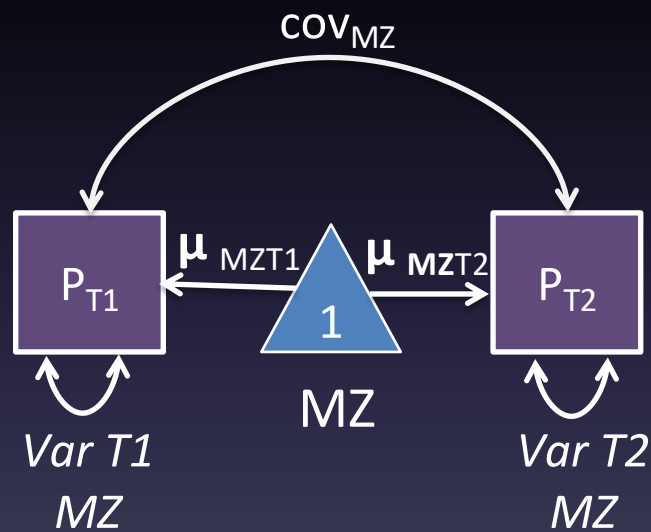
```
bo <- mxMatrix( type="Full", nrow=1, ncol=2,
free=T, values=22, labels=c("betaoA", "betaoB"),
name="Intercept" )
```

T1	T2
AGE_1	AGE_2

```
x <- mxMatrix( type="Full", nrow=1, ncol=2,
free=F, labels=c("data.AGE_1", "data.AGE_2"),
name="Age" )
```



# Stay Tuned... We Will Continue to Build... What Happens When You Add Zygosity?



# Data Preparation Considerations

- The algebra style used in OpenMx expects 1 line per case/family
- (Almost) limitless number of families and variables
- Data needs to be read into R before it can be analyzed
  - (the commands to read the data can be nested within the R script)
- Default missing code is NA



# Resources

- Basic background on general SEM
  - <http://www.nyu.edu/classes/shrout/G89-2247/McDonaldMoon-Ho2002.pdf>
- OpenMx website
  - <https://openmx.ssri.psu.edu/>
- Wiki Page
  - <https://openmx.ssri.psu.edu/wiki/main-page>
- Discussion Forum
  - <https://openmx.ssri.psu.edu/forums>
- YouTube Videos
  - [https://www.youtube.com/playlist?list=PL-eLmogqXe\\_SfwGcCIW41gUXoL2b7U1zT](https://www.youtube.com/playlist?list=PL-eLmogqXe_SfwGcCIW41gUXoL2b7U1zT)

# Reference – Path Tracing Rules for SEM

- 1 Find All Distinct Chains between Variables:
  - A) Go backwards along zero or more single-headed arrows
  - B) Change direction at one and only one Double-headed arrow
  - C) Trace forwards along zero or more Single-headed arrows
- 2 Multiply path coefficients in a chain
- 3 Sum the results of step 2.
  
- Note- For covariance of a variable with itself (Variance), chains are distinct if they have different paths or a different order

# Thank You!

[Elizabeth.Prom-Wormley@vcuhealth.org](mailto:Elizabeth.Prom-Wormley@vcuhealth.org)