Multivariate Modeling

Nathan Gillespie & Irene Rebollo

Files: \\nathan\2008\
      aggression.dat
      multi_aggr_2var.mx

Practica

1. Go thru Mx bivariate script with fine tooth comb
2. Estimate phenotypic correlations, genetic & environmental latent factor correlations, standardized proportions of variance
3. Test AE, CE and E submodels
4. Try your hand at writing a trivariate Cholesky

Bivariate Cholesky

Folder: \nathan\2008

Files: multi_aggr_2var.mx
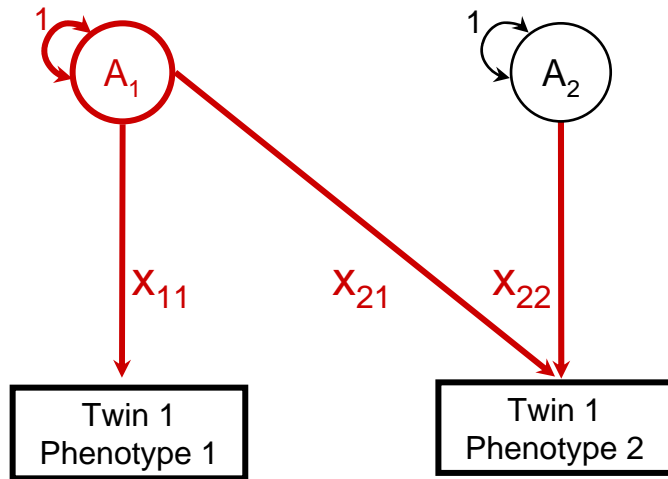
              aggression.dat

Aggression Time 1: 3258 twins in 1991

Aggression Time 2: 3303 twins in 1995

Aggression Time 2: 2984 twins in 1997

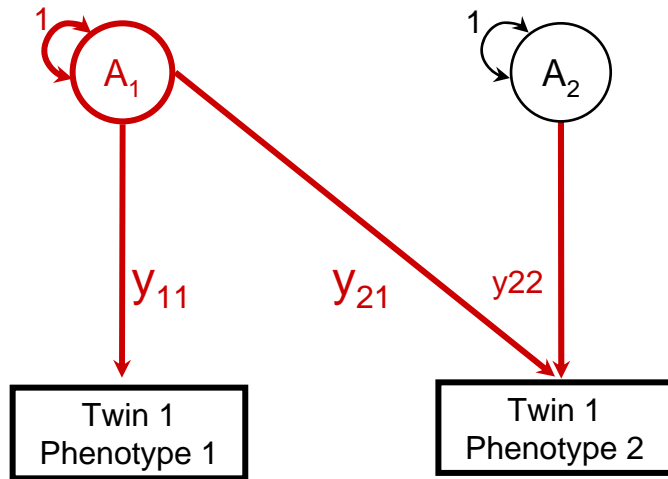Data need to be adjusted for age at interview

Additive genetic path coefficients (X)

Translate path diagram into a matrix

	Twin 1	
	A1	A 2
Phenotype 1	x_{11}	
Phenotype 2	x_{21}	x_{22}

Twin 1

Additive genetic path coefficients (Y)

Translate path diagram into a matrix

	Twin 1	
	C1	C2
Phenotype 1	y_{11}	
Phenotype 2	y_{21}	y_{22}

Twin 1

Additive Genetic Cross-Twin Covariance (DZ)
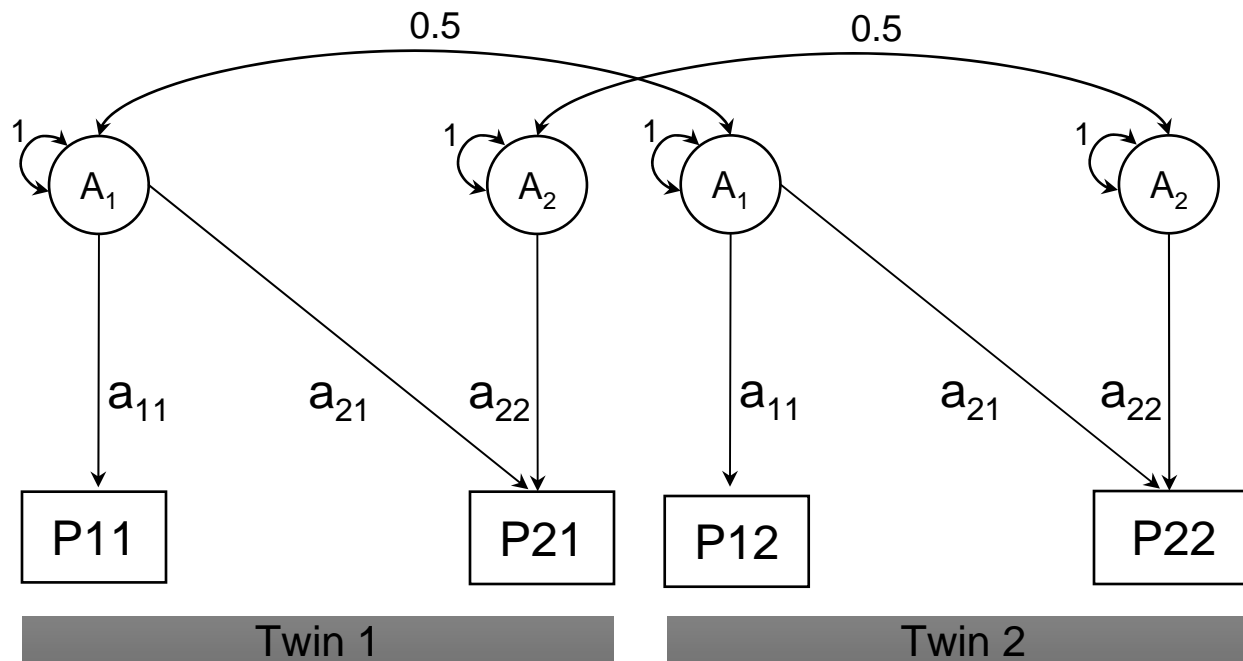
Path Tracing:

Within-traits
P11-P12 =
P21-P22 = $0.5a_{22}^2 + 0.5a_{21}^2$

Cross-traits
P11-P22 = $0.5a_{11}a_{21}$
P21-P12 = $0.5a_{21}a_{11}$

$$0.5 \otimes \Sigma_A = 0.5 \otimes \mathbf{X} * \mathbf{X}' = \begin{bmatrix} & 0.5a_{11}a_{21} \\ 0.5a_{21}a_{11} & 0.5(a_{21}^2 + a_{22}^2) \end{bmatrix}$$

N.B. in Mx $\otimes$ = @

Total Within-Twin Covariance

$$\Sigma_C = \mathbf{Y} * \mathbf{Y}' = \begin{bmatrix} c_{11}^2 & c_{11}c_{21} \\ c_{21}c_{11} & c_{21}^2 + c_{22}^2 \end{bmatrix}$$

$$\Sigma_E = \mathbf{Z} * \mathbf{Z}' = \begin{bmatrix} e_{11}^2 & e_{11}e_{21} \\ e_{21}e_{11} & e_{21}^2 + e_{22}^2 \end{bmatrix}$$

Using matrix addition, the total within-twin covariance for the phenotypes is defined as:

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

$$\Sigma_P = \begin{bmatrix} + c_{11}^2 + e_{11}^2 & + c_{11}c_{21} + e_{11}e_{21} \\ + c_{21}c_{11} + e_{11}e_{21} & + c_{21}^2 + c_{22}^2 + e_{21}^2 + e_{22}^2 \end{bmatrix}$$

Age effects on mean

Means M + B*O | M + B*P ;   1 by 4 matrix

Recall that:

M Full 1 nvar Free   ! grand mean phenotypes [1 2]
B Full 1 ndef Free   ! Age beta [1 1]
O full ndef nvar fix   ! Age for twin 1 at times 1 & 2 [1 2]
P full ndef nvar fix   ! Age for twin 2 at times 1 & 2 [1 2]

$$B*O = [1\ 1] * [1\ 2]$$
$$= [B_{age}] * [age1.1 \times age3.1]$$
$$= [B_{age} \times age1.1 \quad\quad B_{age} \times age3.1]$$
$$B*P = [1\ 1] * [1\ 2]$$
$$= [B_{age}] * [age1.2 \times age3.2]$$
$$= [B_{age} \times age1.2 \quad\quad B_{age} \times age3.2]$$

Age effects on mean

Means M + B*O | M + B*P;   1 by 4 matrix

M + B*O =

$[\mu_{twin1var1} + B_{age} \times age1.1 \quad \mu_{twin1var2} + B_{age} \times age3.1]$

M + B*P =

$[\mu_{twin2var1} + B_{age} \times age1.2 \quad \mu_{twin2var2} + B_{age} \times age3.2]$

Predicted Model

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

Twin 1   Twin 2

		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2

Within-twin covariance

$$\Sigma_P = \boxed{} + \Sigma_C$$

Twin 1

Phenotype 1

$x_{11}^2 + y_{11}^2 + z_{11}^2$

Phenotype 2

$z_{11}z_{21} + y_{11}y_{21} + z_{11}z_{21}$ $x_{22}^2 + x_{21}^2 + y_{22}^2 + y_{21}^2 + z_{22}^2 + z_{21}^2$

Cross-twin covariance   **Within-twin covariance**

Twin 2

Phenotype 1

$1/.5x_{11}^2 + y_{11}^2$ $x_{11}^2 + y_{11}^2 + z_{11}^2$

Phenotype 2

$1/.5x_{11}x_{21} + y_{11}y_{21}$ $1/.5x_{22}^2 + 1/.5x_{21}^2 + y_{22}^2 + y_{21}^2$ $x_{11}x_{21} + y_{11}y_{21} + z_{11}z_{21}$ $x_{22}^2 + x_{21}^2 + y_{22}^2 + y_{21}^2 + z_{22}^2 + z_{21}^2$

Predicted Model

	Twin 1		Twin 2	
	Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
Twin 1 — Phenotype 1, Phenotype 2	**Within-twin covariance** $X*X' + Y*Y' + Z*Z'$		**Cross-twin covariance** $X*X' + Y*Y'$	
Twin 2 — Phenotype 1, Phenotype 2	**Cross-twin covariance** $X*X' + Y*Y'$		**Within-twin covariance** $X*X' + Y*Y' + Z*Z'$	

Estimating genetic and environmental correlations
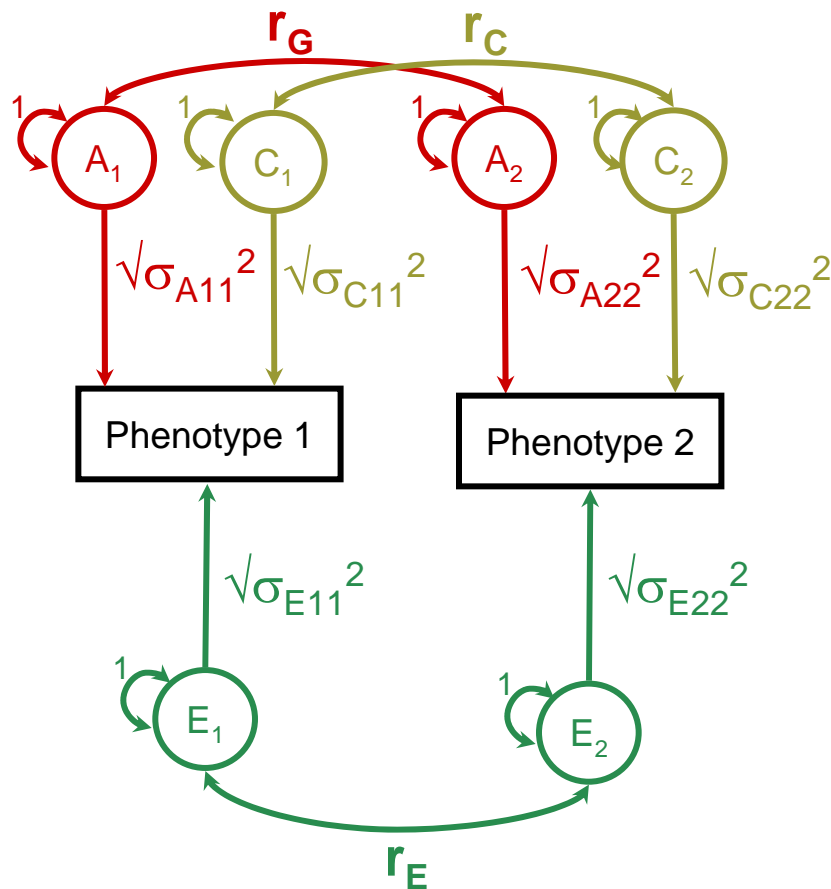
Observed phenotypic correlation is the result of correlation at

- Genetic level

- Common environmental level

- Unique environmental level

Estimate r_g, r_c, and r_e

Estimate contribution of r_g, r_c, and r_e to phenotypic correlation between phenotype 1 and 2

Correlations between latent factors

$$r_{12} = \frac{\sigma_{12}^2}{\sqrt{\sigma_{11}^2 \times \sigma_{22}^2}} = \frac{1}{\sqrt{\sigma_{11}^2}} * \sigma_{12}^2 * \frac{1}{\sqrt{\sigma_{22}^2}}$$

$$\begin{bmatrix} 1 & r_G \\ r_G & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A11}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A22}^2}} \end{bmatrix} * \begin{bmatrix} \sigma_{A11}^2 & \sigma_{A12}^2 \\ \sigma_{A21}^2 & \sigma_{A22}^2 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A11}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A22}^2}} \end{bmatrix}$$

Matrix function in Mx:

R = \sqrt(**I.A**)˜ * **A** * \sqrt(**I.A**)˜;

Where **I** is an identity matrix: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

and **I.A** = $\begin{bmatrix} \sigma_{A11}^2 & 0 \\ 0 & \sigma_{A22}^2 \end{bmatrix}$

R = \stnd(**A**);

Proportion of observed phenotypic correlation explained by A, C and E

- ```
  Begin algebra;
  K = A%P|C%P|E%P;
  S = A%(A+C+E)|C%(A+C+E)|E%(A+C+E);
  End algebra;
  ```
  % is the Mx operator for element division

$$\mathbf{K} = \left[\begin{array}{cc} & \\ & \end{array} \quad \begin{array}{cc} c^2_{P1} & C_{COV} \\ C_{COV} & c^2_{P2} \end{array} \quad \begin{array}{cc} e^2_{P1} & E_{COV} \\ E_{COV} & e^2_{P2} \end{array} \right]$$

Proportion of the phenotypic correlation due to genetic effects

Run ACE Model

ACE Model Fit

Model	LL	df	Δ2LL	Δdf	p	BIC
ACE	37111.99	6549	-	-	-	3476.37

Model comparisons

Model	LL	Df	Δ2LL	Δdf	p	BIC
ACE	37111.99	6549	-	-	-	3476.37
AE	37147.08	6552	35.09	3	<0.001	3487.00
CE	**37112.29**	**6552**	**0.30**	**3**	**0.96**	**3469.61**
E	37376.50	6555	264.51	6	<0.001	3594.806

Write trivariate ACE Model

Change **nvar**

Include starting values for 3rd variables

Change **Select** statement to include 3rd variable

Change **Definition** statement 3rd variable ages

ACE Model Fit

Model	LL	df	Δ2LL	Δdf	p	BIC
ACE	52605.89	9523	-	-	-	3409.126

Mx scripts

http://www.psy.vu.nl/mxbib
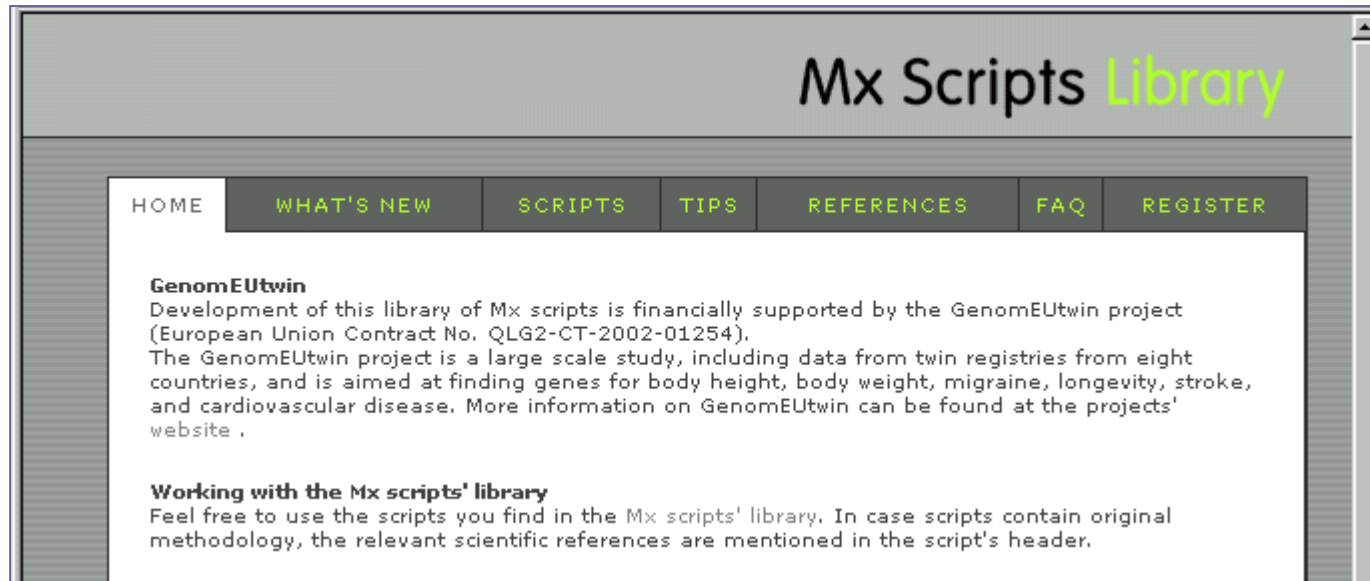
r_{ph} **due to A**

$$h_{x,y}^2 = \sqrt{h_x^2} * r_g * \sqrt{h_y^2}$$

r_{ph} **due to C**

$$c_{x,y}^2 = \sqrt{c_x^2} * r_c * \sqrt{c_y^2}$$

r_{ph} **due to E**

$$e_{x,y}^2 = \sqrt{e_x^2} * r_e * \sqrt{e_y^2}$$
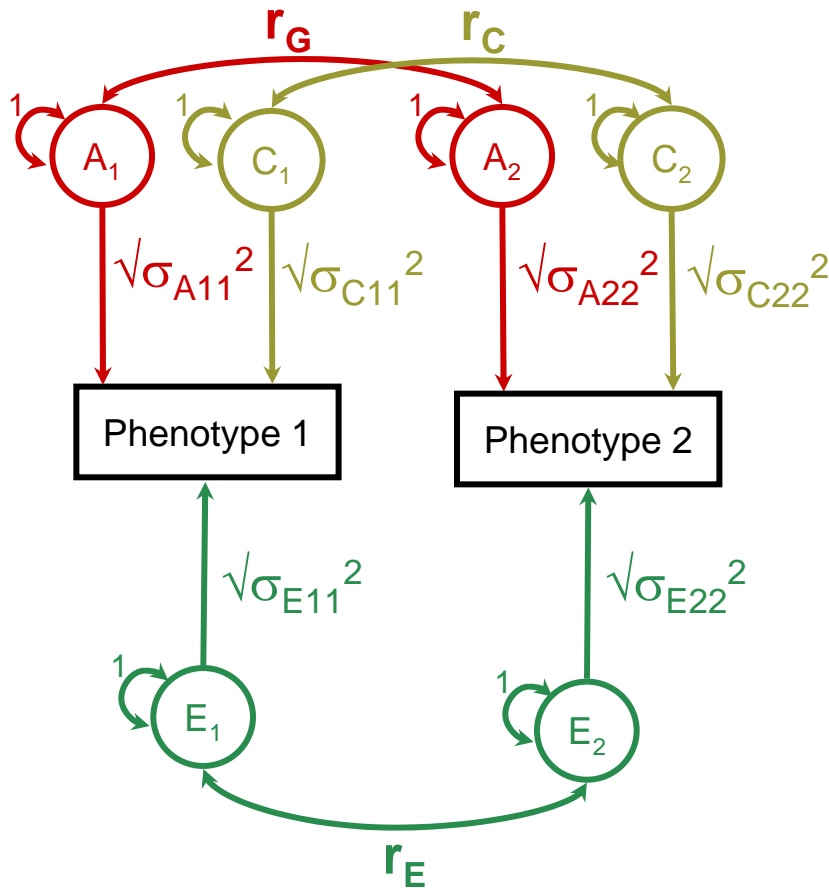
Genetic contribution to observed correlation (h^2_{xy}) is a function of rg and both heritabilities

Observed correlation

$$r = \sqrt{h_x^2} * r_g * \sqrt{h_y^2}$$

$$+ \sqrt{c_x^2} * r_c * \sqrt{c_y^2}$$

$$+ \sqrt{e_x^2} * r_e * \sqrt{e_y^2}$$

Observed correlation and contributions from A, C and E

$$r =$$

$$\sqrt{h_x^2} * r_g * \sqrt{h_y^2}$$

$$+ \sqrt{c_x^2} * r_c * \sqrt{c_y^2}$$

$$+ \sqrt{e_x^2} * r_e * \sqrt{e_y^2}$$

$$.54 =$$

$$\sqrt{.05} * 1.00 * \sqrt{.00} = 0.01$$

$$+ \sqrt{.28} * 0.79 * \sqrt{0.29} = 0.22$$

$$+ \sqrt{.67} * 0.44 * \sqrt{0.71} = 0.31$$

Proportion of the observed correlation (or covariance) explained by correlation at the genetic level: 0.01/0.54 = 0.02

Proportion of the observed correlation (or covariance) explained by correlation at the shared environmental level: 0.22/0.54 = 0.41

Proportion of the observed correlation (or covariance) explained by correlation at the non-shared environmental level: 0.31/0.54 = 0.57