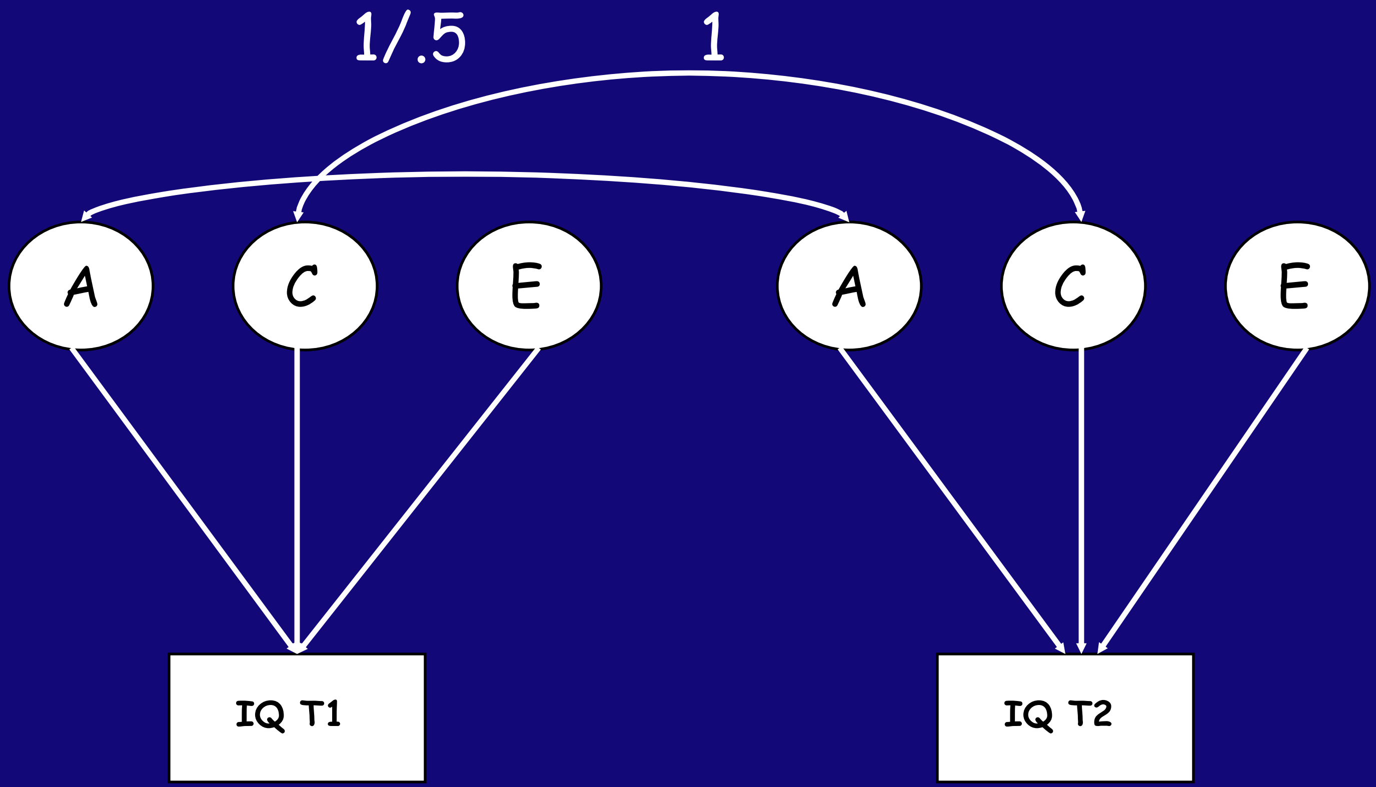
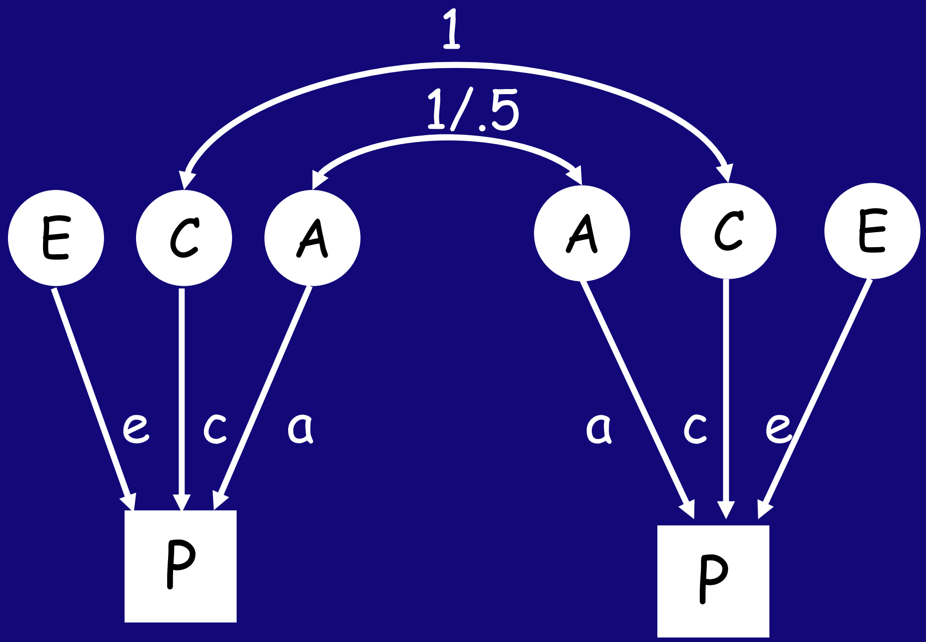
Mx,
Matrix Algebra,
Multivariate Genetic Analyses

Meike Bartels & Kate Morley

Leuven 2008

The Univariate Model



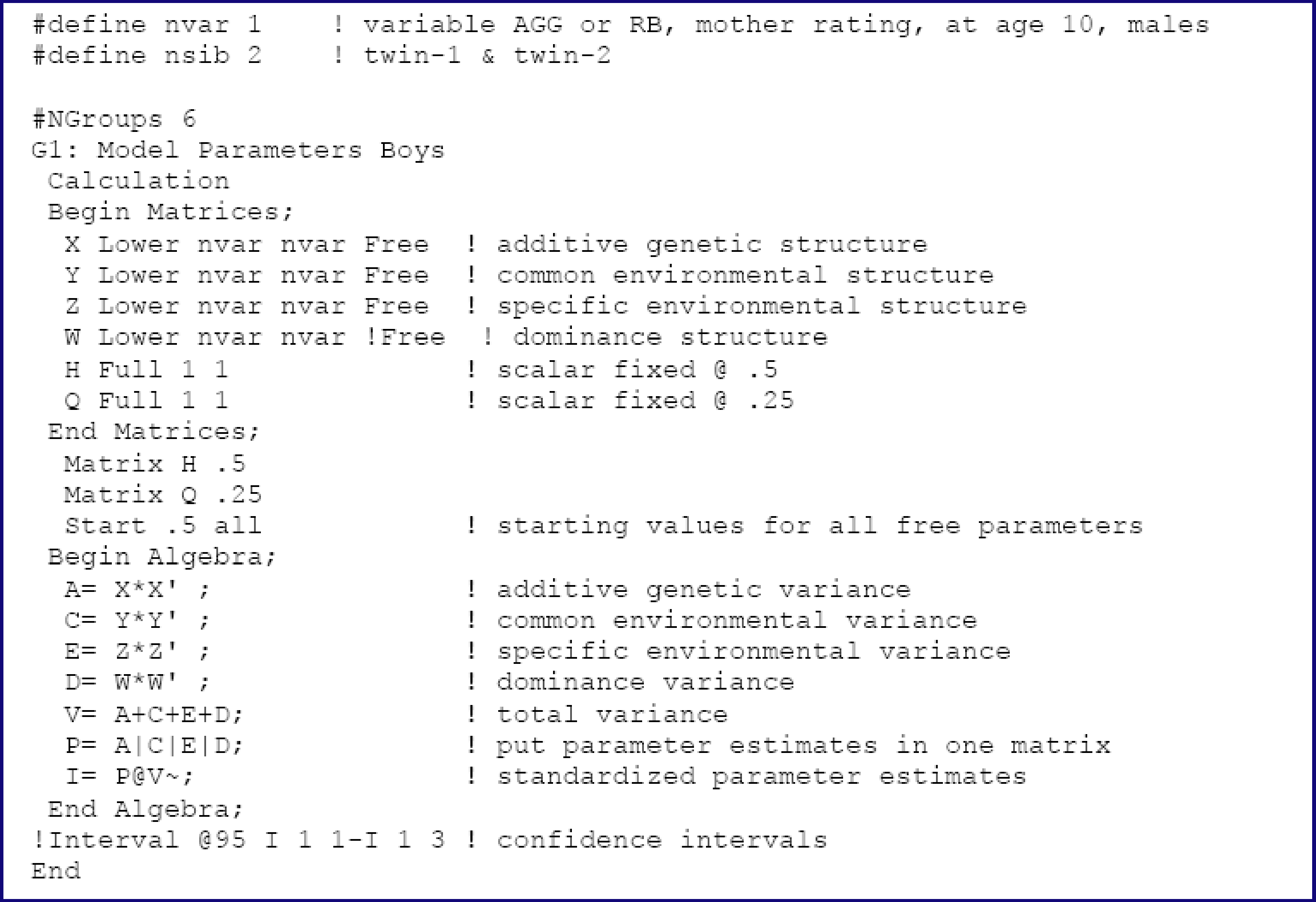


$$\Sigma_{MZ} = \begin{bmatrix} a^2 + c^2 + e^2 & a^2 + c^2 \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

2 × 2

$$\Sigma_{DZ} = \begin{bmatrix} a^2 + c^2 + e^2 & .5a^2 + c^2 \\ .5a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

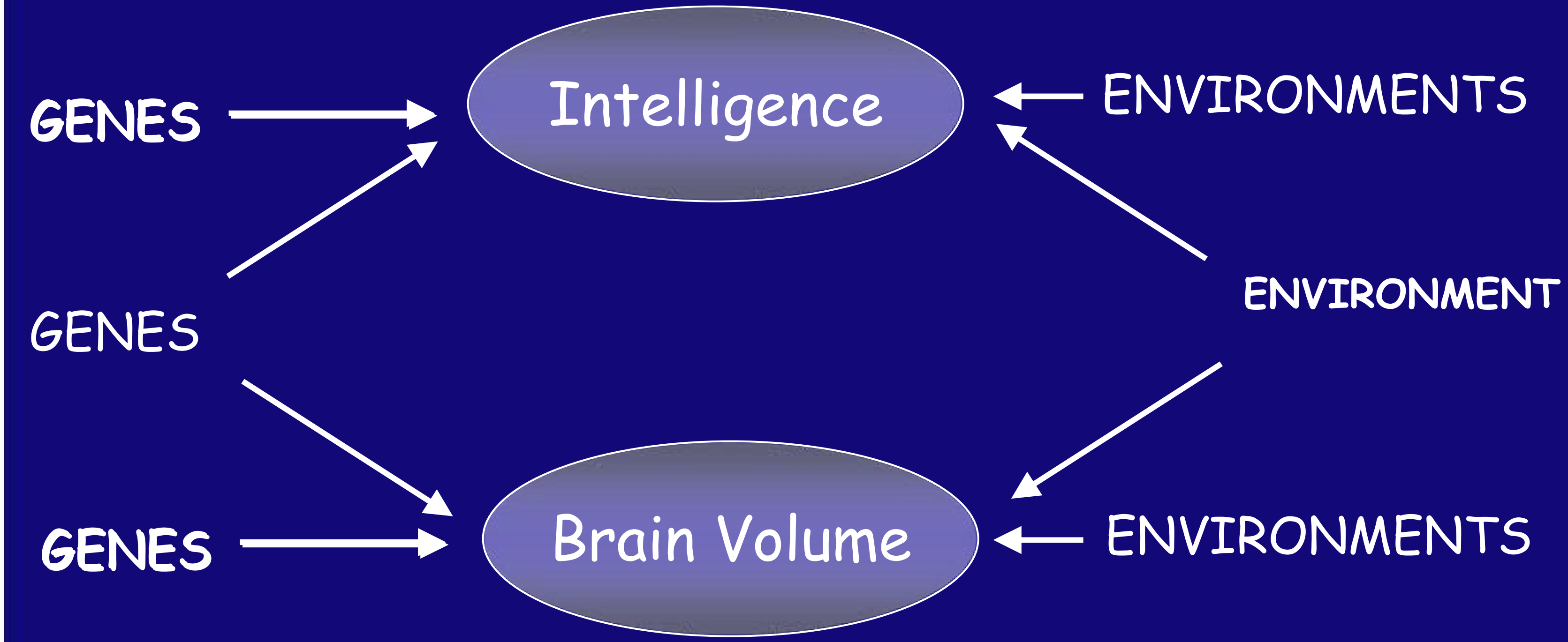
2 × 2



```
#define nvar 1    ! variable AGG or RB, mother rating, at age 10, males
#define nsib 2    ! twin-1 & twin-2

#NGroups 6
G1: Model Parameters Boys
 Calculation
 Begin Matrices;
  X Lower nvar nvar Free  ! additive genetic structure
  Y Lower nvar nvar Free  ! common environmental structure
  Z Lower nvar nvar Free  ! specific environmental structure
  W Lower nvar nvar !Free  ! dominance structure
  H Full 1 1              ! scalar fixed @ .5
  Q Full 1 1              ! scalar fixed @ .25
 End Matrices;
  Matrix H .5
  Matrix Q .25
  Start .5 all            ! starting values for all free parameters
 Begin Algebra;
  A= X*X' ;               ! additive genetic variance
  C= Y*Y' ;               ! common environmental variance
  E= Z*Z' ;               ! specific environmental variance
  D= W*W' ;               ! dominance variance
  V= A+C+E+D;             ! total variance
  P= A|C|E|D;             ! put parameter estimates in one matrix
  I= P@V~;                ! standardized parameter estimates
 End Algebra;
!Interval @95 I 1 1-I 1 3 ! confidence intervals
End
```

Relationships



GENES

Intelligence

ENVIRONMENTS

GENES

ENVIRONMENT

GENES

Brain Volume

ENVIRONMENTS

Multivariate Analyses

- Goal: to understand what factors make sets of variables correlate or co-vary

- Two or more traits can be correlated because they share common genes or common environmental influences

- With twin data on multiple traits it's possible to partition the covariation into it's genetic and environmental components

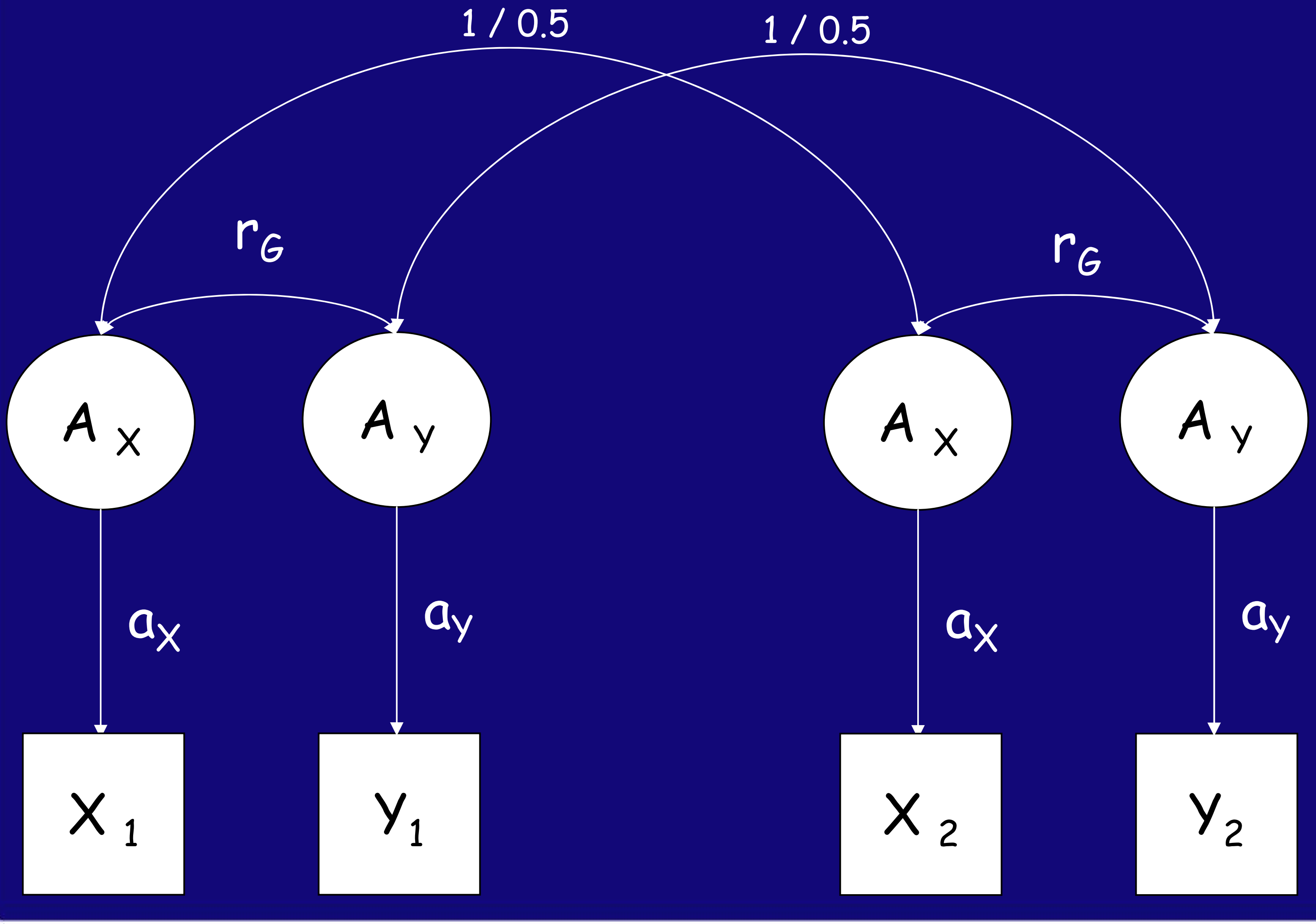
Multivariate Twin Data

		Individual	
		Within	Between
Measure	Within	*Variance*	*Twin covariance*
	Between	*Phenotypic covariance*	*Cross-trait twin covariance*

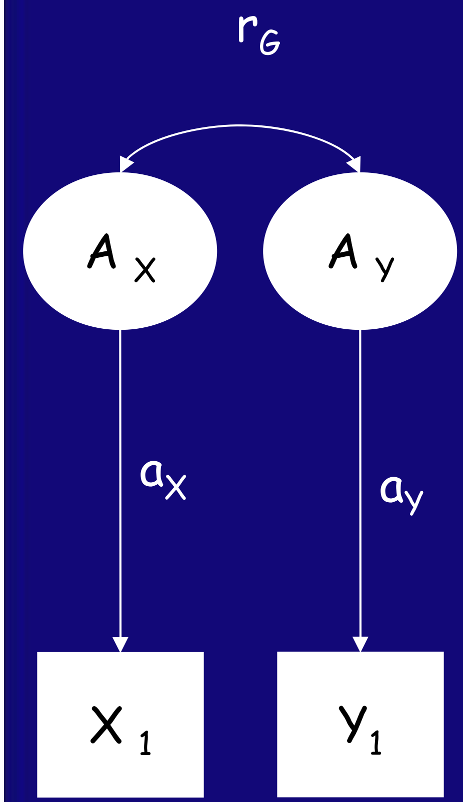
Multivariate Twin Covariance Matrix

$$\begin{array}{c|cccc} & X_1 & Y_1 & X_2 & Y_2 \\ \hline X_1 & V_{X1} & C_{X1Y1} & C_{X1X2} & C_{X1Y2} \\ Y_1 & C_{X1Y1} & V_{Y1} & C_{Y1X2} & C_{Y1Y2} \\ X_2 & C_{X2X1} & C_{X2Y1} & V_{X2} & C_{X2Y2} \\ Y_2 & C_{X1Y2} & C_{Y2Y1} & C_{Y2X2} & V_{Y2} \end{array}$$

Multivariate Model Parameters



Correlated Factors

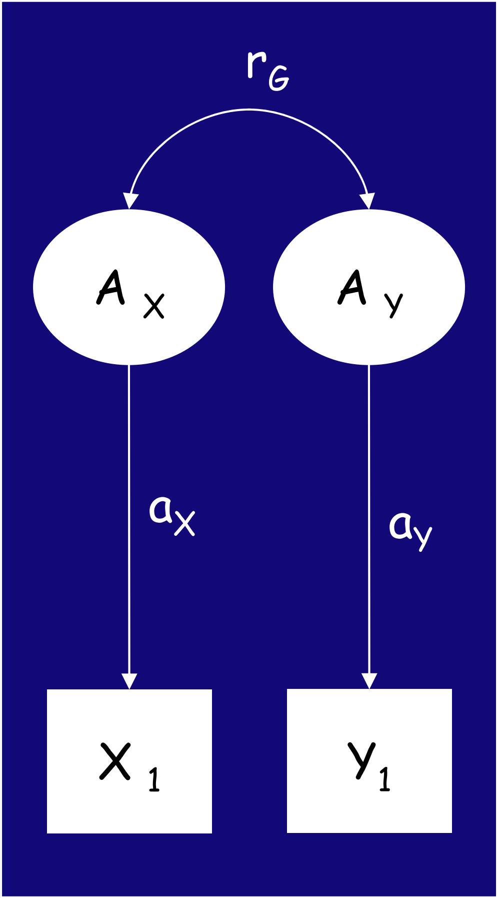


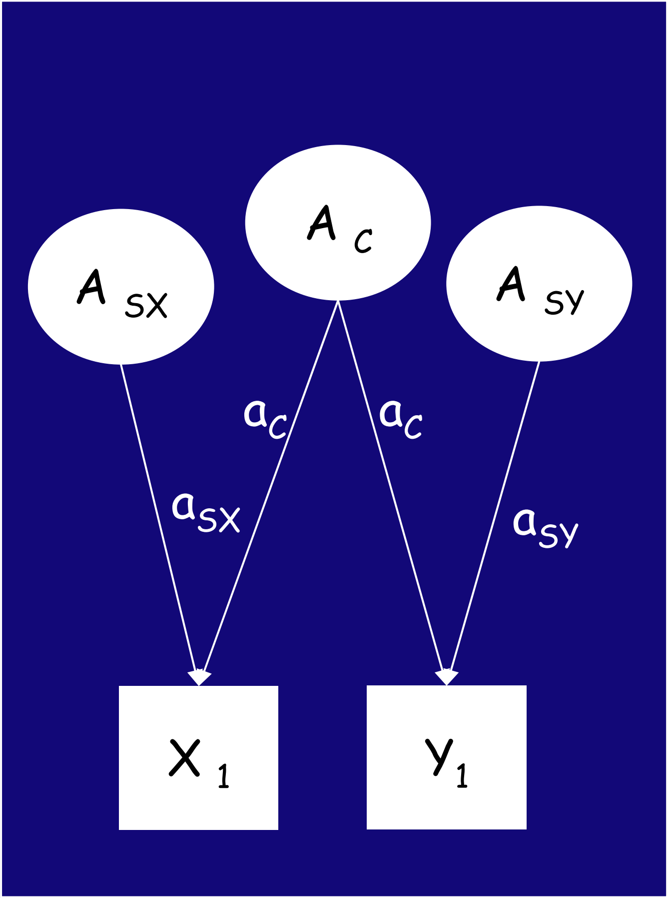
- Genetic correlation r_G
- Chain of paths

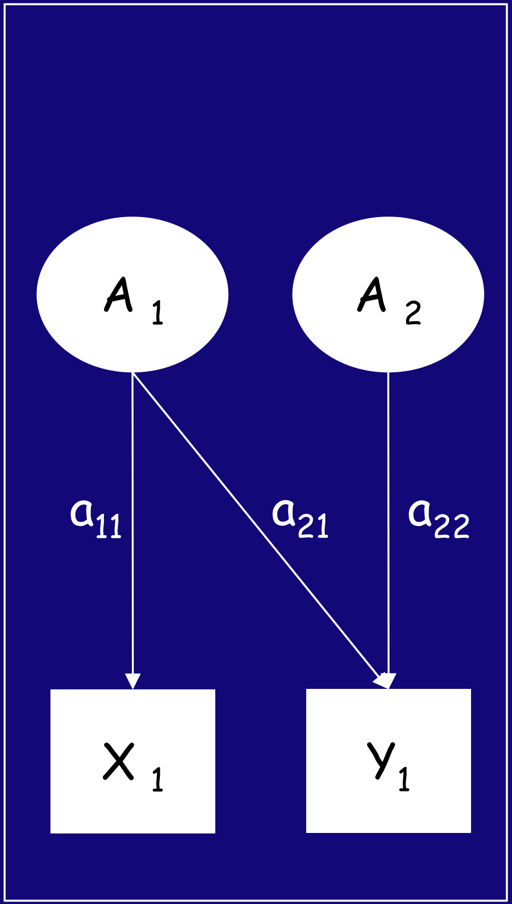
  $a_X r_G a_Y$ *bivariate heritability*
- Component of phenotypic covariance

  $r_{XY} = a_X r_G a_Y + c_X r_C c_Y + e_X r_E e_Y$

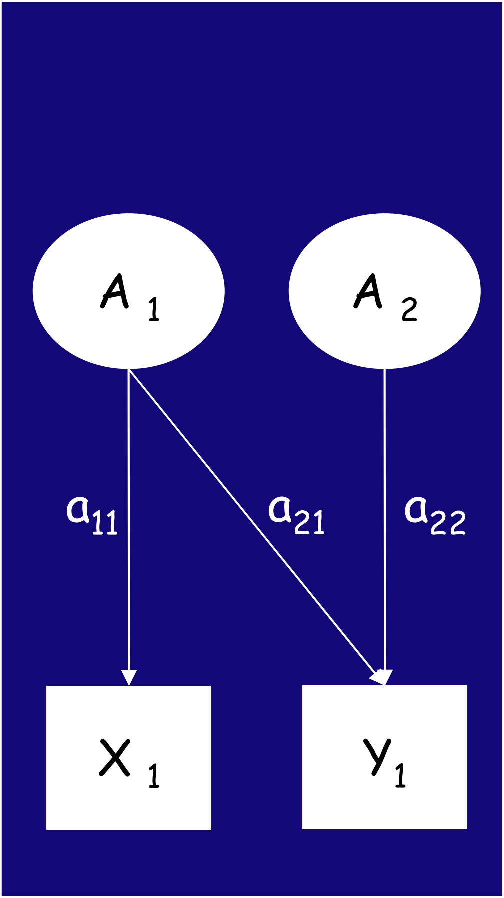
Cholesky Decomposition







Cholesky Decomposition

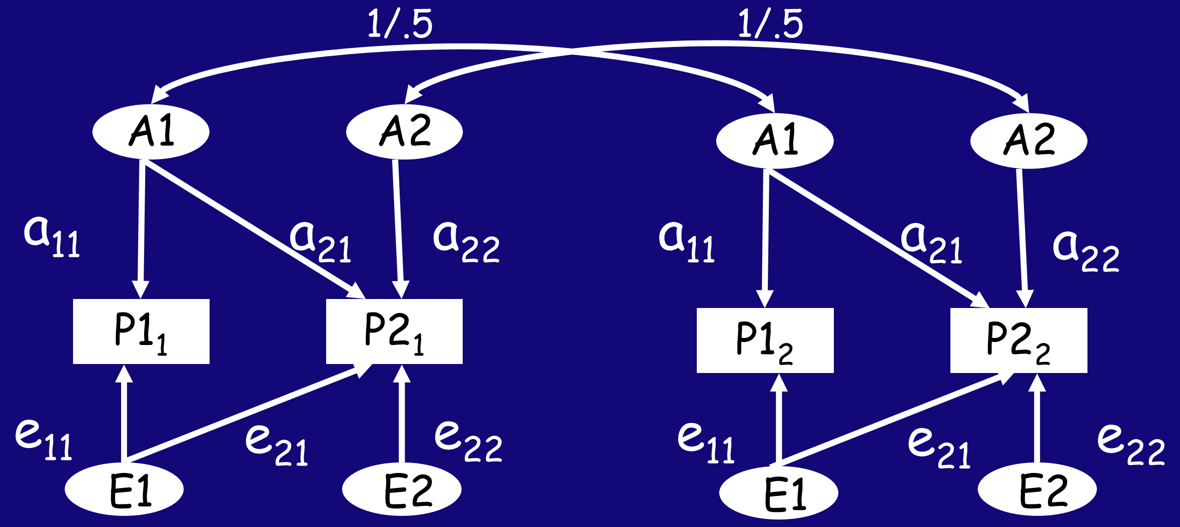


Tests of specificity

- If $a_{22} > 0$
  - genetic influences specific to Y

Multivariate Twin Covariance Matrix

$$\begin{array}{c|cccc}
 & X_1 & Y_1 & X_2 & Y_2 \\
\hline
X_1 & V_{X1} & C_{X1Y1} & C_{X1X2} & C_{X1Y2} \\
Y_1 & C_{X1Y1} & V_{Y1} & C_{Y1X2} & C_{Y1Y2} \\
X_2 & C_{X2X1} & C_{X2Y1} & V_{X2} & C_{X2Y2} \\
Y_2 & C_{X1Y2} & C_{Y2Y1} & C_{Y2X2} & V_{Y2}
\end{array}$$



		Twin1		**Twin2**	
		p1	p2	p1	p2
		Within-Twin Covariances		Cross-Twin Covariances	
Twin1	p1	$a_{11}^2 + e_{11}^2$	$a_{11}*a_{21}+$ $e_{11}*e_{21}$	$1/.5*a_{11}^2$	$1/.5*a_{11}*a_{21}$
	p2	$a_{21}*a_{11}+$ $e_{21}*e_{11}$	$a_{22}^2+a_{21}^2+$ $e_{22}^2+e_{21}^2$	$1/.5*a_{21}*a_{11}$	$1/.5*a_{22}^2+$ $1/.5*a_{21}^2$
		Cross-Twin Covariances		Within-Twin Covariances	
Twin2	p1				
	p2				

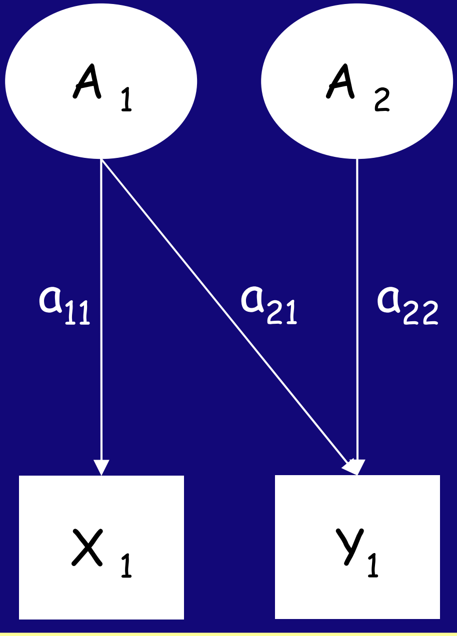
Summary : Cross-traits covariances

- Within-individual cross-traits covariances implies common etiological influences

- Cross-twin cross-traits covariances implies that these common etiological influences are familial

- Whether these common familial etiological influences are genetic or environmental, is reflected in the MZ/DZ ratio of the cross-twin cross-traits covariances

Specification in Mx



$$\Sigma_A = \begin{bmatrix} a^{2}_{11} & a_{11}a_{21} \\ a_{21}a_{11} & a^{2}_{21} + a^{2}_{22} \end{bmatrix}$$

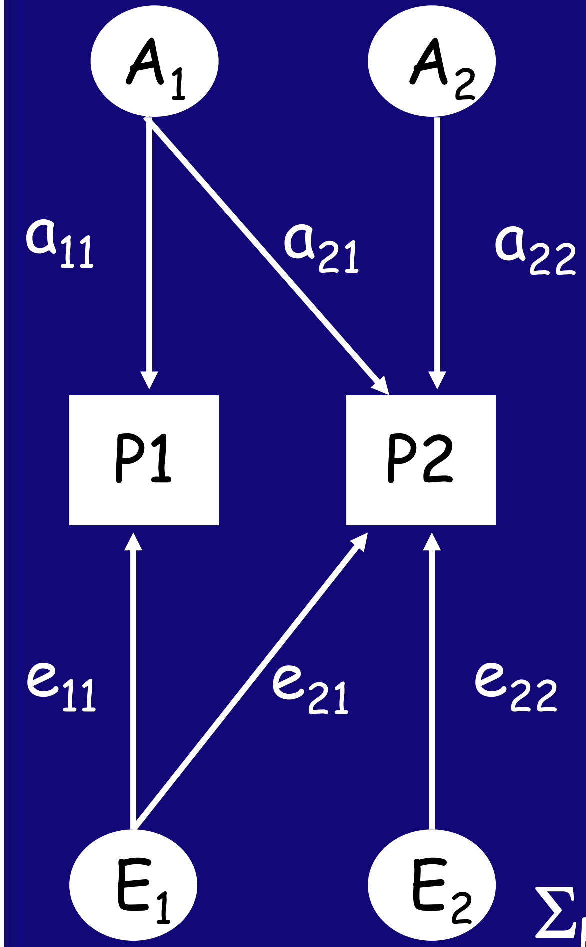
A LOWER 2 × 2

$$\begin{array}{c|cc} & A1 & A2 \\ \hline P1 & a_{11} & 0 \\ P2 & a_{21} & a_{22} \end{array}$$

*** or 'Star' Matrix Multiplication**

$$\Sigma_A = A*A' = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^{2} + 0*0 & a_{11}a_{21} + 0*a_{22} \\ a_{21}a_{11} + a_{22}*0 & a_{21}^{2} + a_{22}^{2} \end{bmatrix}$$

Within-Twin Covariances: Specification in Mx



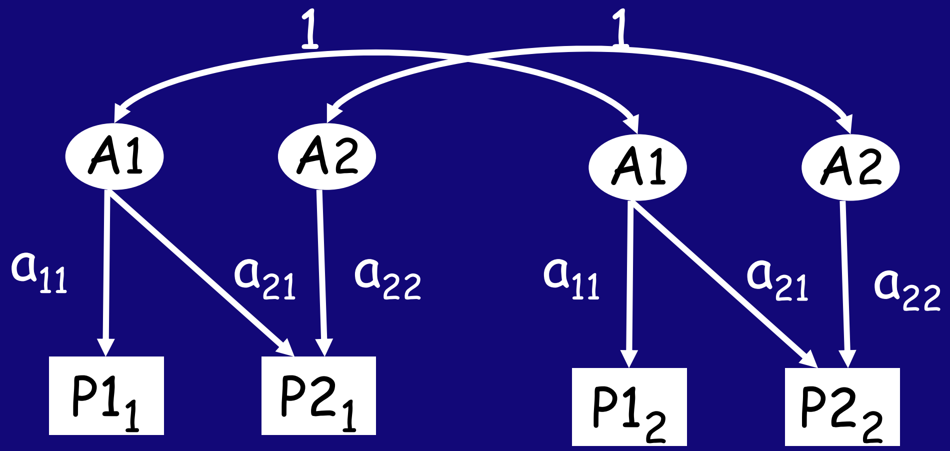
$$\Sigma_A = \begin{bmatrix} a^{2}{}_{11} & a_{11}a_{21} \\ a_{21}a_{11} & a^{2}{}_{21} + a^{2}{}_{22} \end{bmatrix}$$

$$\Sigma_E = \begin{bmatrix} e^{2}{}_{11} & e_{11}e_{21} \\ e_{21}e_{11} & e^{2}{}_{21} + e^{2}{}_{22} \end{bmatrix}$$

$$\boxed{\Sigma_P = \Sigma_A + \Sigma_E}$$

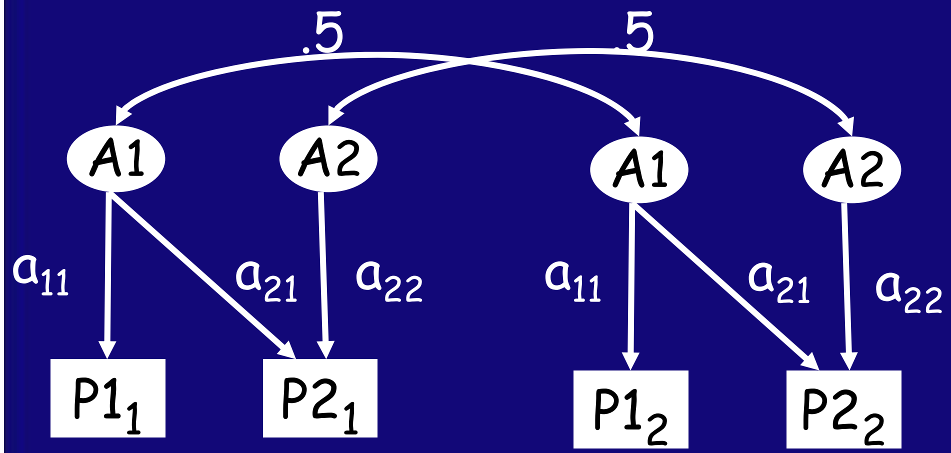
$$\Sigma_P = \begin{pmatrix} a_{11}^2 + e_{11}^2 & a_{11}\,a_{21} + e_{11}\,e_{21} \\ a_{21}\,a_{11} + e_{21}\,e_{11} & a_{21}^2 + a_{22}^2 + a_{21}^2 + a_{22}^2 \end{pmatrix}$$

Cross-Twin Covariances: Specification in Mx (MZ)



$$\Sigma_A = A * A' = \begin{bmatrix} a^{2}11 & a_{11}a_{21} \\ a_{21}a_{11} & a^{2}21 + a^{2}22 \end{bmatrix}$$

Cross-Twin Covariances: Specification in Mx (DZ)



Within-Traits (diagonals):

P_{11}-P_{12}= $a_{11} \times .5 \times a_{11}$

P_{21}-P_{22}= $a_{22} \times .5 \times a_{22} + a_{21} \times .5 \times a_{21}$

Cross-Traits:

P_{11}-P_{22}= $a_{11} \times .5 \times a_{21}$

P_{21}-P_{12}= $a_{21} \times .5 \times a_{11}$

Kronecker Product ⊗

$$.5 \otimes \Sigma_A = .5 \otimes A * A' = \begin{bmatrix} .5a^2{}_{11} & .5a_{11}\,a_{21} \\ .5a_{21}a_{11} & .5a^2{}_{21} + a^2{}_{22} \end{bmatrix}$$

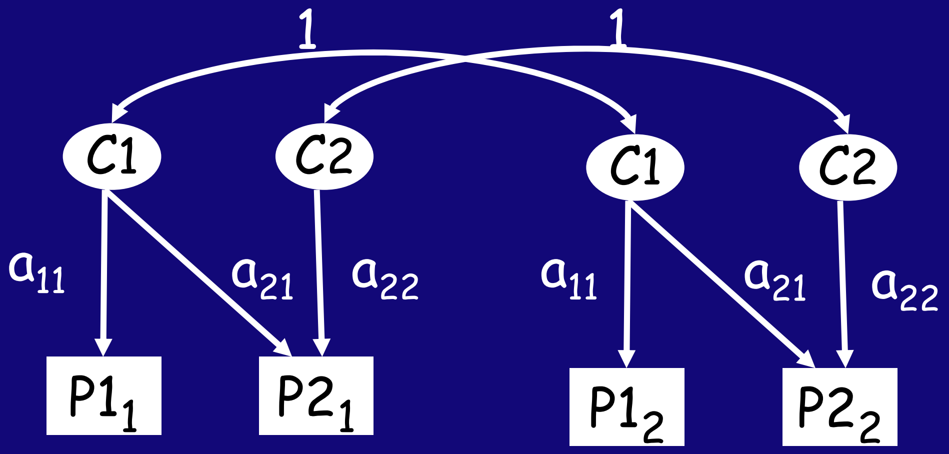
Kronecker Product ⊗

H FULL 1 1
MATRIX H .5

$\otimes = @$

$$.5 \otimes \Sigma_{A} = H@A*A' = \begin{bmatrix} .5a^{2}{}_{11} & .5a_{11}\,a_{21} \\ .5a_{21}a_{11} & .5a^{2}{}_{21} + a^{2}{}_{22} \end{bmatrix}$$

Cross-Twin Covariances: (MZ/DZ)



$$1 \otimes \Sigma_{C} = \quad C * C' = \begin{bmatrix} c^{2}11 & c_{11}c_{21} \\ c_{21}c_{11} & c^{2}21 + c^{2}22 \end{bmatrix}$$

MX

MZ

```
COV     A+C+E  | A+C_
        A+C    | A+C+E /
```

DZ

```
COV     A+C+E  | H@A+C_
        H@A+C  | A+C+E /
```

Practical Session

MRI-IQ dataset

WM = working memory

BBGM = gray matter volume of the cerebrum

BBWM = white matter volume of the cerebrum

(Variables have already been corrected for the effects of age and sex)

Mx job: bivariate_WM+GM.mx (F:\meike\multivariate)

Data file: mri.dat (F:\meike\multivariate)

Posthuma et al. Nature Neuroscience, Feb 2002

Practical Session

MRI-IQ dataset

Phenotypic correlations

Working memory – gray matter volume	0.27
Working memory – white matter volume	0.28

Practical Session

MRI-IQ dataset

Twin correlations

	WM	BBGM	BBWM
MZ	0.72	0.86	0.89
DZ	0.27	0.45	0.34

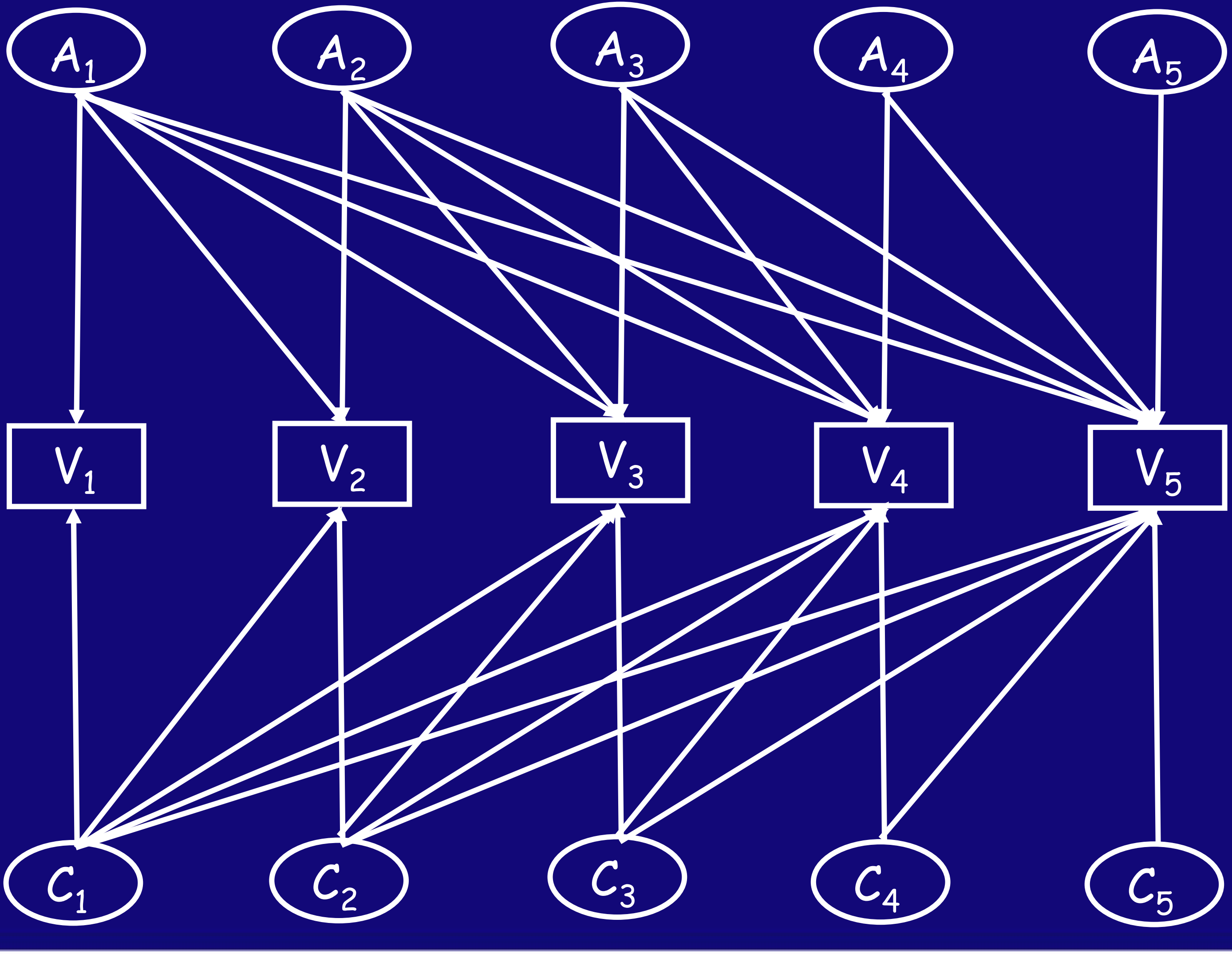
Practical Session

- Run Bivariate Model
- Modify Mx Job to a trivariate model

Practical Session

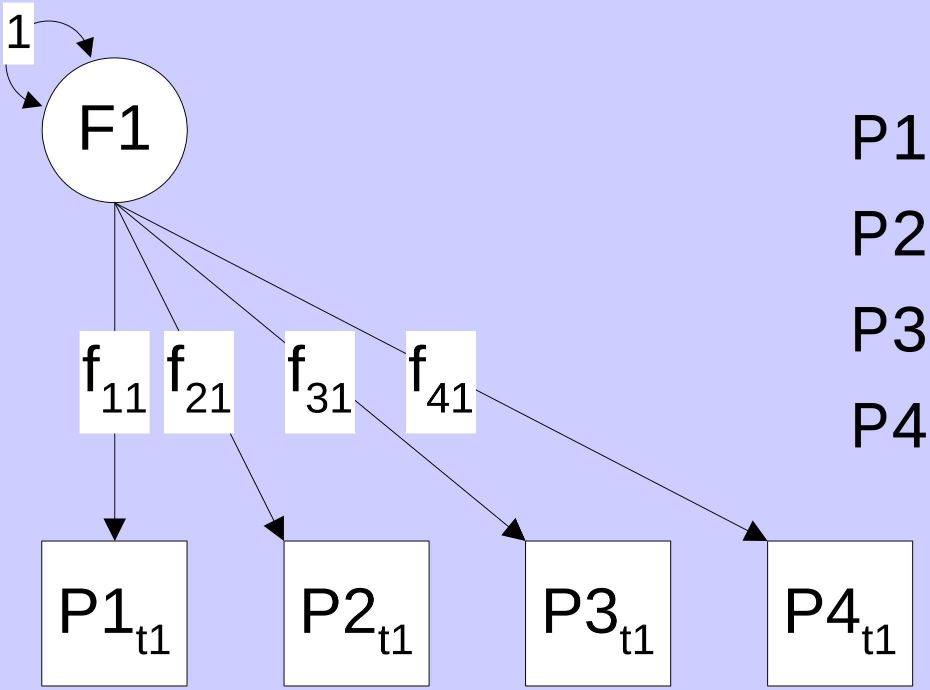
- Do the genes that influence WM also influence Brain Volume?
- Are there genes that are unique to Brain Volume?
- Is the phenotypic correlation caused by genetic correlation?
- What is the genetic correlation?
- The same questions apply to environmental (shared and unique) influences.

More than two variables



More than two variables

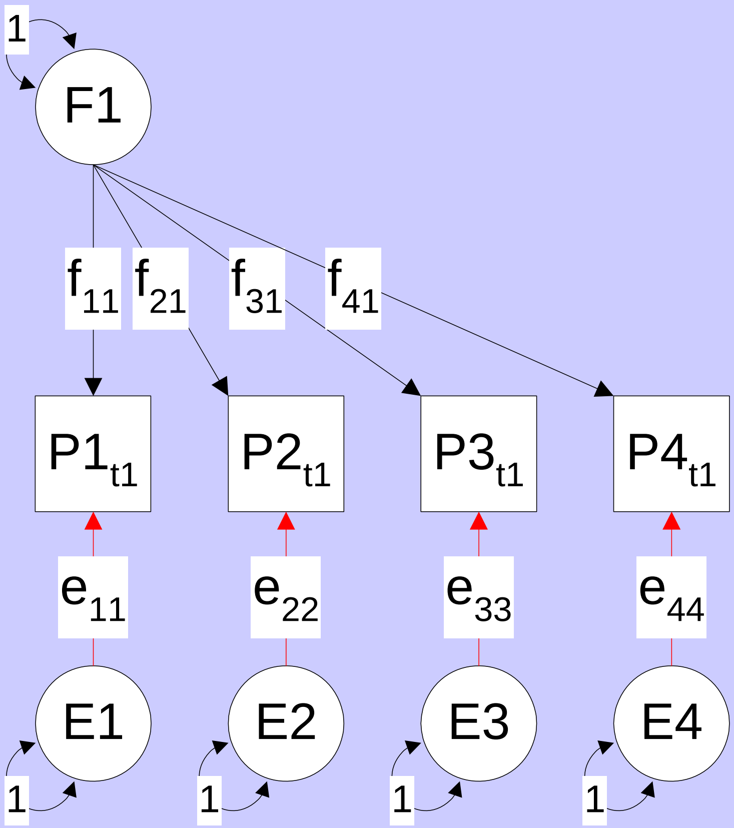
Phenotypic Single Factor



$$\begin{array}{cc} & F1 \\ \begin{matrix} P1 \\ P2 \\ P3 \\ P4 \end{matrix} & \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} \end{array} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} \end{bmatrix}$$

$$\mathbf{F} * \mathbf{F'}$$

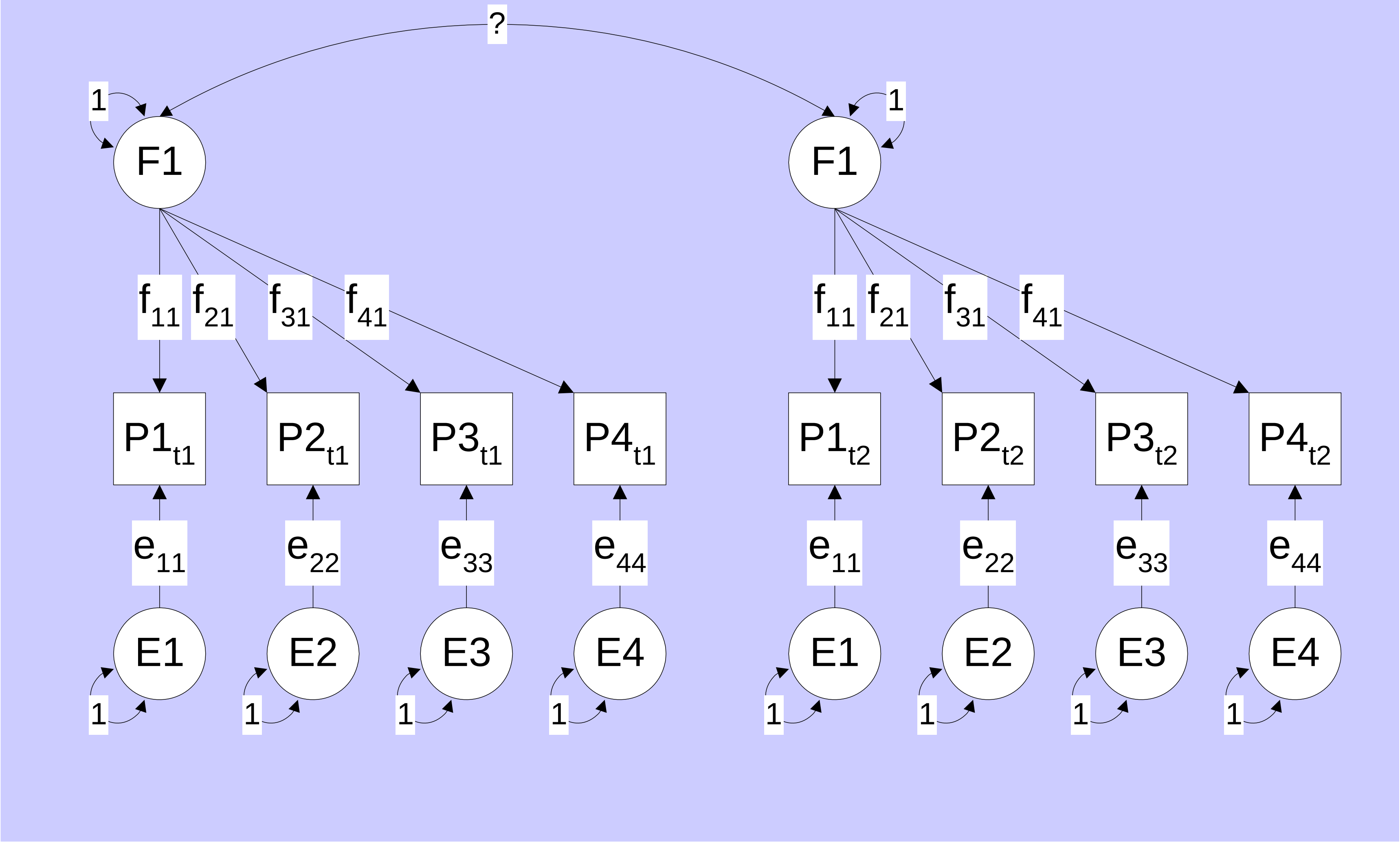
Residual Variances



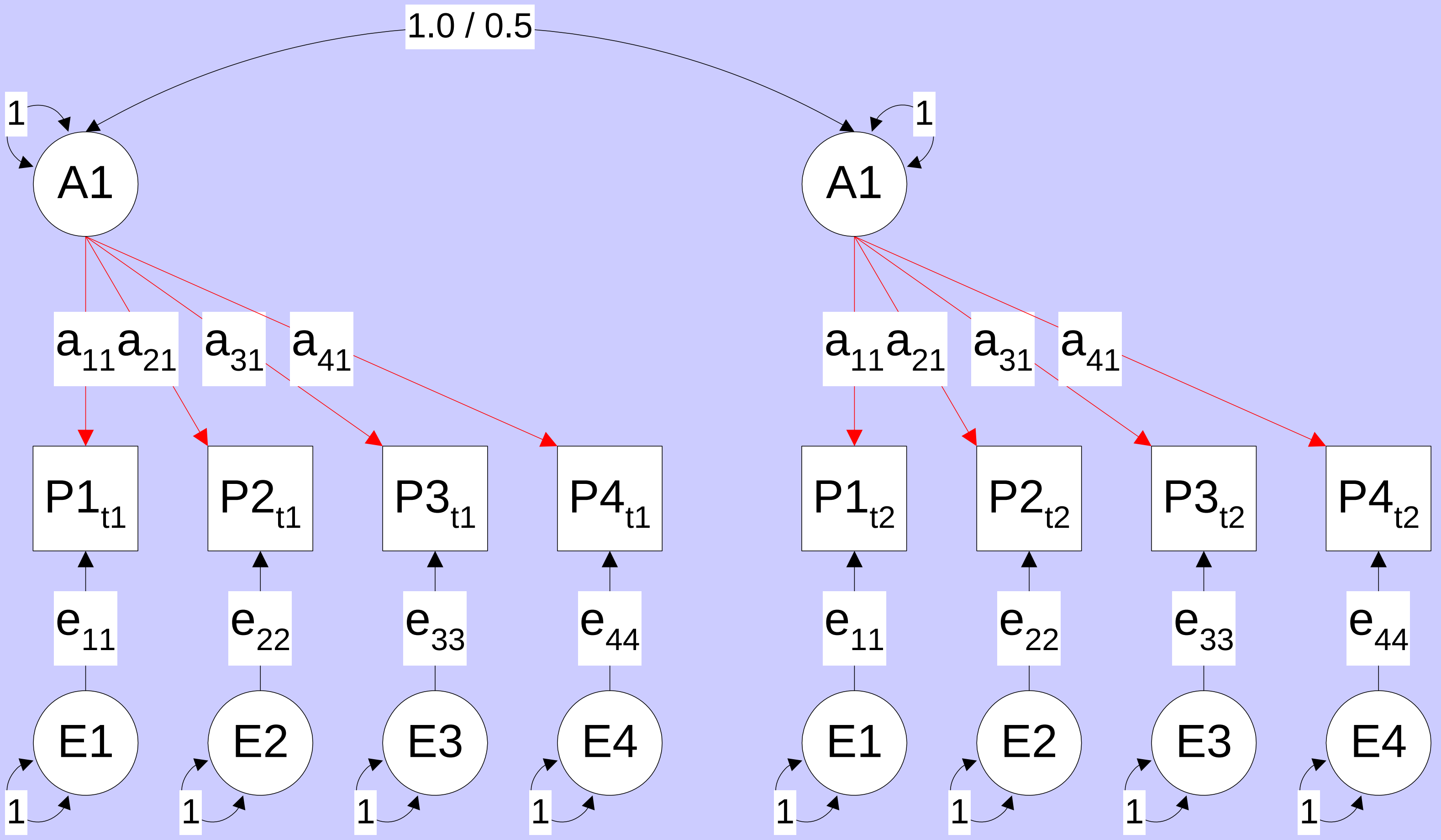
$$\begin{array}{c|cccc|} & E1 & E2 & E3 & E4 \\ P1 & e_{11} & 0 & 0 & 0 \\ P2 & 0 & e_{22} & 0 & 0 \\ P3 & 0 & 0 & e_{33} & 0 \\ P4 & 0 & 0 & 0 & e_{44} \end{array} * \begin{bmatrix} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{bmatrix}$$

$$\mathbf{E} \quad * \quad \mathbf{E'}$$

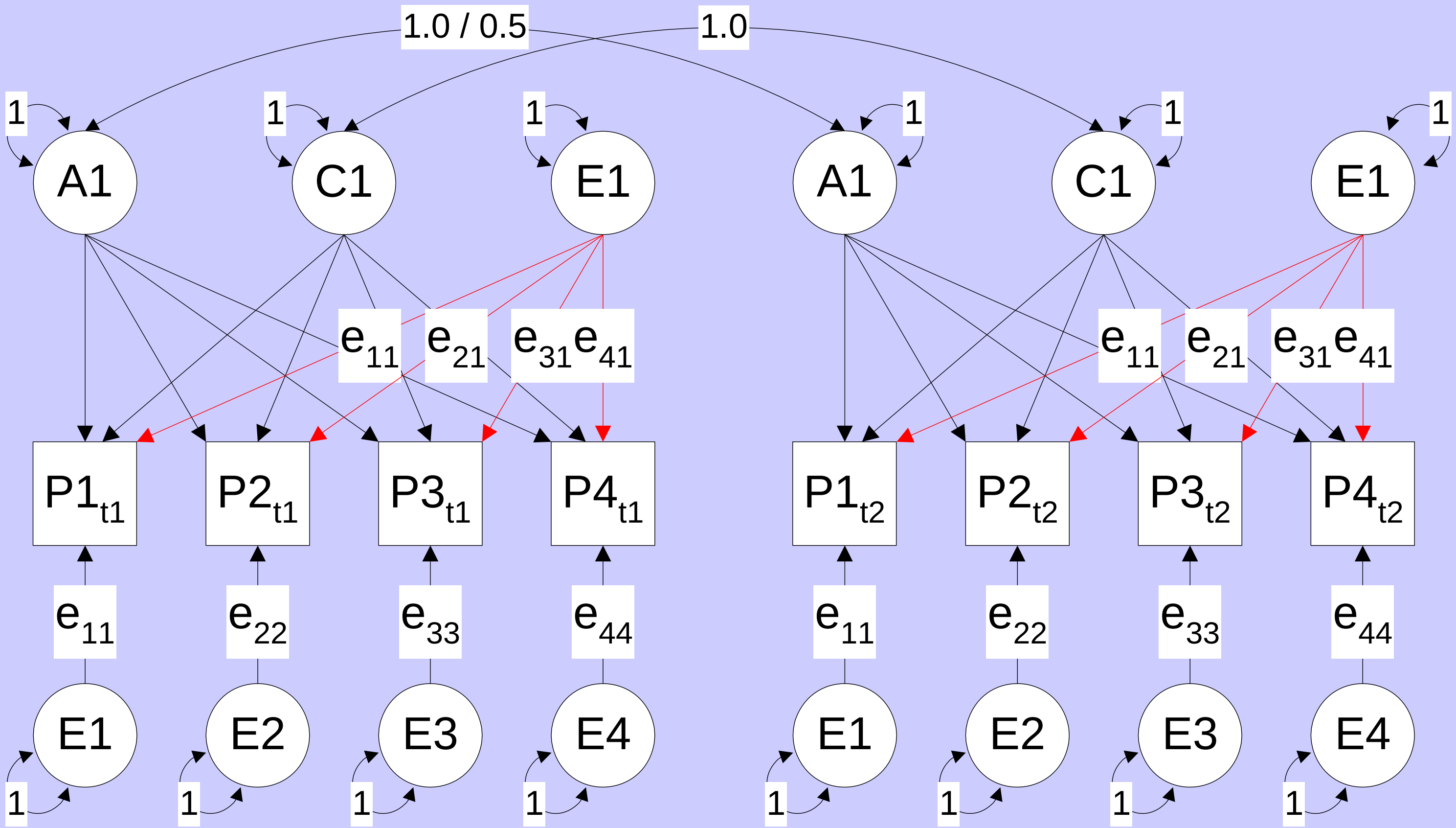
Twin Data


?
1
F1
1
F1
f11
f21
f31
f41
f11
f21
f31
f41
P1t1
P2t1
P3t1
P4t1
P1t2
P2t2
P3t2
P4t2
e11
e22
e33
e44
e11
e22
e33
e44
E1
E2
E3
E4
E1
E2
E3
E4
1
1
1
1
1
1
1
1

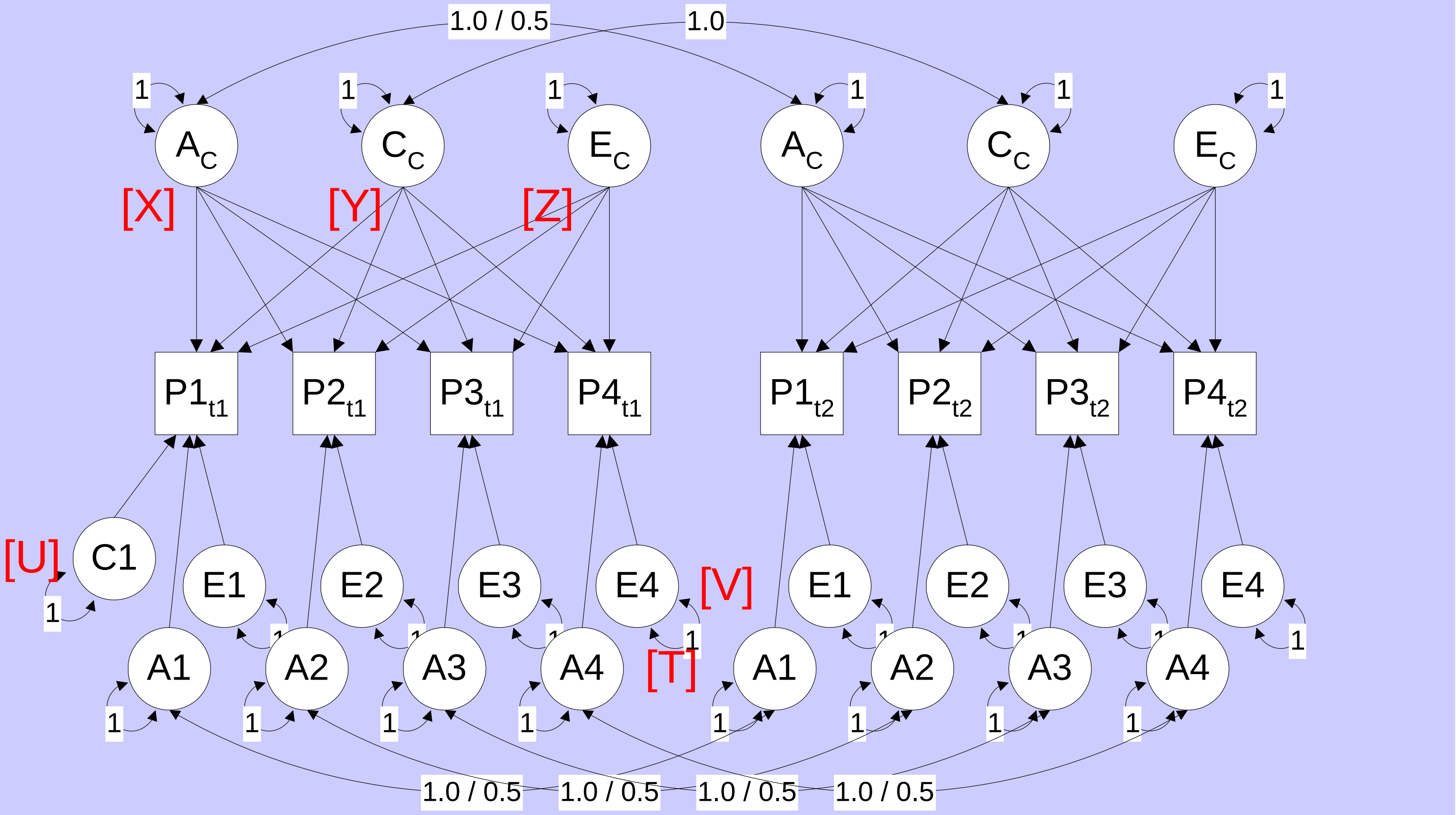
Genetic Single Factor



ACE all Single Factors



Independent Pathway Model



Independent Pathway Model

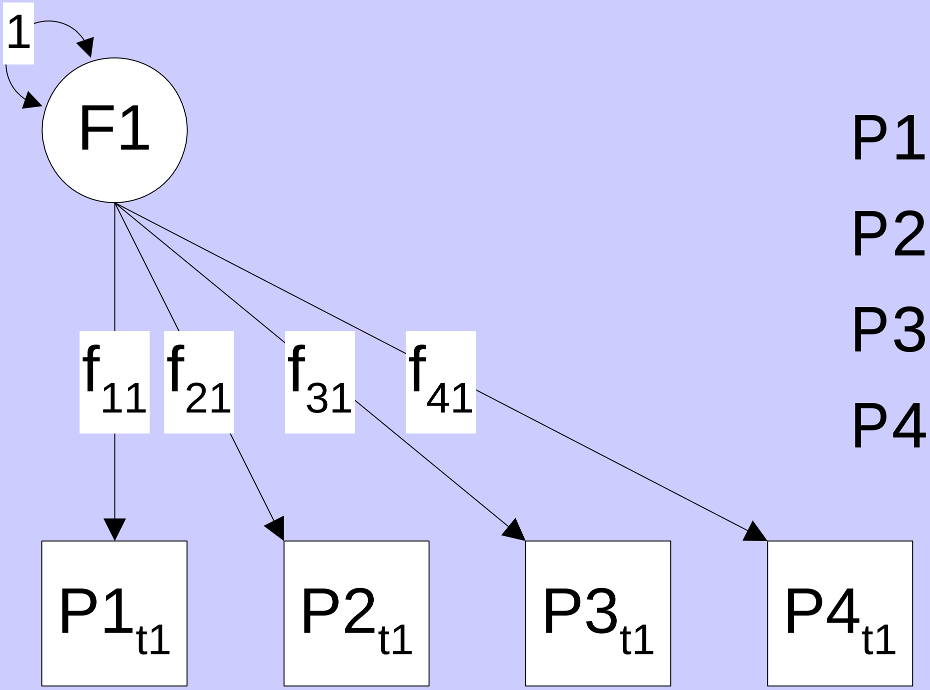
```
G1: Define matrices
 Calculation
 Begin Matrices;
  X full nvar nfac Free     ! common factor genetic path coefficients
  Y full nvar nfac Free     ! common factor shared environment paths
  Z full nvar nfac Free     ! common factor unique environment paths
  T diag nvar nvar Free     ! variable specific genetic paths
  U diag nvar nvar Free     ! variable specific shared env paths
  V diag nvar nvar Free     ! variable specific residual paths
  M full 1 nvar Free        ! means
 End Matrices;
  Start …
 Begin Algebra;
  A= X*X' + T*T';           ! additive genetic variance components
  C= Y*Y' + U*U';           ! shared environment variance components
  E= Z*Z' + V*V';           ! nonshared environment variance components
 End Algebra;
End
```

Independent Pathway Model

```
G2: MZ twins
#include iqnlmz.dat
 Begin Matrices = Group 1;
 Means M | M ;
 Covariance A+C+E | A+C    _
            A+C   | A+C+E  ;
 Option Rsiduals
End

G3: DZ twins
#include iqnldz.dat
 Begin Matrices= Group 1;
  H full 1 1
 End Matrices;
  Matrix H .5
 Means M | M ;
 Covariance A+C+E | H@A+C _
            H@A+C | A+C+E  ;
 Option Rsiduals
End
```

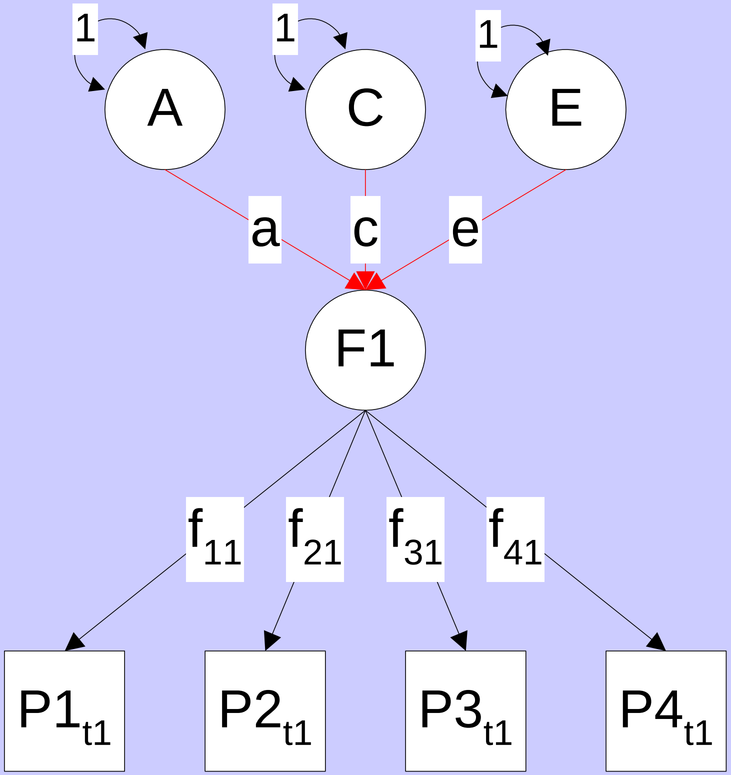
Phenotypic Single Factor



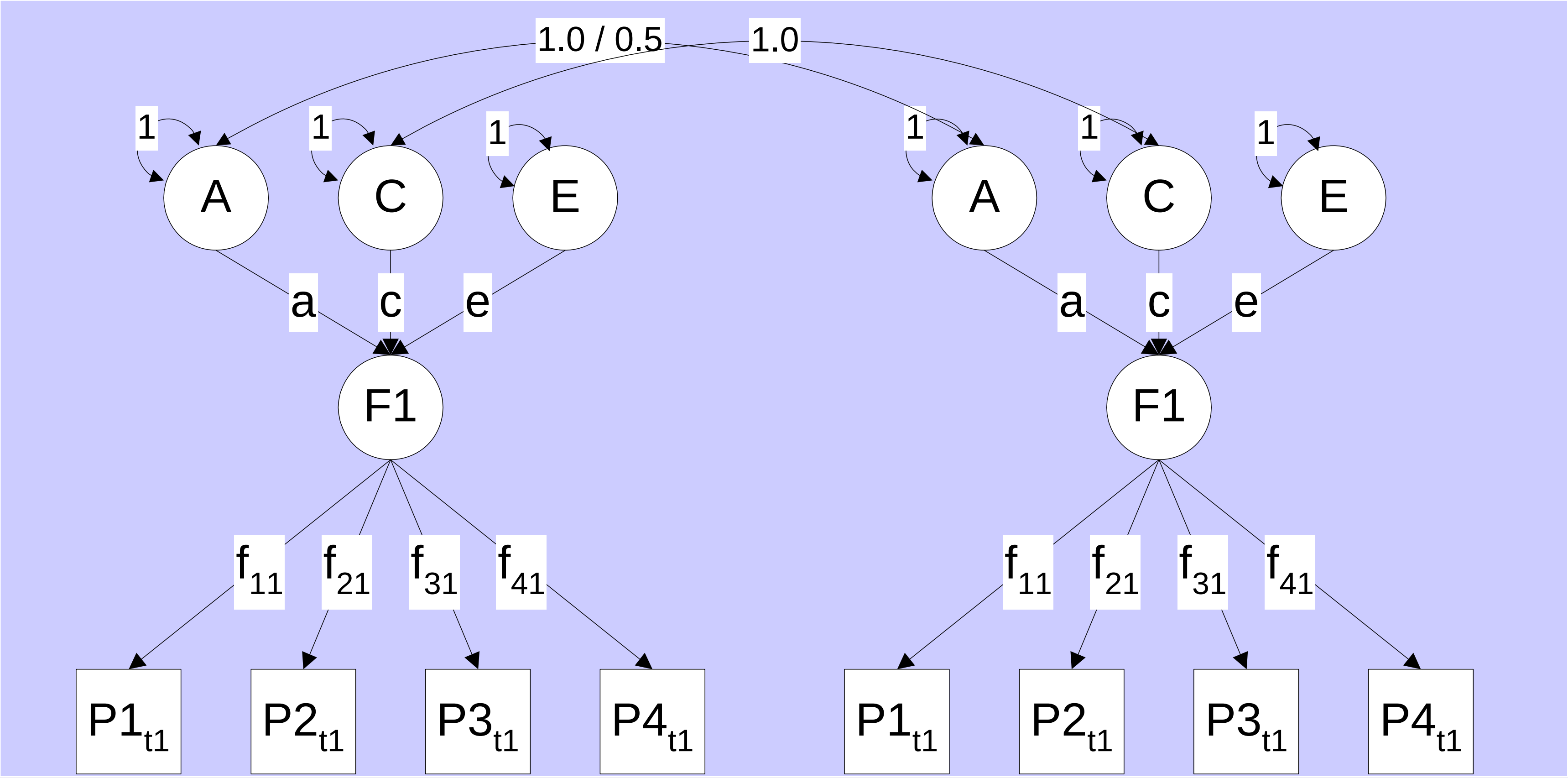
$$\begin{array}{cc} & F1 \\ \begin{array}{c} P1 \\ P2 \\ P3 \\ P4 \end{array} & \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} \end{array} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} \end{bmatrix}$$

$$\mathbf{F} * \mathbf{F'}$$

Latent Phenotype



Twin Data


1.0 / 0.5
1.0
1
1
1
1
1
1
A
C
E
A
C
E
a
c
e
a
c
e
F1
F1
f_{11}
f_{21}
f_{31}
f_{41}
f_{11}
f_{21}
f_{31}
f_{41}
$P1_{t1}$
$P2_{t1}$
$P3_{t1}$
$P4_{t1}$
$P1_{t1}$
$P2_{t1}$
$P3_{t1}$
$P4_{t1}$

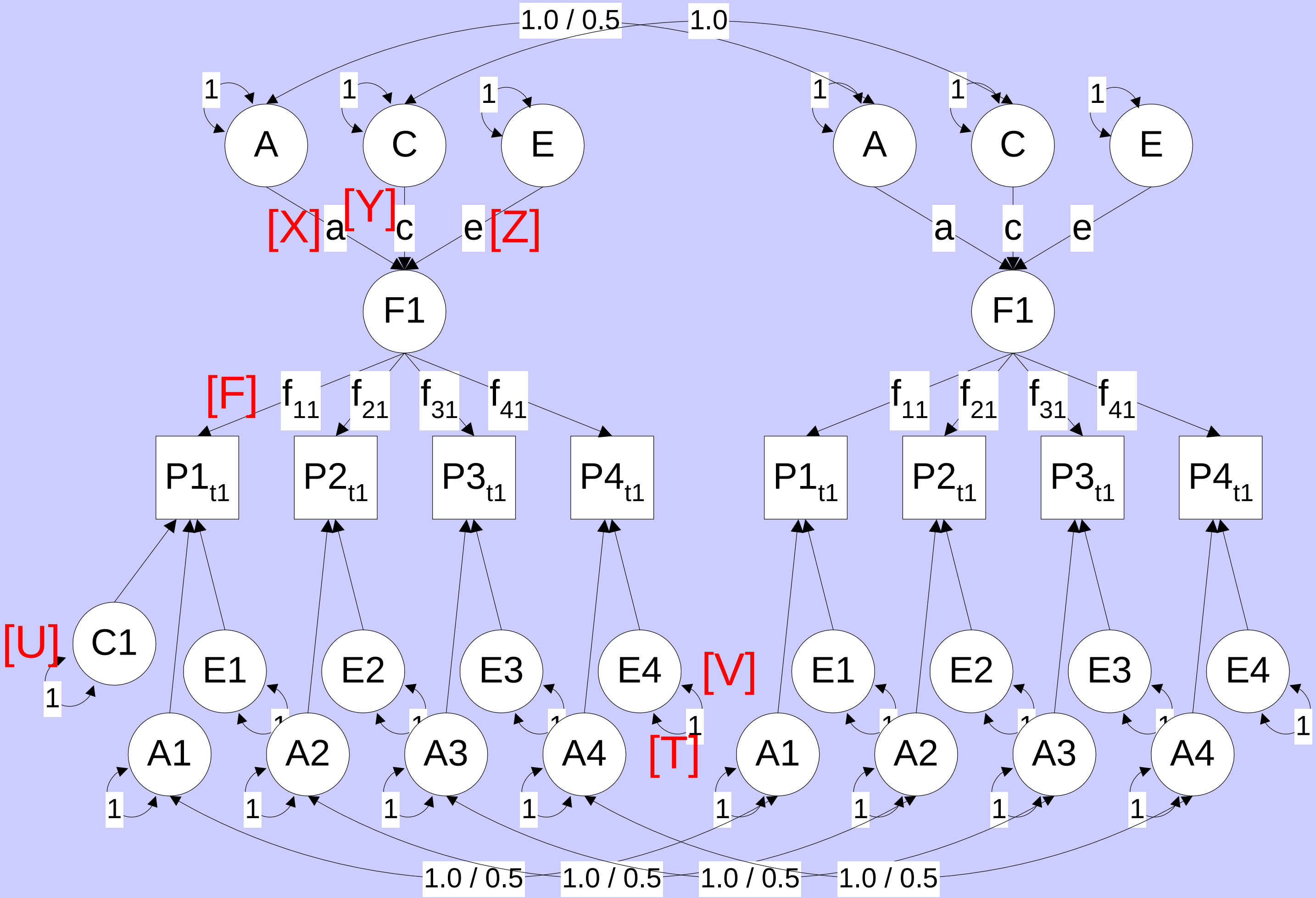
Factor on Latent Phenotype

$$\begin{array}{c} \\ P1 \\ P2 \\ P3 \\ P4 \end{array}\begin{array}{c} F1 \\ \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} \end{array} * \begin{bmatrix} a \end{bmatrix} * \begin{bmatrix} a \end{bmatrix} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} \end{bmatrix}$$

$$\mathbf{F} \; * \; \mathbf{X} \; * \; \mathbf{X'} \; * \; \mathbf{F'}$$

$$= \mathbf{F} \;\&\; (\, \mathbf{X} * \mathbf{X'} \,)$$

Common Pathway Model



Common Pathway Model

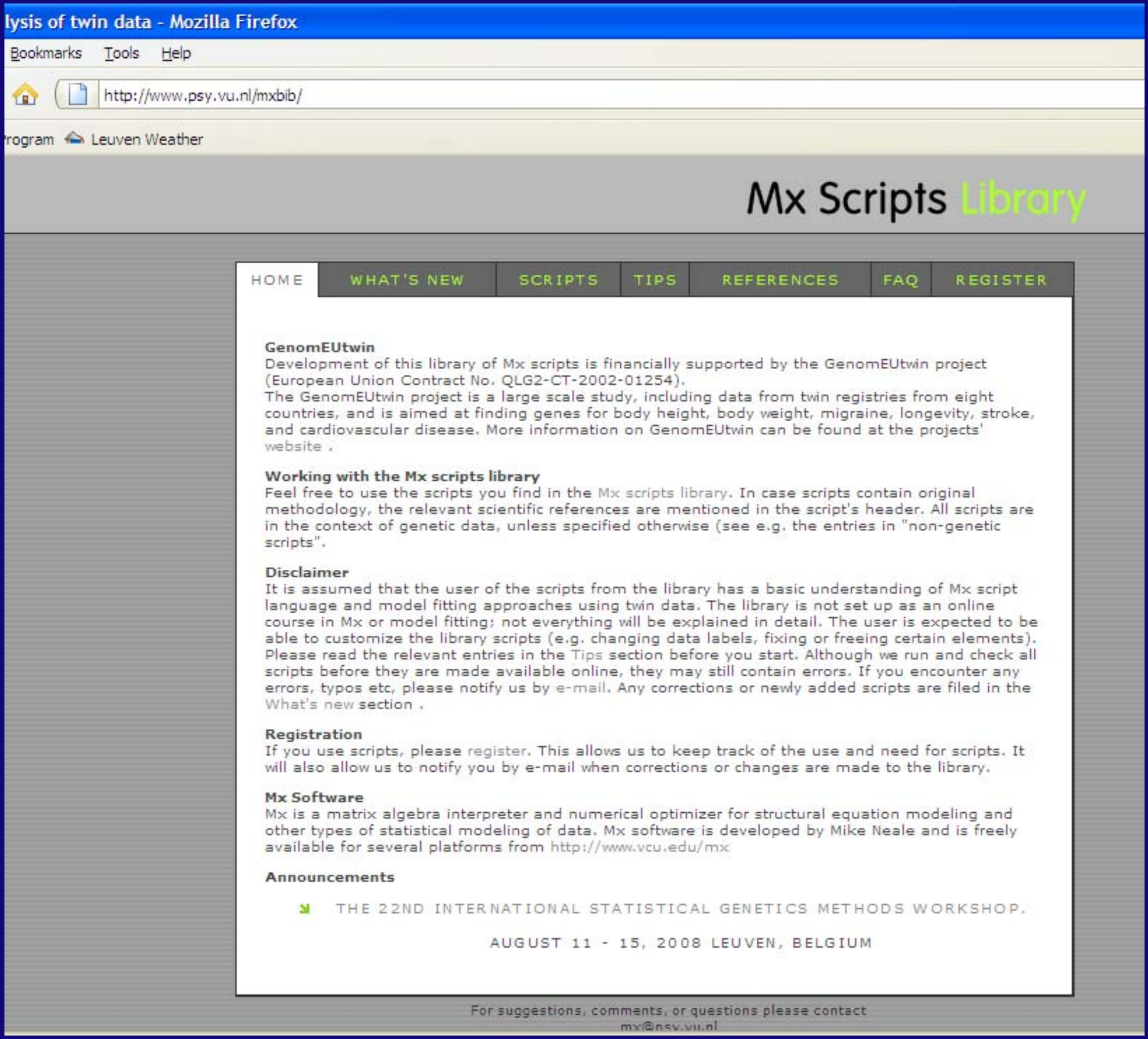
```
G1: Define matrices
 Calculation
 Begin Matrices;
  X full nfac nfac Free     ! latent factor genetic path coefficient
  Y full nfac nfac Free     ! latent factor shared environment path
  Z full nfac nfac Free     ! latent factor unique environment path
  T diag nvar nvar Free     ! variable specific genetic paths
  U diag nvar nvar Free     ! variable specific shared env paths
  V diag nvar nvar Free     ! variable specific residual paths
  F full nvar nfac Free     ! loadings of variables on latent factor
  I Iden 2 2
  M full  1 nvar Free       ! means
 End Matrices;
  Start ..
 Begin Algebra;
  A= F&(X*X') + T*T';       ! genetic variance components
  C= F&(Y*Y') + U*U';       ! shared environment variance components
  E= F&(Z*Z') + V*V';       ! nonshared environment variance components
  L= X*X' + Y*Y' + Z*Z';    ! variance of latent factor
 End Algebra;
End
```

Common Pathway Model

```
G4: Constrain variance of latent factor to 1
 Constraint
 Begin Matrices;
  L computed =L1
  I unit 1 1
 End Matrices;
 Constraint L = I ;
End

G5: Calculate Standardised Solution
 Calculation
 Matrices = Group 1
  D Iden nvar nvar
 End Matrices;
 Begin Algebra;
  R=A+C+E;            ! total variance
  S=(\sqrt(D.R))~;    ! diagonal matrix of standard deviations
  P=S*F;              ! standardized estimates for loadings on F
  Q=S*T_ S*U_ S*V;    ! standardized estimates for specific factors
 End Algebra;
 Options NDecimals=4
End
```

Mx Scripts



GenomEUtwin
Development of this library of Mx scripts is financially supported by the GenomEUtwin project (European Union Contract No. QLG2-CT-2002-01254).
The GenomEUtwin project is a large scale study, including data from twin registries from eight countries, and is aimed at finding genes for body height, body weight, migraine, longevity, stroke, and cardiovascular disease. More information on GenomEUtwin can be found at the projects' website .

Working with the Mx scripts library
Feel free to use the scripts you find in the Mx scripts library. In case scripts contain original methodology, the relevant scientific references are mentioned in the script's header. All scripts are in the context of genetic data, unless specified otherwise (see e.g. the entries in "non-genetic scripts".

Disclaimer
It is assumed that the user of the scripts from the library has a basic understanding of Mx script language and model fitting approaches using twin data. The library is not set up as an online course in Mx or model fitting; not everything will be explained in detail. The user is expected to be able to customize the library scripts (e.g. changing data labels, fixing or freeing certain elements). Please read the relevant entries in the Tips section before you start. Although we run and check all scripts before they are made available online, they may still contain errors. If you encounter any errors, typos etc, please notify us by e-mail. Any corrections or newly added scripts are filed in the What's new section .

Registration
If you use scripts, please register. This allows us to keep track of the use and need for scripts. It will also allow us to notify you by e-mail when corrections or changes are made to the library.

Mx Software
Mx is a matrix algebra interpreter and numerical optimizer for structural equation modeling and other types of statistical modeling of data. Mx software is developed by Mike Neale and is freely available for several platforms from http://www.vcu.edu/mx

Announcements

THE 22ND INTERNATIONAL STATISTICAL GENETICS METHODS WORKSHOP.

AUGUST 11 - 15, 2008 LEUVEN, BELGIUM

For suggestions, comments, or questions please contact
mx@psy.vu.nl