Multivariate Genetic Analysis

Michael D. Hunter with much credit due to
Hermine H. Maes, Meike Bartels, Elizabeth Prom-Wormley,
Michel Nivard, Tim Bates & many others

Department of Pediatrics
University of Oklahoma Health Sciences Center

March 7, 2018
International Workshop on Statistical Genetics and
Methodology of Twin and Family Studies
Boulder, CO
`Faculty/mikeh/2018/Multivariate`

A Distinction

- Earlier: Not so theoretical models
  - Saturated Model
  - Fully Correlated Genetic Factors (Cholesky) Model
- Now: Theoretical models
  - Common Pathway Model
  - Independent Pathway Model

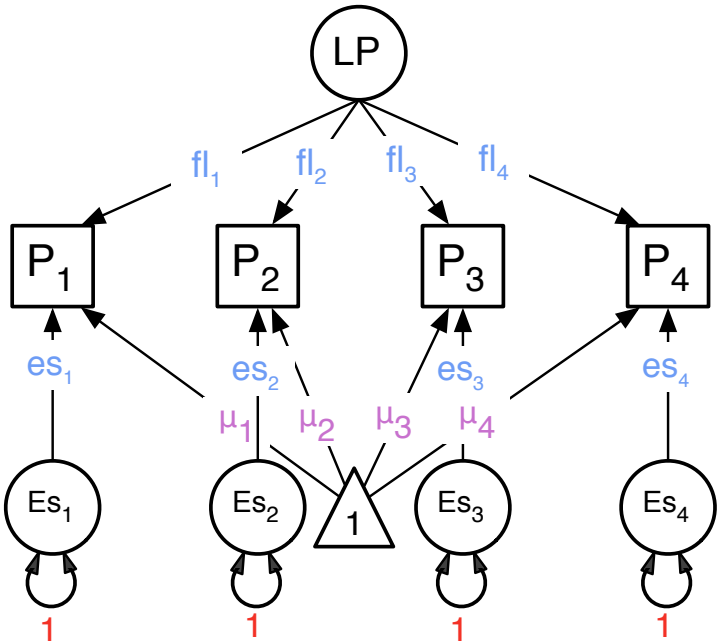
Scientific questions you can ask

- In *univariate* analyses: what are the contributions of additive genetic, dominance genetic, shared environmental, and unique environmental factors to the variance?
- In *multivariate* analyses: what are these contributions to the **covariance** between two or more traits?

Common Pathway Model

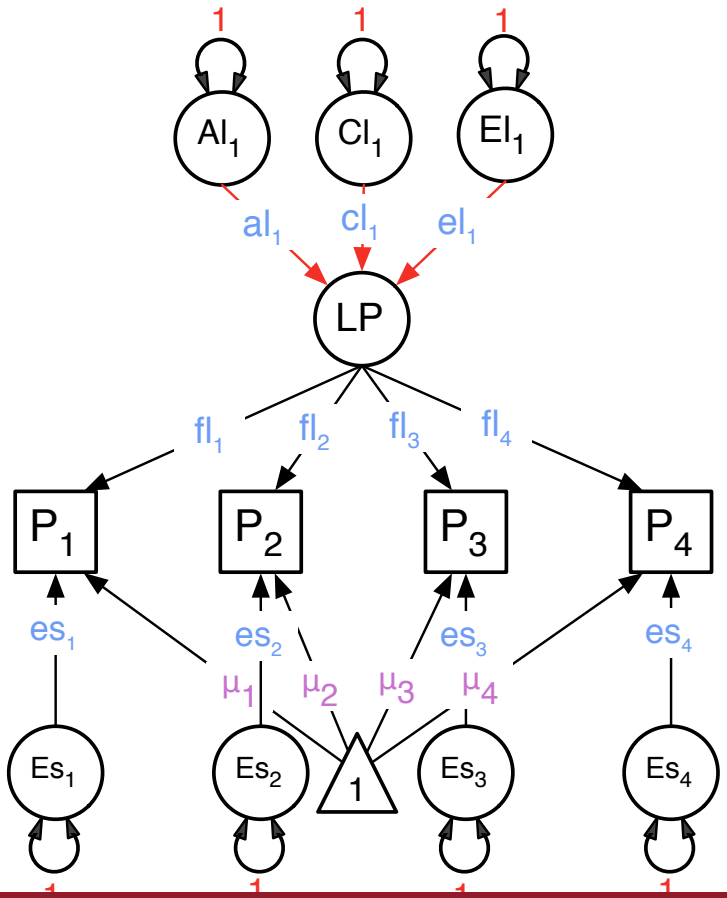
- Theoretical model
- Start with a phenotypic factor model
- Origin is more psychometric than biometric
- Fixes same covariance structure across A, C, and E
- Decompose the factor variance into A, C, E
- Decompose the residual variances into A, C, E

Factor Loadings



$$\begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{matrix} \begin{bmatrix} fl_{11} \\ fl_{21} \\ fl_{31} \\ fl_{41} \end{bmatrix}$$

Latent Phenotype ACE

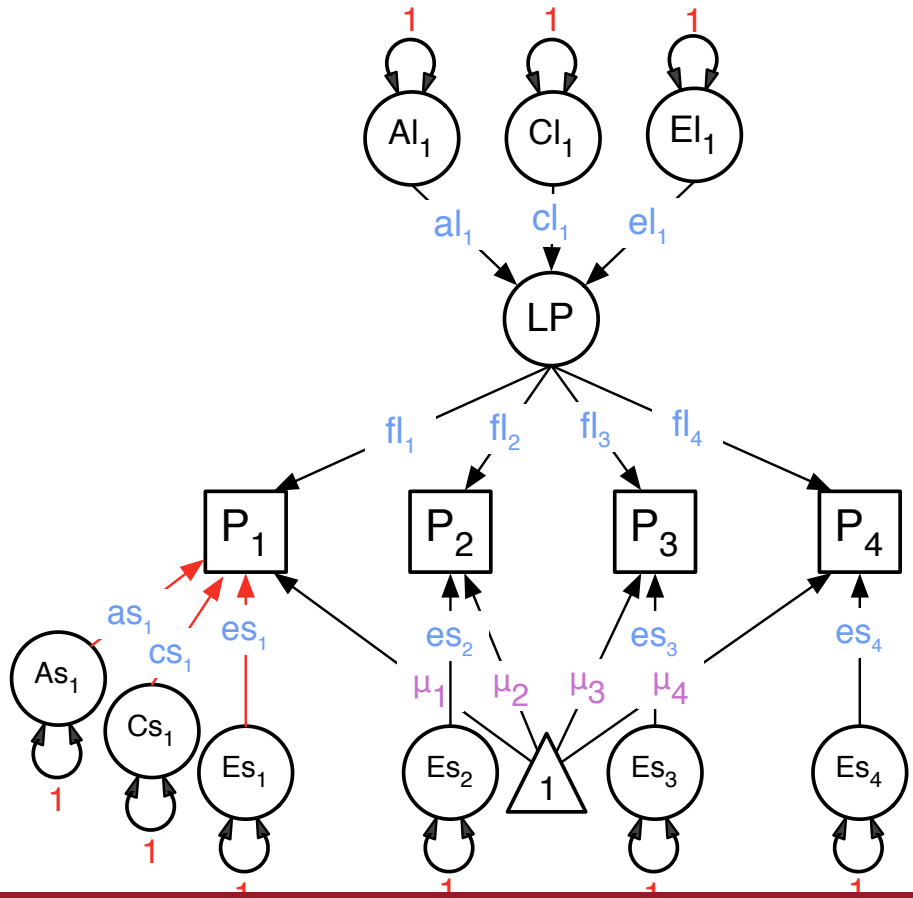


$$\begin{bmatrix} al_{11} \end{bmatrix} \qquad \begin{bmatrix} cl_{11} \end{bmatrix}$$

$$\begin{bmatrix} el_{11} \end{bmatrix}$$

$$\begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{matrix} \begin{bmatrix} fl_{11} \\ fl_{21} \\ fl_{31} \\ fl_{41} \end{bmatrix}$$

ACE Specifics



$$\begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{matrix} \begin{bmatrix} es_{11} & 0 & 0 & 0 \\ 0 & es_{22} & 0 & 0 \\ 0 & 0 & es_{33} & 0 \\ 0 & 0 & 0 & es_{44} \end{bmatrix}$$

$$\begin{bmatrix} as_{11} & 0 & 0 & 0 \\ 0 & as_{22} & 0 & 0 \\ 0 & 0 & as_{33} & 0 \\ 0 & 0 & 0 & as_{44} \end{bmatrix}$$

$$\begin{bmatrix} cs_{11} & 0 & 0 & 0 \\ 0 & cs_{22} & 0 & 0 \\ 0 & 0 & cs_{33} & 0 \\ 0 & 0 & 0 & cs_{44} \end{bmatrix}$$

Common A Factors
Specific A Factors

object: pathFl
matrix name: fl

$$\begin{bmatrix} fl_{11} \\ fl_{21} \\ fl_{31} \\ fl_{41} \end{bmatrix} \times [al_{11}] \times [al_{11}] \times \begin{bmatrix} fl_{11} & fl_{21} & fl_{31} & fl_{41} \end{bmatrix} = \begin{bmatrix} fl_{11}^2 al_{11}^2 & fl_{11} fl_{21} al_{11}^2 & fl_{11} fl_{31} al_{11}^2 & fl_{11} fl_{41} al_{11}^2 \\ fl_{21} fl_{11} al_{11}^2 & fl_{21}^2 al_{11}^2 & fl_{21} fl_{31} al_{11}^2 & fl_{21} fl_{41} al_{11}^2 \\ fl_{31} fl_{11} al_{11}^2 & fl_{31} fl_{21} al_{11}^2 & fl_{31}^2 al_{11}^2 & fl_{31} fl_{41} al_{11}^2 \\ fl_{41} fl_{11} al_{11}^2 & fl_{41} fl_{21} al_{11}^2 & fl_{41} fl_{31} al_{11}^2 & fl_{41}^2 al_{11}^2 \end{bmatrix}$$

object: pathAl
matrix name: al

$$\begin{bmatrix} as_{11} & 0 & 0 & 0 \\ 0 & as_{22} & 0 & 0 \\ 0 & 0 & as_{33} & 0 \\ 0 & 0 & 0 & as_{44} \end{bmatrix} \times \begin{bmatrix} as_{11} & 0 & 0 & 0 \\ 0 & as_{22} & 0 & 0 \\ 0 & 0 & as_{33} & 0 \\ 0 & 0 & 0 & as_{44} \end{bmatrix} = \begin{bmatrix} as_{11}^2 & 0 & 0 & 0 \\ 0 & as_{22}^2 & 0 & 0 \\ 0 & 0 & as_{33}^2 & 0 \\ 0 & 0 & 0 & as_{44}^2 \end{bmatrix}$$

object: pathAs
matrix name: as

```
pathFl    <- mxMatrix( type="Full", nrow=nv, ncol=nl, free=TRUE, values=.2,
                       labels=labFull("fl",nv,nl), name="fl" )
pathAl    <- mxMatrix( type="Lower", nrow=nl, ncol=nl, free=TRUE, values=.6,
                       labels=labLower("al",nl), lbound=.00001, name="al" )
pathAs    <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=.5,
labels=labDiag("as",nv), lbound=.00001, name="as" )
```

Total A Covariance

```
fl %&% (al %*% t(al))
```

$$\begin{bmatrix} fl_{11}^2 al_{11}^2 & fl_{11} fl_{21} al_{11}^2 & fl_{11} fl_{31} al_{11}^2 & fl_{11} fl_{41} al_{11}^2 \\ fl_{21} fl_{11} al_{11}^2 & fl_{21}^2 al_{11}^2 & fl_{21} fl_{31} al_{11}^2 & fl_{21} fl_{41} al_{11}^2 \\ fl_{31} fl_{11} al_{11}^2 & fl_{31} fl_{21} al_{11}^2 & fl_{31}^2 al_{11}^2 & fl_{31} fl_{41} al_{11}^2 \\ fl_{41} fl_{11} al_{11}^2 & fl_{41} fl_{21} al_{11}^2 & fl_{41} fl_{31} al_{11}^2 & fl_{41}^2 al_{11}^2 \end{bmatrix} + \begin{bmatrix} as_{11}^2 & 0 & 0 & 0 \\ 0 & as_{22}^2 & 0 & 0 \\ 0 & 0 & as_{33}^2 & 0 \\ 0 & 0 & 0 & as_{44}^2 \end{bmatrix} = \begin{bmatrix} fl_{11}^2 al_{11}^2 + as_{11}^2 & fl_{11} fl_{21} al_{11}^2 & fl_{11} fl_{31} al_{11}^2 & fl_{11} fl_{41} al_{11}^2 \\ fl_{21} fl_{11} al_{11}^2 & fl_{21}^2 al_{11}^2 + as_{22}^2 & fl_{21} fl_{31} al_{11}^2 & fl_{21} fl_{41} al_{11}^2 \\ fl_{31} fl_{11} al_{11}^2 & fl_{31} fl_{21} al_{11}^2 & fl_{31}^2 al_{11}^2 + as_{33}^2 & fl_{31} fl_{41} al_{11}^2 \\ fl_{41} fl_{11} al_{11}^2 & fl_{41} fl_{21} al_{11}^2 & fl_{41} fl_{31} al_{11}^2 & fl_{41}^2 al_{11}^2 + as_{44}^2 \end{bmatrix}$$

```
as %*% t(as)
```

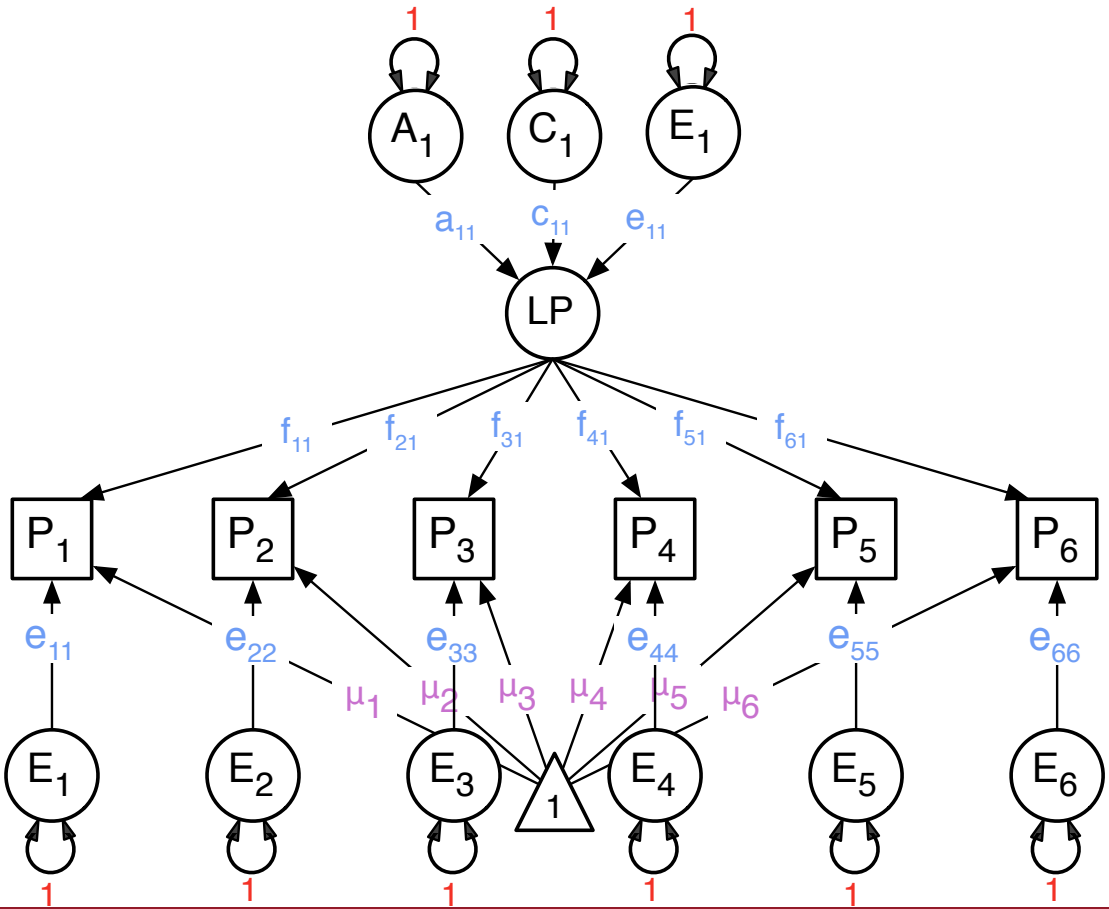
```
object: CovA
matrix name:A
```

```
covA      <- mxAlgebra( expression=fl %&% (al %*% t(al)) + as %*% t(as), name="A" )
```

Common Pathway Model

- Theoretical model
- Start with a phenotypic factor model
- Origin is more psychometric than biometric
- Fixes same covariance structure across A, C, and E
- Decompose the factor variance into A, C, E
- Decompose the residual variances into A, C, E

CP Model


1
1
1
A1
C1
E1
a11
c11
e11
LP
f11
f21
f31
f41
f51
f61
P1
P2
P3
P4
P5
P6
e11
e22
e33
e44
e55
e66
μ1
μ2
μ3
μ4
μ5
μ6
E1
E2
E3
E4
E5
E6
1
1
1
1
1
1
1

Note! For readabilty of the diagrams, we sometimes omit the residual variance decomposition.

Common Pathway

Variance Component	a2	c2	e2	
Common Factors	al 1 x 1	cl 1 x 1	el 1 x 1	fl nv x 1
Residual Factors	as nv x nv	cs nv x nv	es nv x nv	

Constraint on Variance of Latent Phenotype

```
# Fit Common Pathway ACE Model
# ------------------------------------------------------------------------------
nl        <- 1

# Matrices ac, cc, and ec to store a, c, and e path coefficients for latent phenotype(s)
pathAl    <- mxMatrix( type="Lower", nrow=nl, ncol=nl, free=TRUE, values=.6, labels=labLower("al",nl), lbound=.00001, name="al" )
pathCl    <- mxMatrix( type="Lower", nrow=nl, ncol=nl, free=TRUE, values=.6, labels=labLower("cl",nl), lbound=.00001, name="cl" )
pathEl    <- mxMatrix( type="Lower", nrow=nl, ncol=nl, free=TRUE, values=.6, labels=labLower("el",nl), lbound=.00001, name="el" )

# Matrix and Algebra for constraint on variance of latent phenotype
covarLP   <- mxAlgebra( expression=al %*% t(al) + cl %*% t(cl) + el %*% t(el), name="CovarLP" )
varLP     <- mxAlgebra( expression= diag2vec(CovarLP), name="VarLP" )
unit      <- mxMatrix( type="Unit", nrow=nl, ncol=1, name="Unit")
varLP1    <- mxConstraint( expression=VarLP == Unit, name="varLP1")

# Matrix f for factor loadings on latent phenotype
pathFl    <- mxMatrix( type="Full", nrow=nv, ncol=nl, free=TRUE, values=.2, labels=labFull("fl",nv,nl), name="fl" )

# Matrices A, C, and E compute variance components
covA      <- mxAlgebra( expression=fl %&% (al %*% t(al)) + as %*% t(as), name="A" )
covC      <- mxAlgebra( expression=fl %&% (cl %*% t(cl)) + cs %*% t(cs), name="C" )
covE      <- mxAlgebra( expression=fl %&% (el %*% t(el)) + es %*% t(es), name="E" )
```

latent phenotype nf x nf

a^2 + c^2 + e^2 =1

factor loadings

factor loadings x ace on LP
+ specifics

Fitting CP Model

```
# Create Model Objects for Multiple Groups
pars      <- list(meanG, matI, invSD,
                  pathAl, pathCl, pathEl, covarLP, varLP, unit, pathFl, pathAs, pathCs, pathEs, covA, covC, covE, covP)
modelMZ   <- mxModel( name="MZ", pars, covMZ, expCovMZ, dataMZ, expMZ, funML )
modelDZ   <- mxModel( name="DZ", pars, covDZ, expCovDZ, dataDZ, expDZ, funML )
multi     <- mxFitFunctionMultigroup( c("MZ","DZ") )

# Build & Run Model
modelCP   <- mxModel( "mulCPc", pars, varLP1, modelMZ, modelDZ, multi )
fitCP     <- mxRun(modelCP, intervals=F )
sumCP     <- summary( fitCP )
mxCompare( fitACE, fitCP )
parameterSpecifications(fitCP)

# Generate List of Parameter Estimates and Derived Quantities using formatOutputMatrices
matCPpaths <- c("al","cl","el","iSD %*% fl","iSD %*% as","iSD %*% cs","iSD %*% es")
labCPpaths <- c("stPathAl","stPathCl","stPathEl","stPathFl","stPathAs","stPathCs","stPathEs")
formatOutputMatrices(fitCP, matCPpaths, labCPpaths, vars, 4)
```

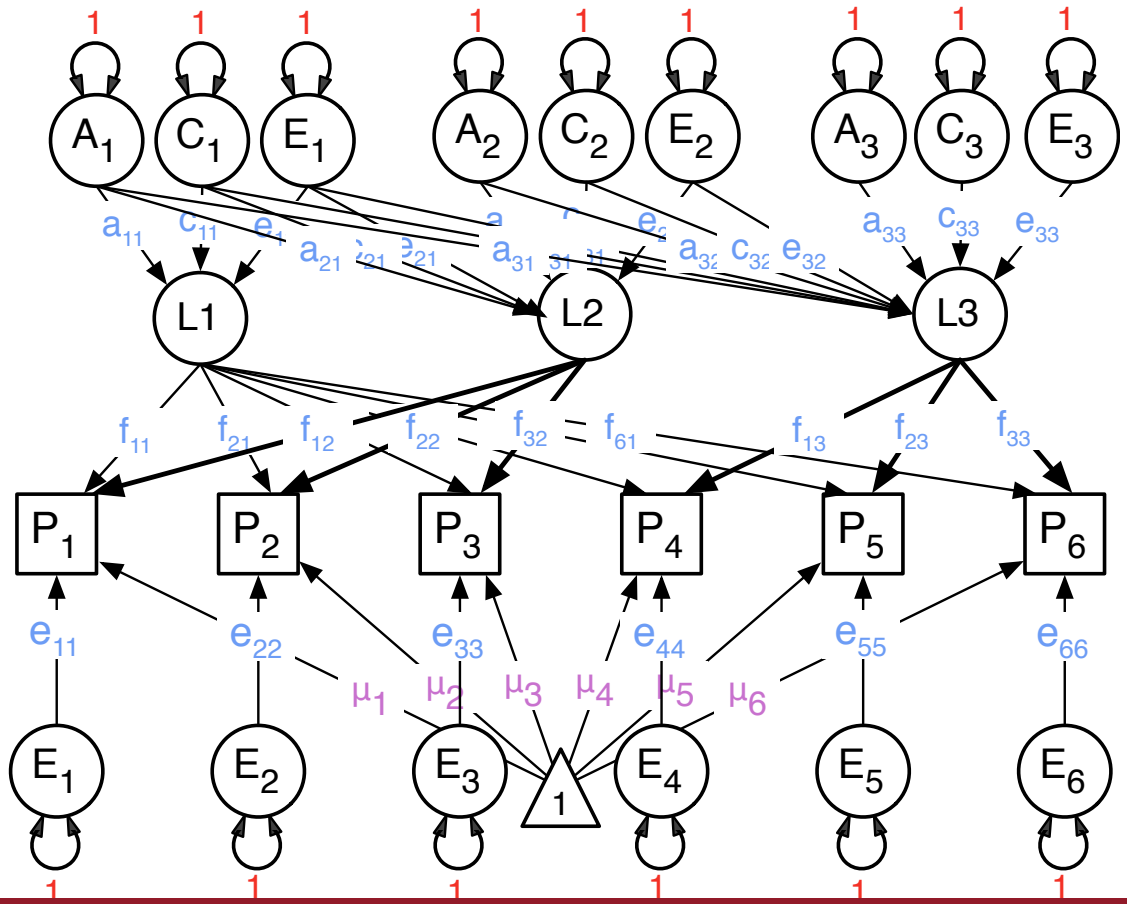
new objects

constraint object in combined model only

already standardized

Can you have multiple latent factors?

CP 3L Model


1
A1
C1
E1
A2
C2
E2
A3
C3
E3
a11
c11
e1
a21
a31
a32
c32
e32
a33
c33
e33
L1
L2
L3
f11
f21
f12
f22
f32
f61
f13
f23
f33
P1
P2
P3
P4
P5
P6
e11
e22
e33
e44
e55
e66
μ1
μ2
μ3
μ4
μ5
μ6
E1
E2
E3
1
E4
E5
E6

Common pathway practical

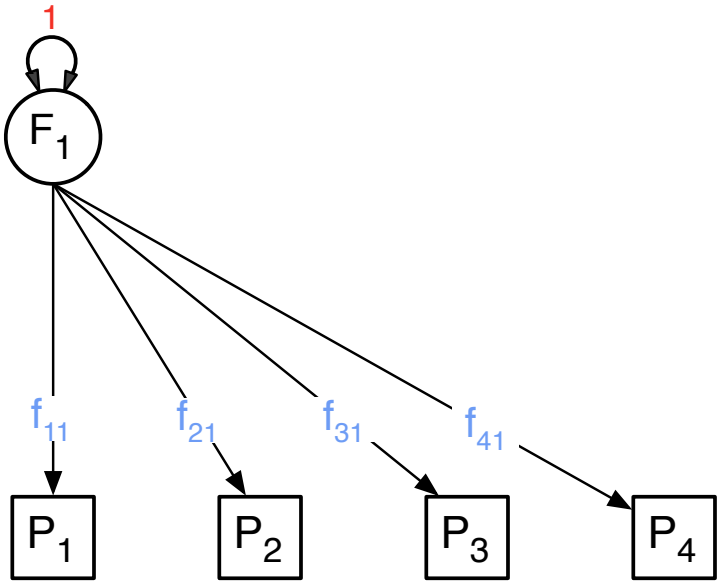
`Faculty/mikeh/2018/Multivariate`

Genetic Factor Model

(Martin & Eaves, 1977)

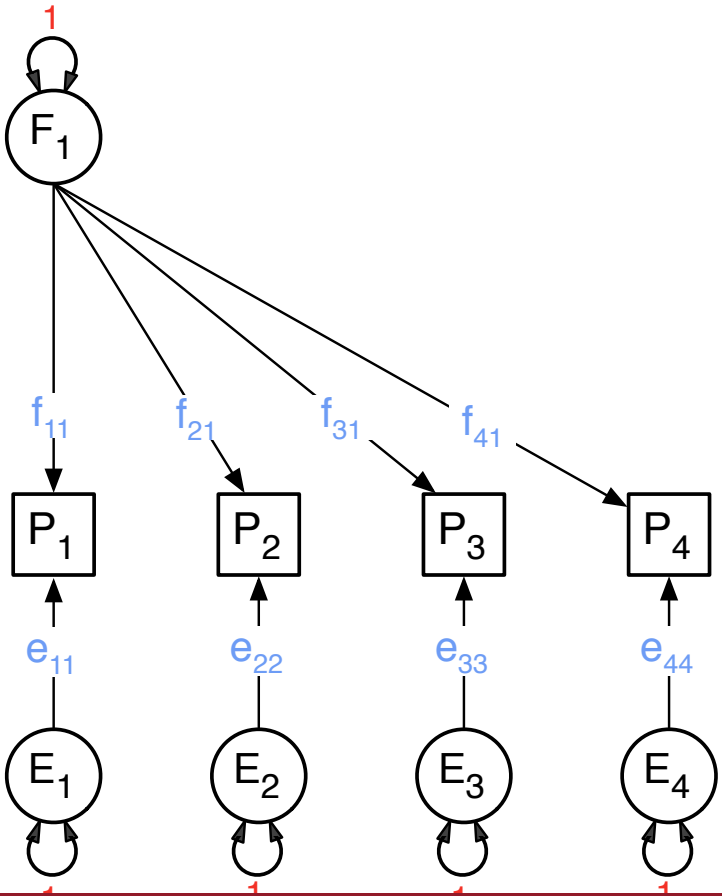
- Theoretical model
- Start with a biometric factor model
- Origin is more biometric than psychometric
- Allows different covariance structure across A, C, and E
- Create latent factors for A, C, E

Common Factor



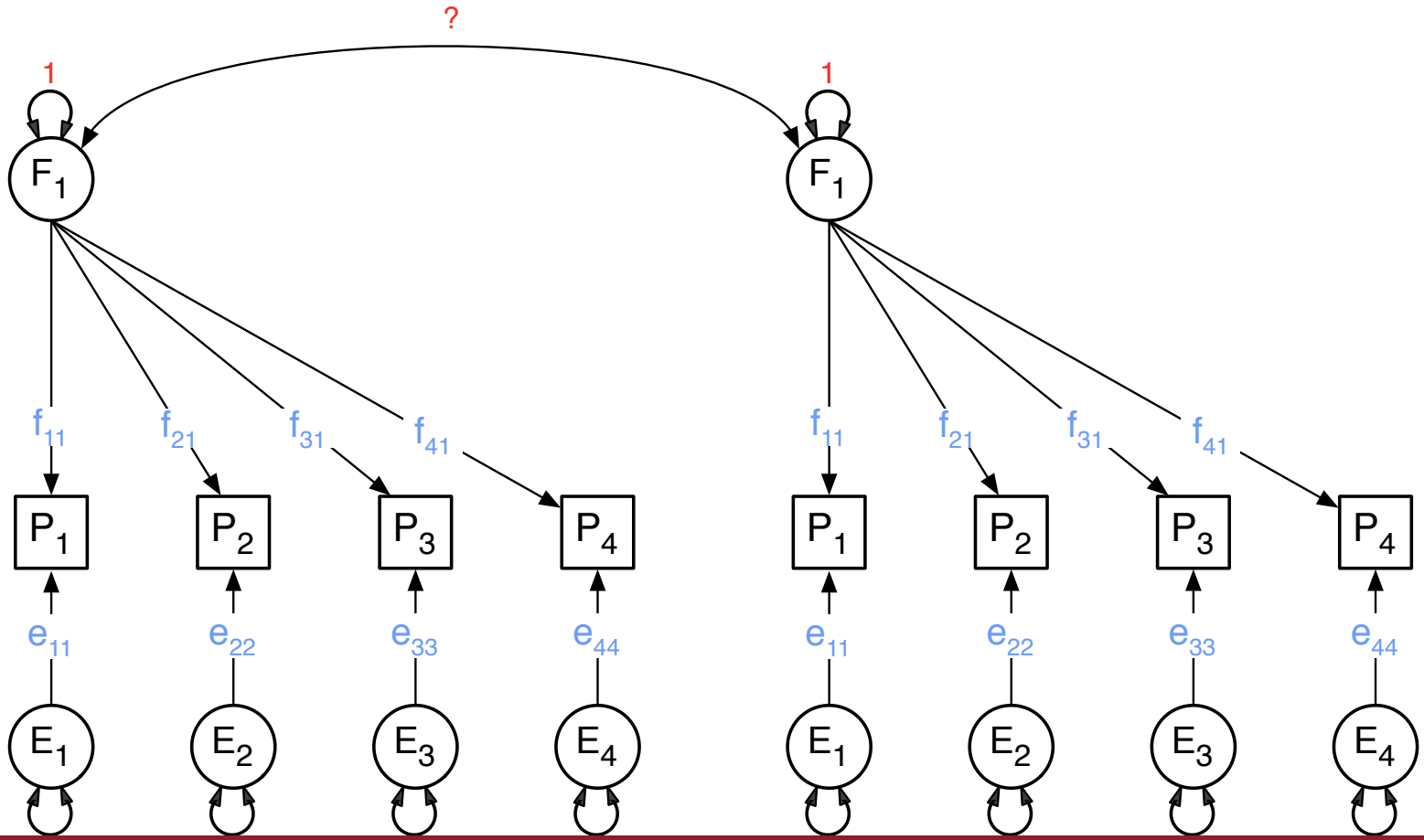
$$
\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array}
\begin{array}{c} F_1 \\ \left[\begin{array}{c} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{array}\right] \end{array}
$$

Residuals

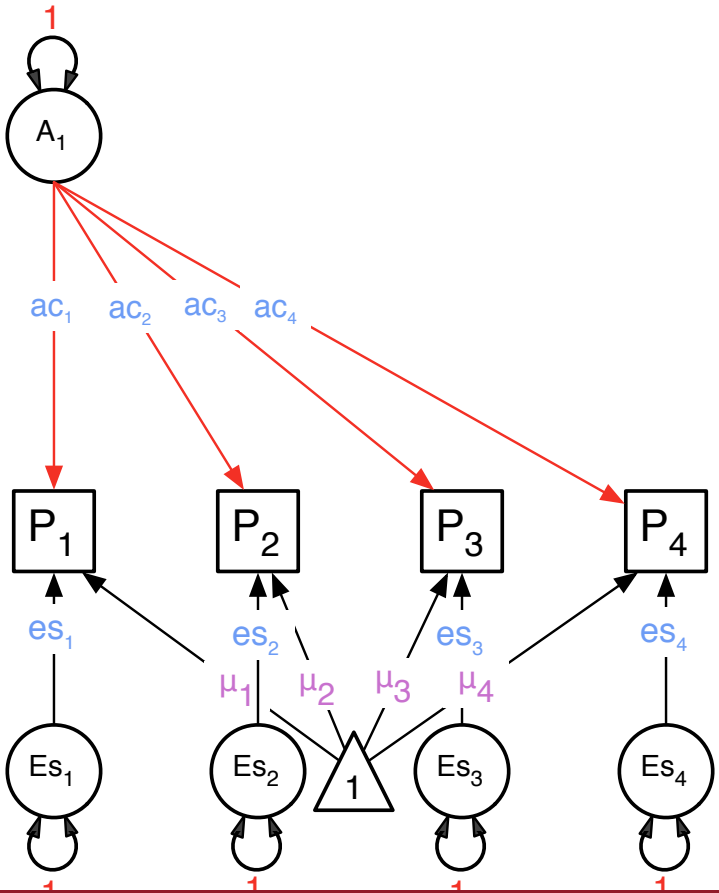


$$\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{bmatrix} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{bmatrix}$$

What about Twins

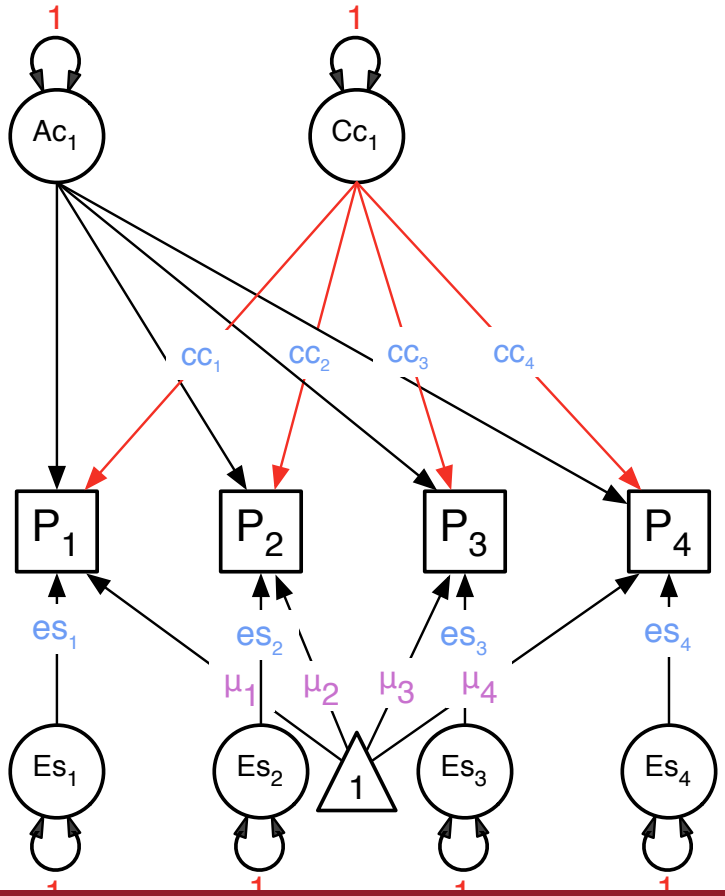


Common A Factor



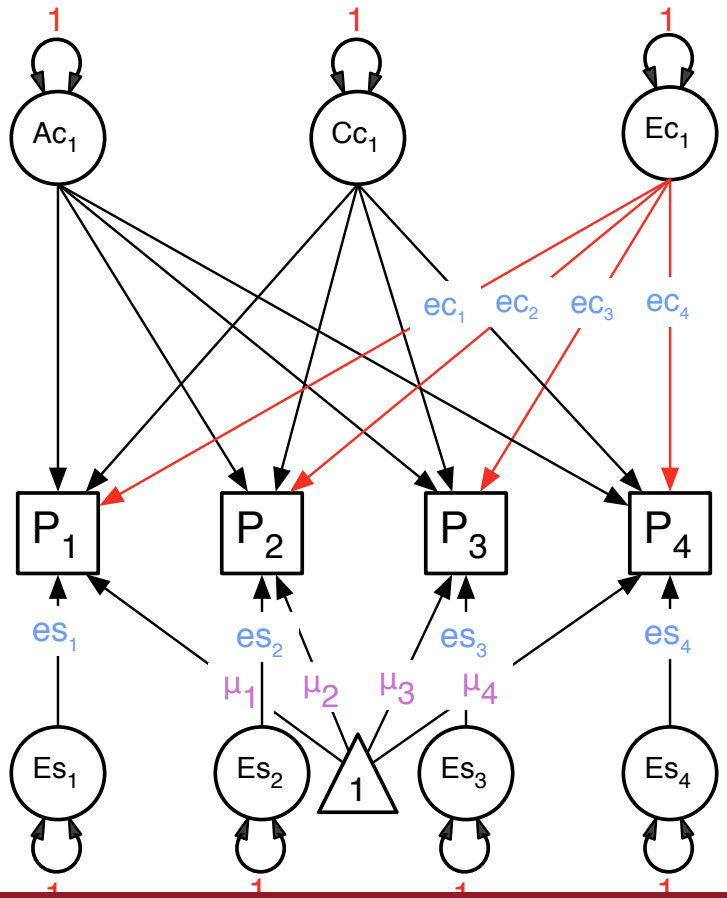
$$\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} A_1 \\ \left[\begin{array}{c} ac_{11} \\ ac_{21} \\ ac_{31} \\ ac_{41} \end{array}\right] \end{array}$$

Common C Factor



$$
\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array}
\begin{array}{c} C_1 \\ \begin{bmatrix} cc_{11} \\ cc_{21} \\ cc_{31} \\ cc_{41} \end{bmatrix} \end{array}
$$

Common E Factor

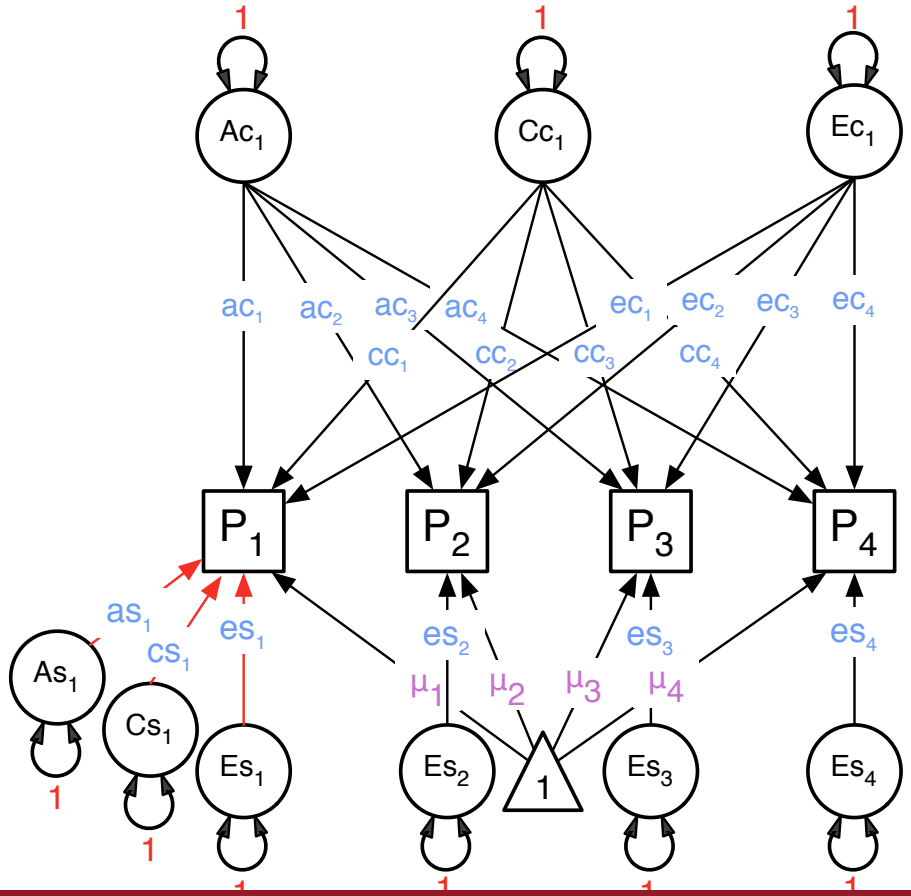


$$
\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array}
\begin{array}{c} E_1 \\ \left[\begin{array}{c} ec_{11} \\ ec_{21} \\ ec_{31} \\ ec_{41} \end{array}\right] \end{array}
$$

Independent Pathway Model

- Theoretical model
- Start with a biometric factor model
- Origin is more biometric than psychometric
- Allows different covariance structure across A, C, and E
- Create latent factors for *common* A, C, E components
- Decompose the residual variances into *specific* A, C, E components

ACE Specifics



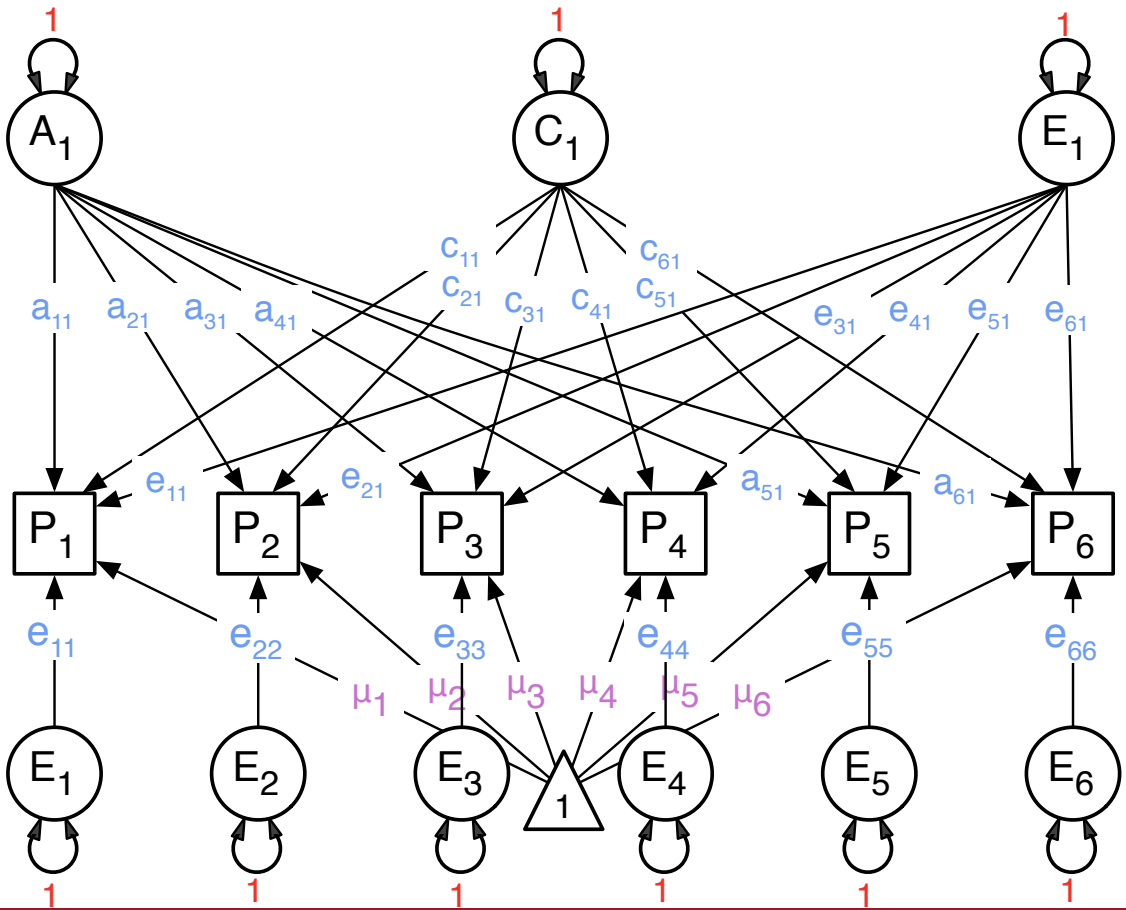
$$P_1 \begin{bmatrix} es_{11} & 0 & 0 & 0 \\ 0 & es_{22} & 0 & 0 \\ 0 & 0 & es_{33} & 0 \\ 0 & 0 & 0 & es_{44} \end{bmatrix}$$

$$\begin{bmatrix} as_{11} & 0 & 0 & 0 \\ 0 & as_{22} & 0 & 0 \\ 0 & 0 & as_{33} & 0 \\ 0 & 0 & 0 & as_{44} \end{bmatrix}$$

$$\begin{bmatrix} cs_{11} & 0 & 0 & 0 \\ 0 & cs_{22} & 0 & 0 \\ 0 & 0 & cs_{33} & 0 \\ 0 & 0 & 0 & cs_{44} \end{bmatrix}$$

Note! For readabilty of the diagrams, we sometimes omit the residual variance decomposition.

IP Model


1
A1
C1
E1
c11
c21
c31
c41
c51
c61
a11
a21
a31
a41
e31
e41
e51
e61
e11
e21
a51
a61
P1
P2
P3
P4
P5
P6
e11
e22
e33
e44
e55
e66
μ1
μ2
μ3
μ4
μ5
μ6
E1
E2
E3
E4
E5
E6
1
1
1
1
1
1
1

Independent Pathway

Variance Component	a2	c2	e2
Common Factors	ac nv x 1	cc nv x 1	ec nv x 1
Residual Factors	as nv x nv	cs nv x nv	es nv x nv

Independent Pathway Model

Identification

- Be careful when adding common factors
- Total parameters per source of variance must be less than $nv * (nv + 1)/2$ for the number of phenotypes, nv
- For a single common factor with only 2 indicators, equate the 2 factor loadings
- Alternatively, remove the common factor and add a correlated residual
- If in doubt, try `mxCheckIdentification`
- When not identified, gives offending parameters

Independent Pathways

```
# Fit Independent Pathway ACE Model
# ------------------------------------------------------------------------------
nf        <- 1        # number of factors

# Matrices ac, cc, and ec to store a, c, and e path coefficients for common factors
pathAc    <- mxMatrix( type="Full", nrow=nv, ncol=nf, free=TRUE, values=.6, labels=labFull("ac",nv,nf), name="ac" )
pathCc    <- mxMatrix( type="Full", nrow=nv, ncol=nf, free=TRUE, values=.6, labels=labFull("cc",nv,nf), name="cc" )
pathEc    <- mxMatrix( type="Full", nrow=nv, ncol=nf, free=TRUE, values=.6, labels=labFull("ec",nv,nf), name="ec" )

# Matrices as, cs, and es to store a, c, and e path coefficients for specific factors
pathAs    <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, labels=labDiag("as",nv), lbound=.00001, name="as" )
pathCs    <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, labels=labDiag("cs",nv), lbound=.00001, name="cs" )
pathEs    <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=5, labels=labDiag("es",nv), lbound=.00001, name="es" )

# Matrices A, C, and E compute variance components
covA      <- mxAlgebra( expression=ac %*% t(ac) + as %*% t(as), name="A" )
covC      <- mxAlgebra( expression=cc %*% t(cc) + cs %*% t(cs), name="C" )
covE      <- mxAlgebra( expression=ec %*% t(ec) + es %*% t(es), name="E" )
```

specific number of independent pathways by source of variance

common factors of size nv x nf

specific factors of size nv x nv (diagonal only

common factors + specifics

Fitting IP Model

```
# Create Model Objects for Multiple Groups
pars      <- list(meanG, matI, invSD,
                  pathAc, pathCc, pathEc, pathAs, pathCs, pathEs, covA, covC, covE, covP, corA, corC, corE)
modelMZ   <- mxModel( name="MZ", pars, covMZ, expCovMZ, dataMZ, expMZ, funML )
modelDZ   <- mxModel( name="DZ", pars, covDZ, expCovDZ, dataDZ, expDZ, funML )
multi     <- mxFitFunctionMultigroup( c("MZ","DZ") )

# Build & Run Model
modelIP   <- mxModel( "mulIPc", pars, modelMZ, modelDZ, multi )
fitIP     <- mxRun( modelIP, intervals=F )
sumIP     <- summary( fitIP )
mxCompare( fitACE, fitIP )
fitGofs(fitIP)

# Generate List
matIPpaths <- c("iSD %*% ac","iSD %*% cc","iSD %*% ec","iSD %*% as","iSD %*% cs","iSD %*% es")
labIPpaths <- c("stPathAc","stPathCc","stPathEc","stPathAs","stPathCs","stPathEs")
formatOutputMatrices(fitIP, matIPpaths, labIPpaths, vars, 4)
```

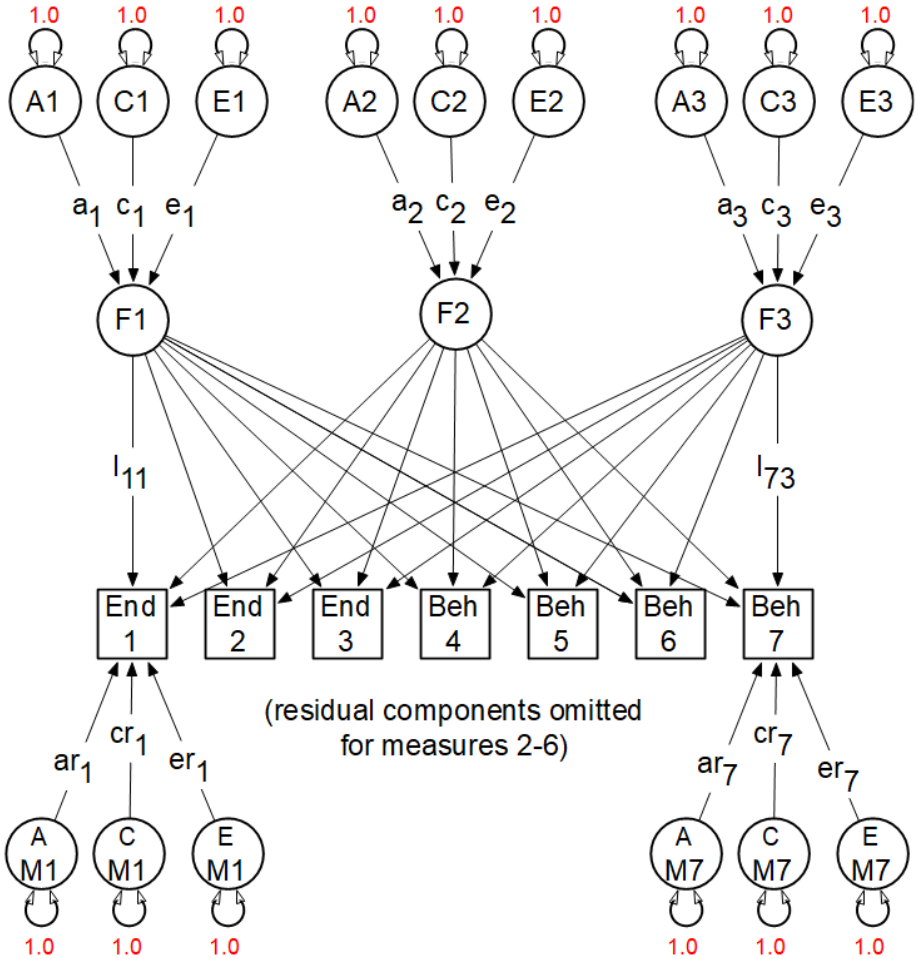
include all relevant matrices

fitted model, list of matrices (in quotes), list of labels (also in quotes), list of variable names, rounding value

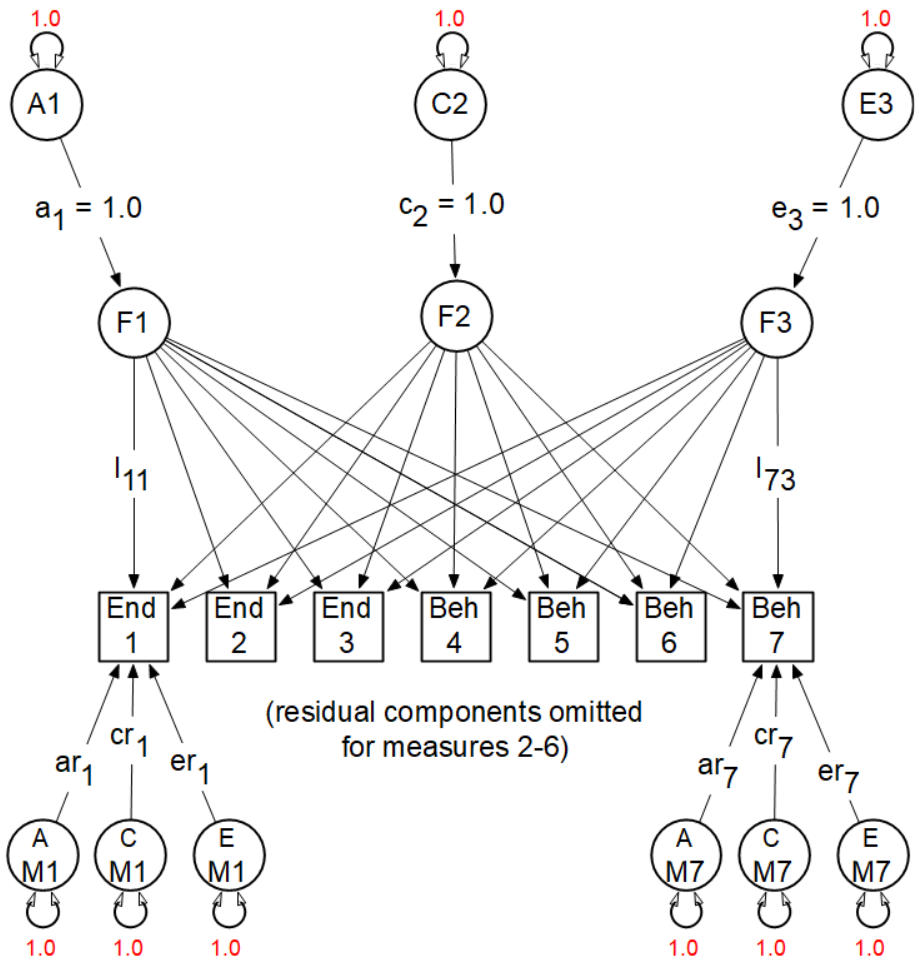
Independent pathway practical

`Faculty/mikeh/2018/Multivariate`

Common Pathway



Independent Pathway



The independent pathway model is nested within the three factor common pathway model.

Scientific questions you can ask

- In *univariate* analyses: what are the contributions of additive genetic, dominance genetic, shared environmental, and unique environmental factors to the variance?
- In *multivariate* analyses: what are these contributions to the **covariance** between two or more traits?

Questions?

- Common Pathway
  - Can you test for a 1 factor vs a 2 factor vs a 3 factor CP?
  - Can you test for every common factor being A and C?
  - Can you test for every specific factor being only E?
  - Can you fit an ADE model?
- Independent Pathway
  - Can you test for a 1 factor vs a 2 factor vs a 3 factor IP?
  - Can you test for every A factor having the same loadings? What does that imply?
  - Can you test for every specific factor being only E?
  - Can you fit an ADE model?

Thank You

mhunter1@ouhsc.edu

Martin, N. G., & Eaves, L. J. (1977, Feb). The genetical analysis of covariance structure. *Heredity*, *38*(1), 79-95. doi: 10.1038/hdy.1977.9