Introduction to Multivariate Genetic Analysis

Meike Bartels, Hermine Maes, Elizabeth Prom-Wormley and Michel Nivard

Aim and Rationale

<u>Aim:</u> to examine the source of factors that make traits correlate or co-vary <u>Rationale:</u> Traits may be correlated due to shared genetic factors (A) or shared environmental factors (C or E)

Can use information on multiple traits from twin pairs to partition covariation into genetic and environmental components

Example 1

Why do traits correlate/covary?

- How can we explain the association?
 - Additive genetic factors (r_G)
 - Shared environment (r_c)

Non-shared environment (r_E)



Example 2

- Associations between phenotypes over time
 - Does anxiety in childhood lead to depression in adolescence?
- How can we explain the association?
 - Additive genetic factors (a₂₁)
 - Shared environment (c₂₁)
 - Non-shared environment (e₂₁)
 - How much is not explained by prior anxiety?



Sources of Information

- For example: two traits measured in twin pairs
- Interested in:
 - Cross-trait covariance within individuals
 - Cross-trait covariance between twins
 - MZ:DZ ratio of cross-trait covariance between twins

Observed Covariance Matrix

		Twin 1		Tw	vin 2
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
/in 1	Phenotype 1	Variance P1			
Ч	Phenotype 2	Covariance P1-P2	Variance P2		
Fwin 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1	
	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2

Observed Covariance Matrix

		Twin 1		Tw	in 2
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
		Within-twin	covariance		
/in 1	Phenotype 1	Variance P1			
Τ«	Phenotype 2	Covariance P1-P2	Variance P2		
				Within-twin	covariance
win 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1	
F	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2

Observed Covariance Matrix

		Twin 1		Tw	in 2
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
		Within-twin	covariance		
/in 1	Phenotype 1	Variance P1			
Ţ	Phenotype 2	Covariance P1-P2	Variance P2		
		Cross-twin covariance		Within-twin covariance	
win 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1	
┝━	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2

SEM: Cholesky Decomposition



SEM: Cholesky Decomposition



SEM: Cholesky Decomposition



Cholesky Decomposition

Path Tracing























		Twin 1		Twin 2	
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
		Within-twin	covariance		
vin 1	Phenotype 1	a ₁₁ ² +c ₁₁ ² +e ₁₁ ²			
ŕ	Phenotype 2	a ₁₁ a ₂₁ +c ₁₁ c ₂₁ + e ₁₁ e ₂₁	$a_{22}^{2}+a_{21}^{2}+c_{22}^{2}+c_{21}^{2}+c_{21}^{2}+c_{21}^{2}+c_{21}^{2}$		
		Cross-twin	covariance	Within-twin	covariance
win 2	Phenotype 1	1/.5a ₁₁ ² +c ₁₁ ²		a ₁₁ ² +c ₁₁ ² +e ₁₁ ²	
 	Phenotype 2	1/.5a ₁₁ a ₂₁ + c ₁₁ c ₂₁	$1/.5a_{22}^{2}+1/.5$ $a_{21}^{2}+C_{22}^{2}+C_{21}^{2}$	a ₁₁ a ₂₁ +c ₁₁ c ₂₁ + e ₁₁ e ₂₁	$a_{22}^{2}+a_{21}^{2}+c_{22}^{2}$ + $c_{21}^{2}+e_{22}^{2}+e_{21}$

		Twin 1		Twin 2	
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
		Within-twin	covariance		
/in 1	Phenotype 1	Variance P1			
Ч	Phenotype 2	Covariance P1-P2	Variance P2		
		Cross-twin	covariance	Within-twin	covariance
win 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1	
Ļ	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2









Example Covariance Matrix



Example Covariance Matrix



Example Covariance Matrix



Summary

- Within-individual cross-trait covariance implies common aetiological influences
- Cross-twin cross-trait covariance implies common aetiological influences are familial
- Whether familial influences genetic or environmental shown by MZ:DZ ratio of crosstwin cross-trait covariances

Cholesky Decomposition Bivariate Genetic analyses

Specification in OpenMx



Path Tracing: $\Sigma_A = \begin{bmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 \end{bmatrix}$



Path Tracing:			
$\Sigma_A = \left[egin{array}{c} a_{11}^2\ a_{21}a_{11} \end{array} ight.$	$a_{11} a_{21}^2 + a_{21}^2 a$	$a_{21} - a_{22}^2$	
a Lower 2 x 2:	a1	a2	_
P1	a_{11}	0	
P2	a_{21}	a_{22}	



$$S_{A} = a^{*} a^{T}$$

$$S_{A} = a^{\%} * \% t(a) = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^{2} + 0 \times 0 & a_{11}a_{21} + 0 \times a_{22} \\ a_{21}a_{11} + 0 \times a_{22} & a_{21}^{2} + a_{22}^{2} \end{bmatrix}$$



Total Within-Twin Covar.

$$S_{A} = a \% *\% t(a) = \begin{bmatrix} a_{11}^{2} & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^{2} + a_{22}^{2} \end{bmatrix} S_{C} = c \% *\% t(c) = \begin{bmatrix} c_{11}^{2} & c_{11}c_{21} \\ c_{21}c_{11} & c_{21}^{2} + c_{22}^{2} \end{bmatrix}$$
$$S_{E} = e \% *\% t(e) = \begin{bmatrix} e_{11}^{2} & e_{11}e_{21} \\ e_{21}e_{11} & e_{21}^{2} + e_{22}^{2} \end{bmatrix}$$

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

$$S_{V} = \Sigma_{A} + \Sigma_{C} + \Sigma_{E}$$

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

OpenMx Matrices & Algebra

```
OpenMx
multACEModel <- mxModel("multACE", mxModel("ACE",
# Matrices a, c, and e to store a, c, and e path coefficients
mxMatrix(type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="a"),
                                                                               mxMatrix(
type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="c"), mxMatrix(type="Lower",
nrow=nv, ncol=nv, free=TRUE, values=.6, name="e"),
# Matrices A, C, and E compute variance components
                                                       mxAlgebra( expression=a \%*% t(a),
name="A"),
mxAlgebra( expression=c %*% t(c), name="C" ),
mxAlgebra( expression=e \% \% t(e), name="E"),
# Algebra to compute total variances and standard deviations (diagonal only)
                                                                            mxAlgebra(
expression=A+C+E, name="V"),
mxMatrix( type="lden", nrow=nv, ncol=nv, name="l"),
                                                      mxAlgebra( expression=solve(sqrt(I^*V)),
name="isd"),
```

Additive Genetic Cross-Twin Covariance (DZ)



Additive Genetic Cross-Twin Covariance (MZ)



 $1 \dot{A} S_A = 1\% x\% (a\% *\% t(a)) =$

$$egin{array}{ccc} a_{11}^2 & a_{11}a_{21}\ a_{21}a_{11} & (a_{21}^2+a_{22}^2) \end{array}$$

Common Environment Cross-Twin Covariance



$$1\ddot{\mathsf{A}}\,\mathsf{S}_{c} = 1\%x\%(c\%*\%t(c)) = \begin{bmatrix} c_{11}^{2} & c_{11}c_{21} \\ c_{21}c_{11} & (c_{21}^{2} + c_{22}^{2}) \end{bmatrix}$$

Covariance Model for Twin Pairs



Obtaining Standardised Estimates

Three Important Results

- 1. Variance Decomposition -> Heritability, (Shared) environmental influences
- Covariance Decomposition -> The influences of genes and environment on the covariance between the two variables

"how much of the phenotypic correlation is accounted for by genetic and environmental influences"

3. Genetic and Environmental correlations -> the overlap in genes and environmental effects

"is there a large overlap in gene/ environmental sets"

OpenMx Output

- 1. Variance Decomposition -> Heritability, (Shared) environmental influences
- Covariance Decomposition -> The influences of genes and environment on the covariance between the two variables

[1] "Matrix A/V"
 stCovA1 stCovA2
 family 0.4809 0.7154
 happy 0.7154 0.3995

 [1] "Matrix C/V" stCovC1 stCovC2
 family 0.0919 0.1121
 happy 0.1121 0.0250

[1] "Matrix E/V" stCovE1 stCovE2 family 0.4272 0.1725 happy 0.1725 0.5755

Genetic correlation



OpenMx Output

[1] "Matrix solve(sqrt(I*A)) %&% A" corA1 corA2
family 1.0000 0.6985
happy 0.6985 1.0000

[1] "Matrix solve(sqrt(I*E)) %&% E" corE1 corE2
family 1.0000 0.1488
happy 0.1488 1.0000

Genetic correlation & contribution to observed correlation



If the rg = 1, the two sets of genes overlap completely

If however a11 and a22 are near to zero, genes do not contribute to the observed correlation

Twin 1

Interpreting Results

- High genetic correlation = large overlap in genetic effects on the two phenotypes
- Does it mean that the phenotypic correlation between the traits is largely due to genetic effects?
 - No: the substantive importance of a particular r_G depends the value of the correlation **and** the value of the $\sqrt{\sigma_A^2}$ paths i.e. importance is also determined by the heritability of each phenotype

Example



Proportion of r_P due to additive genetic factors:

 $\left(\sqrt{a_x^2} * r_g * \sqrt{a_y^2}\right) / r_p$



Interpretation of Correlations

Consider two traits with a phenotypic correlation of 0.40 :

- h_{P1}^{2} = 0.7 and h_{P2}^{2} = 0.6 with r_G = .3
- Correlation due to additive genetic effects = ?
- Proportion of phenotypic correlation attributable to additive genetic effects = ?

$$h_{P1}^2 = 0.2$$
 and $h_{P2}^2 = 0.3$ with $r_G = 0.8$

- Correlation due to additive genetic effects = ?
- Proportion of phenotypic correlation attributable to additive genetic effects = ?

Correlation due to A: $\sqrt{h_{P1}^2 * r_G * \sqrt{h_{P2}^2}}$ Divide by r_P to find proportion of phenotypic correlation.

More Variables...



More Variables...



Expanded Matrices



OpenMx Parameter Matrices



Practical

SCRIPT:

F:\meike\2014\Multivariate\Bivariate

DATA:

DHBQ_bs.dat

DATA

- General Family Functioning, Subjective Happiness
- T1, T2, brother, sister
- missing -999



Observed Cross-twin Cross-trait Correlations

describe(mzData, skew=F)

describe(dzData, skew=F)

dim(mzData)

dim(dzData)

```
colMeans(mzData,na.rm=TRUE)
colMeans(dzData,na.rm=TRUE)
cov(mzData,use="complete")
cov(dzData,use="complete")
cor(mzData,use="complete")
cor(dzData,use="complete")
```

Observed Cross-twin Cross-trait Correlations

> cor(mzData,use="complete")

	family1	happy1	family2	happy2
family1	1.00	0.40	0.58	0.37
happy1	0.40	1.00	0.35	0.46
family2	0.58	0.35	1.00	0.41
happy2	0.37	0.46	0.41	1.00

> cor(dzData,use="complete")

	family1	happy1	family2	happy2
family1	1.00	0.47	0.30	0.061
happy1	0.47	1.00	0.30	0.217
family2	0.30	0.30	1.00	0.379
happy2	0.06	0.21	0.37	1.000

PRAC I. The ACE model and its estimates

- 1. Run the ACE model
- 2. What is the heritability of FAM and HAP?
- 3. What is the genetic influence on the covariance?
- 4. What is the genetic correlation?

OpenMx Output

- 1. Variance Decomposition -> Heritability, (Shared) environmental influences
- Covariance Decomposition -> The influences of genes and environment on the covariance between the two variables

[1] "Matrix A/V"
 stCovA1 stCovA2
 family 0.4809 0.7154
 happy 0.7154 0.3995

 [1] "Matrix C/V" stCovC1 stCovC2
 family 0.0919 0.1121
 happy 0.1121 0.0250

[1] "Matrix E/V" stCovE1 stCovE2 family 0.4272 0.1725 happy 0.1725 0.5755

OpenMx Output

[1] "Matrix solve(sqrt(I*A)) %&% A" corA1 corA2
family 1.0000 0.6985
happy 0.6985 1.0000

[1] "Matrix solve(sqrt(I*E)) %&% E" corE1 corE2
family 1.0000 0.1488
happy 0.1488 1.0000

PRAC II. Trivariate Model

- 1. Add a third variable (Satisfaction with Life) to the model
- 2. Run model
- 3. What are the parameter estimates?
- 4. What is the genetic correlation?

Changes that had to be made

Select Variables for Analysis

- Vars <- c('family','happy','life')
- nv <- 3 # number of variables
- ntv <- nv*2 # number of total variables
- selVars <- paste(Vars,c(rep(1,nv),rep(2,nv)),sep="")</pre>

svMe <- c(7,5,5)

Genetic and Environmental Influences

0.5501

0.6348

[1] "Matrix A/V"

happy

life

0.1742

0.1299

	stCovA1	stCovA2	stCovA3			
family	0.4117	0.6549	0.5697			
happy	0.6549	0.3297	0.3224			
life	0.5697	0.3224	0.2246			
[1] "Matr	ix C/V"					
	stCovC1	stCovC2	stCovC3			
Family	0.1517	0.1709	0.3004			
happy	0.1709	0.0681	0.1275			
life	0.3004	0.1275	0.1406			
[1] "Matrix E/V"						
	stCovE1	stCovE2	stCovE3			
Family	0.4366	0.1742	0.1299			

0.6022

0.5501

Genetic and Environmental Correlations

[1] "Matrix solve(sqrt(I*A)) %&% A"

	corA1	corA2	corA3
Family	1.0000	0.7620	0.7817
Нарру	0.7620	1.0000	0.8856
ife	0.7817	0.8856	1.0000

[1] "Matrix solve(sqrt(I*C)) %&% C"

	corC1	corC2	corC3
Family	1.0000	0.7208	0.8585
happy	0.7208	1.0000	0.9743
life	0.8585	0.9743	1.0000

[1] "Matrix solve(sqrt(I*E)) %&% E"

	corE1	corE2	corE3
Family	1.0000	0.1456	0.1029
happy	0.1456	1.0000	0.6651
life	0.1029	0.6651	1.0000

Genetic Models

Genetic and environmental factor analysis examples





Common Pathway Model

Independent Pathway Model