2020 International Statistical Genetics Workshop Introduction to Multivariate Genetic Analysis PRACTICAL

Lucía Colodro Conde, Meike Bartels, Elizabeth Prom-Wormley, and Hermine Maes with thanks to Katrina Grasby, Jose Morosoli, Sarah Medland, and Dorret Boomsma

In \\workshop\Faculty\lucia\2020\Wednesday\twoACEvc_practical_2traits.R

The dataset: Simulated data

- Children age ~ 8 years old, 50 % females
- Phenotypes are scores on reading, grammar, writing, and math skills.
- Available both raw and scaled (mean 0, SD 1) before restructuring the file to one row per family!
- Means and Variances could be equated across Twin Order and Zygosity
- No need of sex-limitation model
- Age and sex should be included as covariates
- We'll analyse "Read" and "Writ"

Three Concepts

1.- What is the variance due to genetic and (shared and unique) environmental contributions for each trait?Variance Decomposition

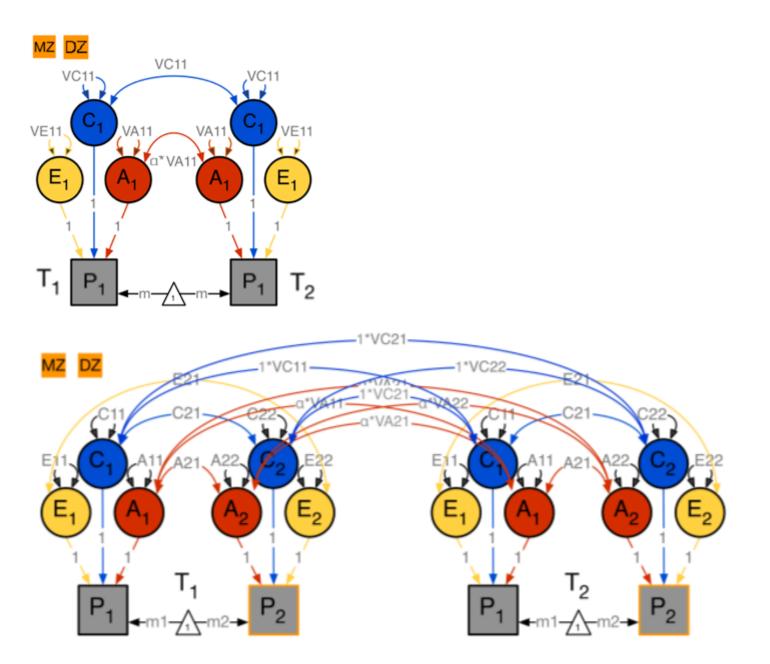
2.- How much of the phenotypic correlation between the traits is accounted for by genetic and (shared and unique) environmental factors?*Covariance Decomposition*

3.- To what extent do the genetic factors underlying each trait overlap?And the (shared and unique) environmental factors?Genetic and Environmental correlations

From univariate

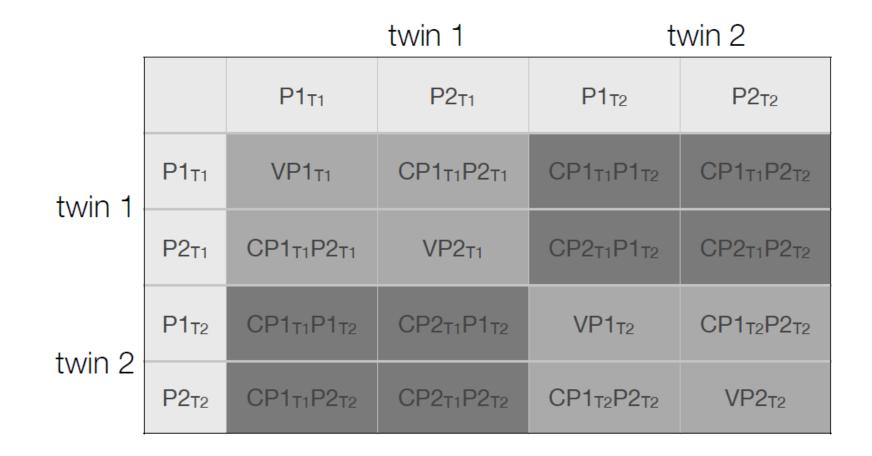
to

multivariate

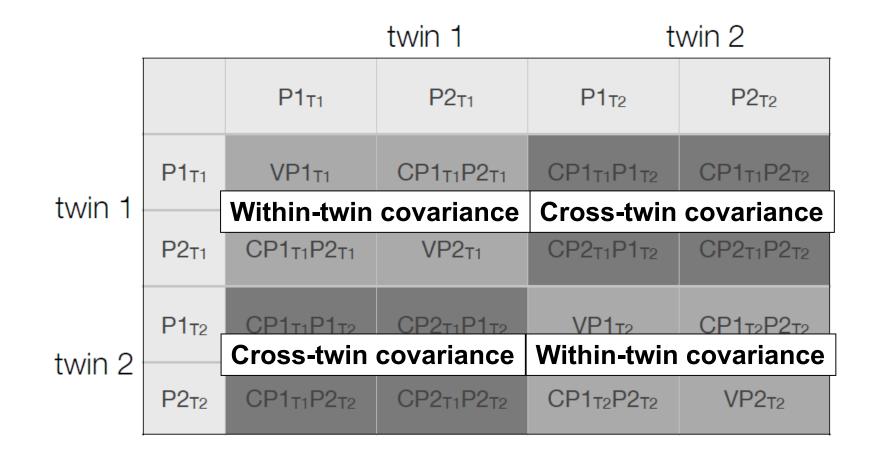


 α = assumed degree of genetic relatedness (1, 0.5)

Twin Covariances / Correlations



Twin Covariances / Correlations



Variance/covariance matrices

Correlation matrices

ΜZ

ΜZ

		Tw	in 1	Twin	2			1	Twin 1	Т	win 2
Twin 1 Twin 2	read_1 writ_1 read_2 writ_2	read_1 10.44 5.01 7.60 4.07	writ_1 5.01 8.21 4.53 4.01	read_2 7.60 4.53 10.38 4.88	writ_2 4.07 4.01 4.88 8.18	Twin 1 Twin 2	read_1 writ_1 read_2 writ_2	read_1 1.00 0.54 0.73 0.44	writ_1 0.54 1.00 0.49 0.49	read_2 0.73 0.49 1.00 0.53	writ_2 0.44 0.49 0.53 1.00
DZ		Tw	in 1	Twin	2	DZ		т	win 1	Τv	vin 2

Variance/covariance matrices

Correlation matrices

ΜZ

MZ

		Tw	vin 1	Twin	2			-	Twin 1	T	win 2
		read_1	writ_1	read_2	writ_2			read_1	writ_1	read_2	writ_2
Twin 1	read_1	10.44				Twin 1	read_1	1.00			
	writ_1	5.01	8.21				writ_1	0.54	1.00		
Twin 2	read_2	7.60	4.53	10.38		Twin 2	read_2		0.49	1.00	
	writ_2	4.07	4.01	4.88	8.18		writ_2	0.44	0.49	0.53	1.00
D7		-	• •	- ·.	2	DZ		т	win 1	Ти	vin 2
DZ		Тм	vin 1	Twin	12	DZ		Т	win 1	Τv	vin 2
DZ						DZ				Tv read_2	
	read 1	read_1		Twin read_2			read_1	read_1 1.00			
DZ Twin 1	read_1 writ_1	read_1 9.75	writ_1			DZ Twin 1	writ_1	read_1 1.00 0.50			
	read_1 writ_1 read_2	read_1					_	read_1 1.00 0.50	writ_1		

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

- Are there common etiological influences?

read_1 writ_1 read_2 writ_2 Twin1 read_1 1.00 writ_1 0.54 1.00 ____ read_2 0.73 0.49 1.00

Twin 1

Twin 2 writ_2 0.44 0.49 0.53 1.00

Twin 2

DZ

Twin 2 Twin 1 read_1 writ_1 read_2 writ_2 1.00 read 1 Twin 1 writ_1 0.50 1.00 read_2 0.43 0.34 1.00 Twin 2 writ_2 0.26 0.35 0.59 1.00

MZ

MZ

Twin 2 Twin 1 read_1 writ_1 read_2 writ_2 read_1 1.00 Twin 1 0.54 writ_1 1.00 0.49 read_2 0.73 1.00 Twin 2 writ_2 0.44 0.49 0.53 1.00

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

Are there common etiological influences?
 Within-twin Cross-trait covariances

DZ

		Twi	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	read_1 writ_1	0.50	1.00			
	read 2	0.43	0.34	1.00		
Twin 2	writ_2	0.26	0.35	0.59	1.00	

ΜZ

		Tw	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	read_1 writ_1	0.54	1.00			
Twin 2	read_2	0.73	0.49	1.00		
	writ_2	0.44	0.49	0.53	1.00	

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

Are there common etiological influences?
 Within-twin Cross-trait covariances

- Do familial common etiological influences contribute to the covariation?

DZ

		Twi	in 1	Twin 2		
			writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	writ_1	0.50	1.00			
	read_2	0.43	0.34	1.00		
Twin 2	writ_2	0.26	0.34 0.35	0.59	1.00	

ΜZ

		Tw	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1 writ_1	1.00				
	writ_1	0.54	1.00			
Twin 2	read_2	0.73	0.49	1.00		
	read_2 writ_2	0.44	0.49	0.53	1.00	

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

Are there common etiological influences?
 Within-twin Cross-trait covariances

 Do familial common etiological influences contribute to the covariation? Cross-twin Cross-trait covariances

DZ

	Twi	in 1	Twin 2		
	read_1	writ_1	read_2	writ_2	
read_1	1.00				
		1.00			
read_2	0.43	0.34	1.00		
writ_2	0.26	0.35	0.59	1.00	
	writ_1	read_1 read_1 1.00 writ_1 0.50	read_1 1.00 writ_1 0.50 1.00	read_1 writ_1 read_2 read_1 1.00	

MZ

		Tw	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	read_1 writ_1	0.54	1.00			
Twin 2	read_2	0.73	0.49	1.00		
	writ_2	0.44	0.49	0.53	1.00	

DZ

UL		Twi	in 1	Twin 2		
			writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	writ_1	0.50	1.00			
	read_2	0.43	0.34	1.00		
Twin 2	writ_2	0.26	0.35	0.59	1.00	

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

- Are there common etiological influences?
 Within-twin Cross-trait covariances
- Do familial common etiological influences contribute to the covariation? Cross-twin Cross-trait covariances
- Are the common etiological influences genetic and/or environmental (A and/or C)?

MZ

Twin 2 Twin 1 read_1 writ_1 read_2 writ_2 1.00 read 1 Twin 1 writ 1 0.54 1.00 0.73 read 2 0.49 1.00 Twin 2 writ_2 0.44 0.49 0.53 1.00 DZ Twin 2 Twin 1 read_1 writ_1 read_2 writ_2 1.00 read 1 Twin 1 writ_1 0.50 1.00 read 2 0.43 0.34 1.00 Twin 2 writ_2 0.26 0.35 0.59 1.00 What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

- Are there common etiological influences?
 Within-twin Cross-trait covariances
- Do familial common etiological influences contribute to the covariation? Cross-twin Cross-trait covariances

 Are the common etiological influences genetic and/or environmental (A and/or C)?
 MZ/DZ ratio of Cross-twin Cross-trait covariances ΜZ

		Tw	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1	1.00 0.54				
	writ_1	0.54	1.00			
Truin 2	read_2	0.73	0.49 0.49	1.00		
Twin 2	writ_2	0.44	0.49	0.53	1.00	

DZ Twin 2 Twin 1 read_1 writ_1 read_2 writ_2 1.00 read 1 Twin 1 writ_1 0.50 1.00 0.43 read_2 0.34 1.00 Twin 2 writ 2 0.26 0.35 0.59 1.00 What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

- Are there common etiological influences?
 Within-twin Cross-trait covariances
- Do familial common etiological influences contribute to the covariation? Cross-twin Cross-trait covariances
- Are the common etiological influences genetic and/or environmental (A and/or C)?
 MZ/DZ ratio of Cross-twin Cross-trait covariances
- Is there evidence of covariation due to unique environment (E)?

MZ

		Tw	in 1	Twin 2		
		read_1	writ_1	read_2	writ_2	
Twin 1	read_1	1.00				
	read_1 writ_1	0.54	1.00			
Twin 2				1.00		
	writ_2	0.44	0.49 0.49	0.53	1.00	

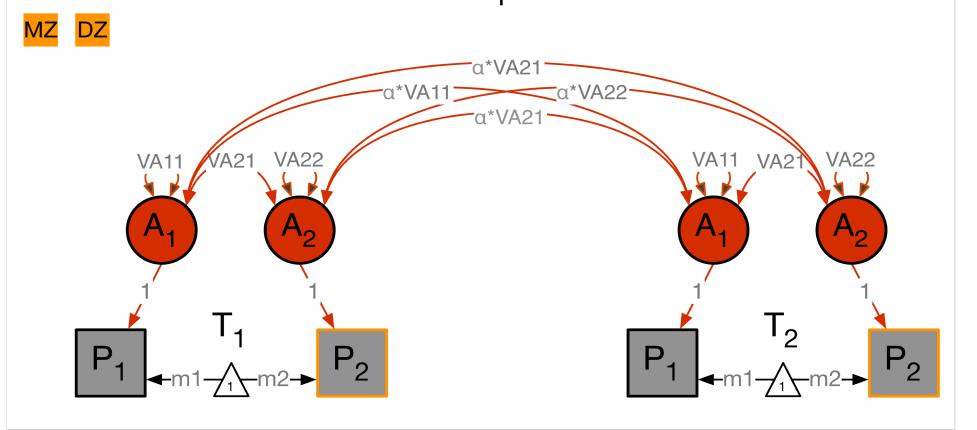
DZ

υz		Twi	in 1	Twin 2	
		read_1	writ_1	read_2	writ_2
Twin 1	read_1	1.00			
	writ_1	0.50	1.00		
	read_2	0.43	0.34	1.00	
Twin 2	read_2 writ_2	0.26	0.35	0.59	1.00

What are our initial expectations after looking at the covariance and correlation matrices by zygosity?

- Are there common etiological influences?
 Within-twin Cross-trait covariances
- Do familial common etiological influences contribute to the covariation? Cross-twin Cross-trait covariances
- Are the common etiological influences genetic and/or environmental (A and/or C)?
 MZ/DZ ratio of Cross-twin Cross-trait covariances
- Is there evidence of covariation due to unique environment (E)?
 Ratio of cross-twin to within-twin cross-trait covariance

Variance Components

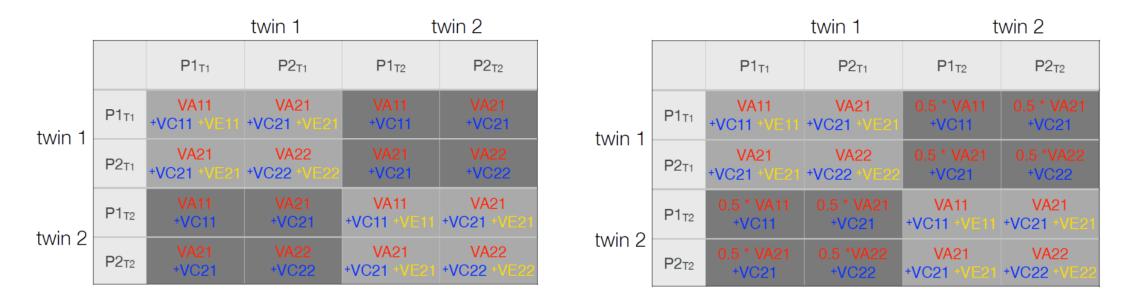


covA <- mxMatrix(type="Symm", nrow=nv, ncol=nv, free=TRUE, [,1] [,2] values=valDiag(svPa,nv), labels=labLower("VA",nv), name="VA") [2,] "va11" "va21"

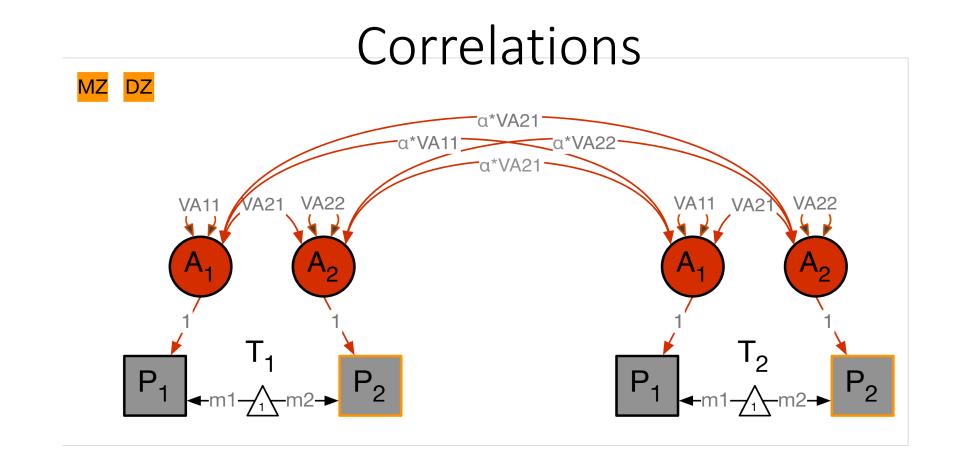
 α = assumed degree of genetic relatedness (1, 0.5)

Predicted MZ Twin Covariance

Predicted DZ Twin Covariance



Create Algebra for expected Variance/Covariance Matrices in MZ & DZ twins



corA <- mxAlgebra(expression=solve(sqrt(I*VA))%&%VA, name ="rA")</pre>

$$\mathbf{r}_{g} = \frac{VA_{12}}{\sqrt{VA_{11} * VA_{22}}} \begin{bmatrix} 1 & r_{g} \\ r_{g} & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^{2}}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^{2}}} \end{bmatrix} * \begin{bmatrix} \sigma_{A_{11}}^{2} & \sigma_{A_{12}}^{2} \\ \sigma_{A_{21}}^{2} & \sigma_{A_{22}}^{2} \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^{2}}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^{2}}} \end{bmatrix}$$

Three Important Concepts

1.- What is the variance due to genetic and environmental contributions for each trait?Variance Decomposition

2.- How much of the phenotypic correlation between the traits is accounted for by genetic and environmental factors? *Covariance Decomposition*

3.- To what extent do the genetic factors underlying each trait overlap?And the environmental factors?Genetic and Environmental correlations

	SA	SC	SE
Read			
Write			

mxAlgebra 'SV'
\$formula: cbind(VA/V, VC/V, VE/V)
\$result:

SA SA SC SC SE SE SE SV 0.57018782 0.60226513 0.14994850 0.25553311 0.27986369 0.14220176 SV 0.60226513 0.25562130 0.25553311 0.22731236 0.14220176 0.51706634

	SA	SC	SE
Read			
Write			

mxAlgebra 'SV'
\$formula: cbind(VA/V, VC/V, VE/V)
\$result:

SASASCSCSESESESV0.570187820.602265130.149948500.255533110.279863690.14220176SV0.602265130.255621300.255533110.227312360.142201760.51706634

	SA	SC	SE
Read	0.57 (0.44,0.70)	0.15 (0.03 <i>,</i> 0.26)	0.28 (0.24 <i>,</i> 0.32)
Write			

mxAlgebra 'SV'
\$formula: cbind(VA/V, VC/V, VE/V)
\$result:

SASASCSCSESESV 0.570187820.602265130.149948500.255533110.279863690.14220176SV 0.602265130.255621300.255533110.227312360.142201760.51706634

	SA	SC	SE
Read	0.57	0.15	0.28
	(0.44 <i>,</i> 0.70)	(0.03 <i>,</i> 0.26)	(0.24,0.32)
Write	0.26	0.23	0.52
	(0.08,0.42)	(0.09 <i>,</i> 0.36)	(0.45 <i>,</i> 0.59)

Three Important Concepts

1.- What is the variance due to genetic and (shared and unique) environmental contributions for each trait?Variance Decomposition

Genetic, common environmental and unique genetic factors (including measurement error) contribute to 57%, 15%, and 28% of the individual differences in read, respectively.

Genetic, common environmental and unique genetic factors (including measurement error) contribute to 26%, 23%, and 52% of the individual differences in write, respectively.

Three Important Concepts

1.- What is the variance due to genetic and (shared and unique) environmental contributions for each trait?*Variance Decomposition*

2.- How much of the phenotypic correlation between the traits is accounted for by genetic and (shared and unique) environmental factors?*Covariance Decomposition*

3.- To what extent do the genetic factors underlying each trait overlap?And the (shared and unique) environmental factors?Genetic and Environmental correlations

	SA12	SC12	SE12
Read-write			

> fitACE\$algebras\$SV mxAlgebra 'SV' \$formula: cbind(VA/V, VC/V, VE/V) \$result:

SA SA SC SC SE SE SE SV 0.57018782 0.60226513 0.14994850 0.25553311 0.27986369 0.14220176 SV 0.60226513 0.25562130 0.25553311 0.22731236 0.14220176 0.51706634

	SA12	SC12	SE12
Read-write			

mxAlgebra 'SV'
\$formula: cbind(VA/V, VC/V, VE/V)
\$result:

SASASCSCSESESV0.570187820.602265130.149948500.255533110.279863690.14220176SV0.602265130.255621300.255533110.227312360.142201760.51706634

	SA12	SC12	SE12
Read-write	0.60	0.26	0.14
	(0.40,0.81)	(0.08,0.42)	(0.08,0.21)

mxAlgebra 'SV'
\$formula: cbind(VA/V, VC/V, VE/V)
\$result:

SASASCSCSESESV0.570187820.602265130.149948500.255533110.279863690.14220176SV0.602265130.255621300.255533110.227312360.142201760.51706634

	SA12	SC12	SE12
Read-write	0.60	0.26	0.14
	(0.40,0.81)	(0.08,0.42)	(0.08,0.21)

Remember the relative importance of the covariance of A, C, and E depends on the r_{ph} . If r_{ph} is very low, a covariance of 0.60 due to genetic factors means nothing.

Three Important Concepts

2.- How much of the phenotypic correlation between the traits is accounted for by genetic and (shared and unique) environmental factors?Covariance Decomposition

Genetic, common environmental and unique genetic factors contribute to 60%, 26%, and 14% of the covariance between read and write, respectively.

Three Important Concepts

1.- What is the variance due to genetic and (shared and unique) environmental contributions for each trait?Variance Decomposition

2.- How much of the phenotypic correlation between the traits is accounted for by genetic and (shared and unique) environmental factors? *Covariance Decomposition*

3.- To what extent do the genetic factors underlying each trait overlap?And the (shared and unique) environmental factors?Genetic and Environmental correlations

	r _a	r _c	r _e	r _{ph}
Read-write				

> fitACE\$algebras\$corr

```
mxAlgebra 'corr'
$formula: cbind(rA, rC, rE, rP)
$result:
```

rArArCrCrErErPrPcorr 1.0000000 0.85377382 1.0000000 0.74907828 1.0000000 0.20231178 1.000000 0.541206corr 0.85377382 1.0000000 0.74907828 1.0000000 0.20231178 1.0000000 0.541206 1.000000

	r _a	r _c	r _e	r _{ph}
Read-write	0.85	0.75	0.20	0.54
	(0.62 <i>,</i> 1)	(0.33,1)	(0.11,0.29)	(0.51,0.57)

Three Important Concepts

3.- To what extent do the genetic factors underlying each trait overlap?And the (shared and unique) environmental factors?Genetic and Environmental correlations

The genetic and common shared environment correlations between read and write are large ($r_a = .85$ and $r_c = .75$); the unique environment correlation is low ($r_e = .20$).

85%, 75%, and 20% of the genetic factors, shared environmental factors, and unique environmental factors between read and write overlap.

Correlations. Interpreting Results

- High r_a = large overlap in genetic effects on the two traits
 - If r_a = 1, the two sets of genes influencing each trait overlap completely
- If there is a high r_a, does it mean that the phenotypic correlation between the traits is largely due to genetic effects?

No: The contribution to r_{ph} is a function of both h^2 of the traits and r_a

 i.e. the substantive importance of genetic effects depends on the value of the r_a and the heritability of each phenotype (h²)

Proportion of r_{ph} due to A: –

$$\sqrt{h^2}_{P1} * r_g * \sqrt{h^2}_{P2}$$

 r_{ph}

• Same for r_c and r_e

Consider write-read $r_{ph} = 0.54$ and $r_a = 0.85$ $SA_{read}=0.58$ and $SA_{write}=0.26$

What is the contribution to phenotypic correlation attributable to additive genetic effects?

Proportion of r_{ph} due to A:

$$\frac{\sqrt{h_{P1}^2 * r_g * \sqrt{h_{P2}^2}}}{r_{ph}}$$

• Same for r_c and r_e

Consider write-read $r_{ph} = 0.54$ and $r_a = 0.85$ $SA_{read}=0.58$ and $SA_{write}=0.26$

What is the contribution to phenotypic correlation attributable to additive genetic effects?

Proportion of r_{ph} due to A:

$$\frac{\sqrt{h^2}_{P1} * r_g * \sqrt{h^2}_{P2}}{r_{ph}}$$

• Same for r_c and r_e

Prop of r_{ph} due to A: (sqrt(0.58)*0.85*sqrt(0.26))/0.54 = 0.61 (61%)Prop of r_{ph} due to C: (sqrt(0.15)*0.75*sqrt(0.23))/0.54 = 0.22 (26%)Prop of r_{ph} due to E: (sqrt(0.28)*0.20*sqrt(0.52))/0.54 = 0.15 (15%) Consider write-read $r_{ph} = 0.54$ and $r_a = 0.85$ $SA_{read}=0.58$ and $SA_{write}=0.26$

What is the contribution to phenotypic correlation attributable to additive genetic effects?

Proportion of r_{ph} due to A:

$$\frac{\sqrt{h^2}_{P1} * r_g * \sqrt{h^2}_{P2}}{r_{ph}}$$

• Same for r_c and r_e

Prop of r_{ph} due to A: (sqrt(0.58)*0.85*sqrt(0.26))/0.54 = 0.61 (61%)Prop of r_{ph} due to C: (sqrt(0.15)*0.75*sqrt(0.23))/0.54 = 0.22 (26%)Prop of r_{ph} due to E: (sqrt(0.28)*0.20*sqrt(0.52))/0.54 = 0.15 (15%)

5)		SA12	SC12	SE12
5) 5) 5)	Read-write	0.60 (0.40,0.81)	0.26 (0.08,0.42)	0.14 (0.08,0.21)

Consider write-read $r_{ph} = 0.54$ and $r_a = 0.85$ $SA_{read} = 0.58$ and $SA_{write} = 0.26$

This only makes sense if all the covariances have the same sign

What is the contribution to phenotypic correlation attributable to additive genetic effects?

Proportion of r_{ph} due to A:

$$\frac{\sqrt{h^2}_{P1} * r_g * \sqrt{h^2}_{P2}}{r_{ph}}$$

• Same for r_c and r_e

Prop of r_{ph} due to A: (sqrt(0.58)*0.85*sqrt(0.26))/0.54 = 0.61 (61%)Prop of r_{ph} due to C: (sqrt(0.15)*0.75*sqrt(0.23))/0.54 = 0.22 (26%)Prop of r_{ph} due to E: (sqrt(0.28)*0.20*sqrt(0.52))/0.54 = 0.15 (15%)

()		SA12	SC12	SE12
6)	Read-write	0.60	0.26	0.14
6)		(0.40,0.81)	(0.08,0.42)	(0.08,0.21)

Consider other variables with $r_{ph} = 0.54$ and $r_a = 0.85$ $SA_{P1}=0.10$ and $SA_{P2}=0.15$

What is the contribution to phenotypic correlation attributable to additive genetic effects?

Proportion of r_{ph} due to A:

$$\frac{\sqrt{h^2}_{P1} * r_g * \sqrt{h^2}_{P2}}{r_{ph}}$$

• Same for r_c and r_e

Prop of r_{ph} due to A: (sqrt(0.10)*0.85*sqrt(0.15))/0.54 = 0.20 (20%)

What's the best fitting model?

• Check submodels

Now let's add more variables!

In \\workshop\Faculty\lucia\2020\Wednesday\twoACEvc_practical_3traits.R

		SA			SC		SE			
	Read	ead Write Math_R		Read Write		Math_R	Read	Write	Math_R	
Read										
Write										
Math_R										

	r _a			r _c			r _e			r _{ph}		
	Read	Write	Math_R	Read	Write	Math_R	Read	Write	Math_R	Write	Read	Math_R
Read												
Write												
Math_R												

		SA			SC		SE			
	Read	Write	Math_R	Read	Write	Math_R	Read	Write	Math_R	
Read	0.58			0.15			0.28			
Write	0.62	0.27		0.24	0.22		0.14	0.51		
Math_R	0.69	0.60	0.58	0.22	0.31	0.15	0.09	0.08	0.27	

	r _a			r _c			r _e			r _{ph}		
	Read	Write	Math_R	Read	Write	Math_R	Read	Write	Math_R	Write	Read	Math_R
Read	1			1			1			1		
Write	0.85	1		0.74	1		0.19	1		0.54	1	
Math_R	0.79	0.80	1	0.98	0.86	1	0.22	0.12	1	0.66	0.51	1