

# Resemblance between relatives

Mike Goddard

[mike.goddard@ecodev.vic.gov.au](mailto:mike.goddard@ecodev.vic.gov.au)

# What do we mean by resemble?

Similar values of quantitative traits

Measure by correlation

$$= \text{Covariance}(y_i, y_j) / \text{variance}(y)$$

Why do relatives resemble each other?

# Why do relatives resemble each other?

Similar

Genes

Family environment

Country

School

# Model phenotype

Phenotype = genetic effect

+ country

+ year of birth

+ family environment

Fixed effects

Country, year of birth

Random effects

Genetic effect, family environment

We need a model of the covariances between terms

# Intermezzo: fixed and random effects

# Model phenotype

Phenotype = genetic effect

+ country

+ year of birth

+ family environment

+ individual environment

$$V(\text{phenotype}) = V(\text{genetic effects}) + V(\text{family environment}) \\ + V(\text{individual environment})$$

$$\text{Cov}(\text{phenotype}_i, \text{phenotype}_j) = \text{Cov}(\text{genetic effects}) \\ + \text{Cov}(\text{family environments})$$

# Model phenotype

Random effects

Genetic effect, family environment

We need a model of the covariances between terms

$C(\text{family environments}) = \begin{cases} 0 & \text{if different families} \\ 1 * V_{CE} & \text{if same family} \end{cases}$



# Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive gene action

We need genetic variances and covariances

Genotype	BB	Bb	bb	
Effect	a	0	-a	
Frequency	$p^2$	$2pq$	$q^2$	$(p+q=1)$

$$\text{Mean} = a * p^2 + 0 * 2pq - a * q^2 = (p-q) * a$$

$$\text{Variance (genetic effect)} = \text{genetic variance} = V_G$$

$$= E(\text{effect}^2) - E(\text{effect})^2$$

$$V_G = a^2 * p^2 + 0 * 2pq + a^2 * q^2 - [(p-q) * a]^2 = 2pqa^2$$

# Model with 1 gene, 2 alleles and additive gene action

Covariance between parent and offspring

Parent			Offspring			
Genotype		frequency	BB	Bb	bb	mean
BB	a	$p^2$	p	q		pa
Bb	0	$2pq$	0.5p	0.5	0.5q	$0.5(p-q)a$
bb	-a	$q^2$		p	q	-qa

Cov(parent genetic value, offspring genetic value)

$$= p^2 * a * pa + q^2 * (-a) * (-qa) - [(p-q)a] * [(p-q)a] = pqa^2 = 0.5 V_G$$

# Model with 1 gene, 2 alleles and additive gene action

Covariance between parent and offspring (another way)

Model genetic value as sum of gametic effects from mother and father

$$g = x_m + x_f$$

$$V(g) = V(x_m) + V(x_f) = 2V(x)$$

$$\begin{aligned} C(g_p, g_o) &= C(x_{mp} + x_{fp}, x_{mo} + x_{fo}) \\ &= C(x_{mp}, x_{mo}) + C(x_{mp}, x_{fo}) + C(x_{fp}, x_{mo}) + C(x_{fp}, x_{fo}) \\ &= 0 \qquad \qquad \qquad + ? \qquad \qquad \qquad + 0 \qquad \qquad \qquad + ? \end{aligned}$$

$$\begin{aligned} C(x_{mp}, x_{fo}) &= V(x) \text{ if } x_{mp} \text{ is ibd to } x_{fo} \\ &= 0 \text{ otherwise} \end{aligned}$$

$$C(x_{mp}, x_{fo}) = C(x_{fp}, x_{fo}) = 0.5 V(x)$$

$$C(g_p, g_o) = 0 + 0.5V(x) + 0 + 0.5V(x) = V(x) = 0.5 V_G$$

# Probability that relatives share alleles IBD

Covariance between relatives depends on probability that their alleles are IBD

This probability can be calculated from pedigrees

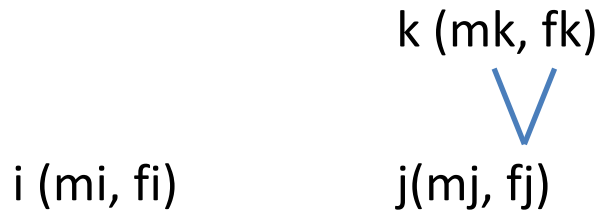
Assume that base individuals at the top of the pedigree (ie those without a pedigree) have unrelated alleles ie the individuals are unrelated

Recurrence formulae for  $P(\text{IBD})$

if  $i$  and  $j$  are base individuals,  $P(x_i \equiv x_j) = 0$

Otherwise,  $P(x_i \equiv x_{fj}) = 0.5 [P(x_i \equiv x_{fk}) + P(x_i \equiv x_{mk})]$  where  $k$  is the father of  $j$

# Probability that relatives share alleles IBD



# Relationships between individuals

Usually we analyse measurements on diploid individuals

$$C(g_i, g_j) = A(i,j) V_G = [C(m_i, m_j) + C(m_i, f_j) + C(f_i, m_j) + C(f_i, f_j)]$$

where  $A$  is the numerator relationship matrix

$$A = 2 * \text{Kinship matrix} = 2 * \text{co-ancestry matrix}$$

# Estimating genetic variance

Data on phenotypes ( $y$ ) of related subjects

$y = \text{fixed effects} + g + e$

$$V(g) = \mathbf{A} V_G$$

$$V(e) = \mathbf{I} V_E$$

Use ML or REML to estimate variances

# Estimating genetic variance

Use ML or REML to estimate variances

ML finds the value of  $V_G$  that maximises the probability of observing the data

ML estimates all parameters together

= estimates variances assuming that fixed effects have been estimated without error

REML allows for loss of df in estimating fixed effects

$$\text{ML } \sigma^2 = \Sigma(y - \text{mean})^2 / N$$

$$\text{REML } \sigma^2 = \Sigma(y - \text{mean})^2 / (N - 1)$$

Little difference unless many fixed effects

Use REML computer programs such as ASREML, GCTA



# Estimating genetic variance

Example: Data on phenotypes ( $y$ ) of full sibs

$y = \text{fixed effects} = g + e$

$\text{Cov}(g_i, g_j) = A(i,j) V_G = 0.5 V_G$  if  $i$  and  $j$  are sibs

Therefore estimate  $V_G$  by  $2\text{cov}(\text{full-sibs})$

$h^2$  by  $2$  correlation between full-sibs

What is the covariance between twins?

# Model with dominance

# Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

We need genetic variances and covariances

Genotype	BB	Bb	bb	
Effect	a	d	-a	
Frequency	$p^2$	$2pq$	$q^2$	$(p+q=1)$

$$\text{Mean} = a * p^2 + d * 2pq - a * q^2 = (p-q) * a + 2pqd$$

$$\text{Variance (genetic effect) = genetic variance} = V_G$$

$$= E(\text{effect}^2) - E(\text{effect})^2$$

$$V_G = a^2 * p^2 + d^2 * 2pq + a^2 * q^2 - [(p-q) * a + 2pqd]^2 = 2pq\alpha^2 + (2pqd)^2$$

$$\text{where } \alpha = a + (q-p)d$$

# Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

but the covariance between relatives doesn't depend directly on  $V_G$ . We need to decompose  $V_G$  into an additive and dominance variance.

Parameterise the genetic value as

$$g = \text{mean} + \text{additive effect} + \text{dominance deviation}$$

$$g = \text{mean} + \text{paternal allele effect} + \text{maternal allele effect} + \text{interaction of alleles}$$

Genotype	BB	Bb	bb	
Effect	a	d	-a	
Frequency	$p^2$	$2pq$	$q^2$	$(p+q=1)$
mean	$(p-q)a + 2pqd$	$(p-q)a + 2pqd$	$(p-q)a + 2pqd$	
additive	$2q\alpha$	$(q-p)\alpha$	$-2p\alpha$	$\alpha=a+(q-p)d$
dominance dev.	$-q^2d$	$2pqd$	$-p^2d$	

Mean(additive effect) = 0, mean(dominance deviation) = 0, cov(additive effect, dominance dev) = 0

$$\begin{aligned} \text{Genetic variance} = V_G &= 2pq\alpha^2 + (2pqd)^2 \\ &= V_A + V_D \end{aligned}$$

# Intermezzo: Genetic variance for the 1-locus Fisher/Falconer model

- Go to

<http://cnsgenomics.com/shiny/Falconer/>

# Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

$$\begin{aligned}\text{Cov}(g_i, g_j) &= \text{Cov}(a_i + d_i, a_j + d_j) = \text{Cov}(a_i, a_j) + \text{cov}(d_i, d_j) \\ &= A(i, j) V_A + D(i, j) V_D\end{aligned}$$

$D(i, j) = \text{prob}(i \text{ and } j \text{ inherit the same genotype IBD})$

Eg

$D(i, j) = 1$  for MZ twins,  $0.25$  for full-sibs,  $0$  for parent and offspring

# Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

Relationships	MZ twins	full-sibs	1/2sibs	P-O
A	1	0.5	0.25	0.5
D	1	0.25	0	0

Therefore can estimate both  $V_A$  and  $V_D$  by using multiple relationships

# Covariance between environmental effects of relatives

$y = \text{mean} + \text{genetic effect} + \text{common environment effect} + \text{individual environment effect}$

$$y = \text{mean} + g + e_c + e$$

Model with a common environmental effect within the same family

$\text{Cov}(e_{ci}, e_{cj}) = V_C$  if  $i$  and  $j$  in same family, zero otherwise

Relationships	MZ twins	full-sibs	1/2sibs	P-O
A	1	0.5	0.25	0.5
D	1	0.25	0	0
E common	1	1	?	?



# Covariance between relatives

Estimating  $V_A$ ,  $V_D$  and  $V_C$

Difficult!

Assume  $V_D = 0$

$$V_A = 2(\text{cov}(\text{MZ twins}) - \text{cov}(\text{full-sibs}))$$

Relationships	MZ twins	full-sibs	1/2sibs	P-O
A	1	0.5	0.25	0.5
D	1	0.25	0	0
E common	1	1	?	?

# Covariance between relatives

Can add epistatic interactions to model

$g = \text{mean} + \text{additive} + \text{dominance} + \text{epistasis}$

eg  $g = \text{mean} + a + d + aa$

$$\text{Cov}(g_i, g_j) = A(i,j) V_A + D(i,j) V_D + A(i,j)^2 V_{AA}$$

Relationships	MZ twins	full-sibs	1/2sibs	P-O
A	1	0.5	0.25	0.5
D	1	0.25	0	0
AxA	1	0.25	0.0625	0.25

# Summary

## Resemblance between relatives

Model phenotypes by fixed effects and random effects including genetic value (additive, dominance, epistatic)

Model covariance of genetic effects by relationship estimated from pedigree (or SNP genotypes)

Estimate genetic variance by REML

# Empirical results on the resemblance between relatives in humans

1. Pearson & Lee 1903
2. From 1903 to 2010: correlations are stable over time
3. Height in Sweden: simple models don't explain empirical results

# ON THE LAWS OF INHERITANCE IN MAN\*.

## I. INHERITANCE OF PHYSICAL CHARACTERS.

By KARL PEARSON, F.R.S., assisted by ALICE LEE, D.Sc.

University College, London.

364

### *On the Laws of Inheritance in Man*

DIAGRAM IV. *Distribution of Stature.*

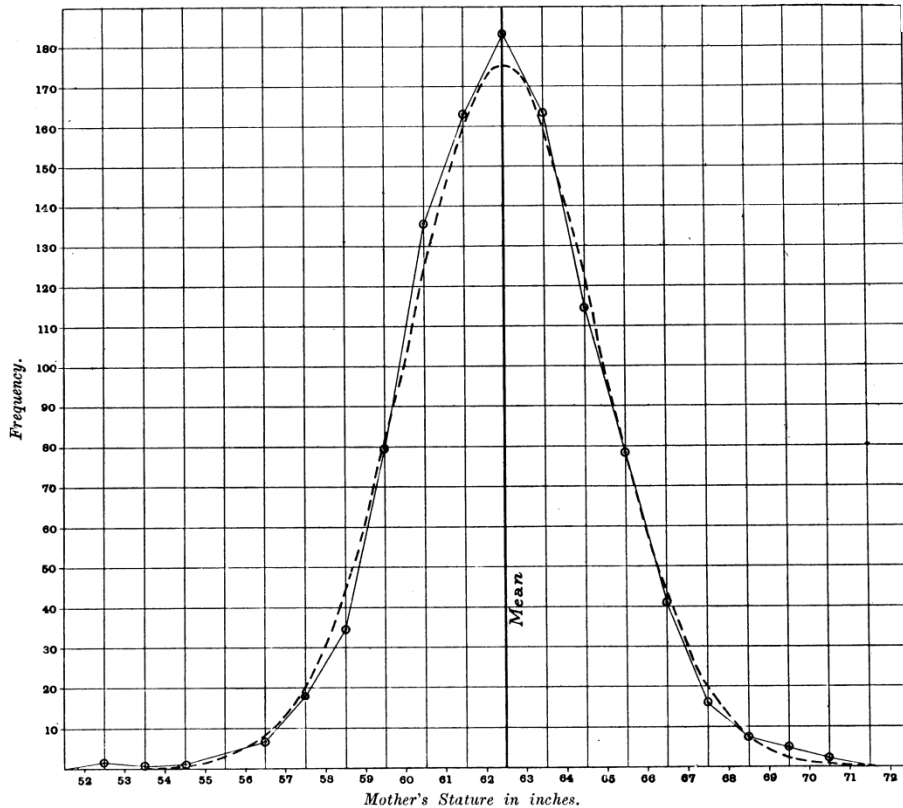
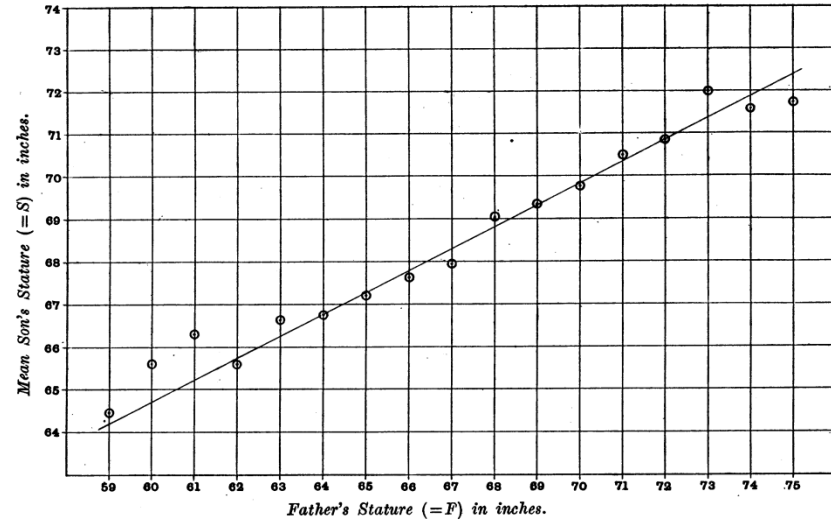


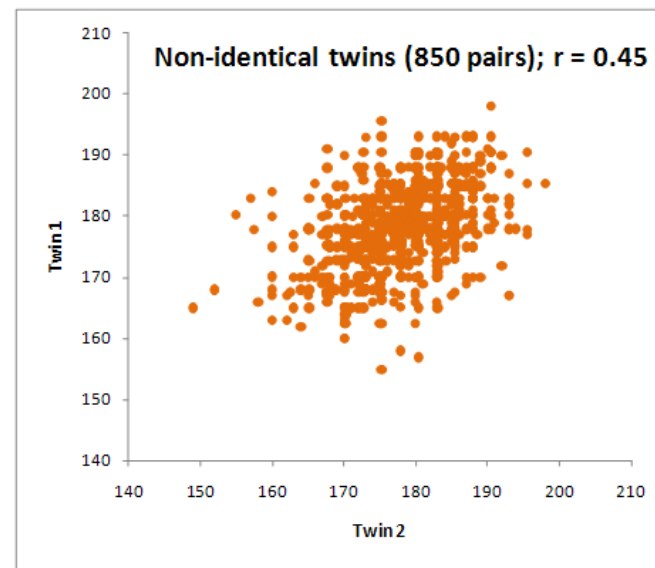
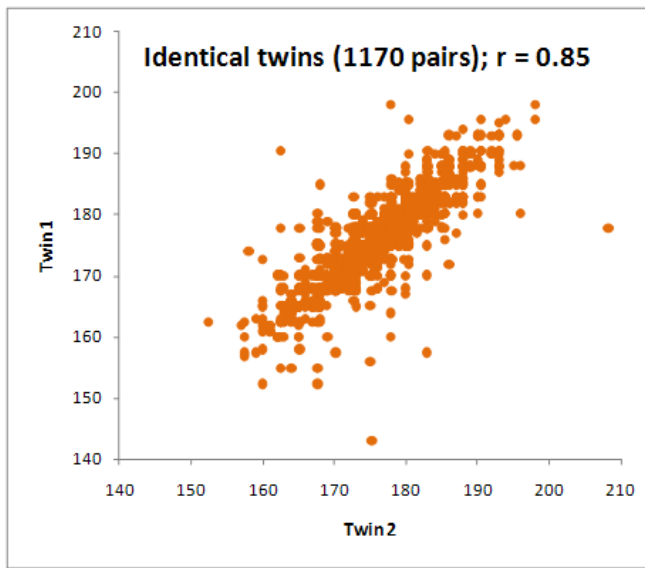
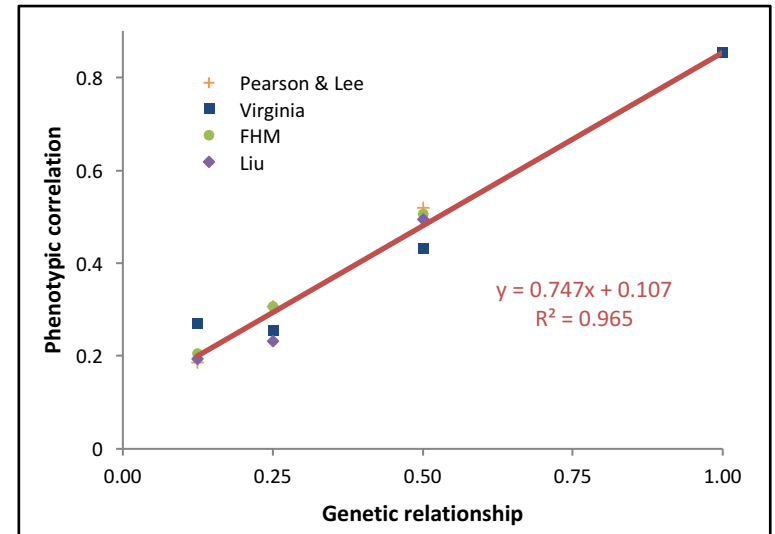
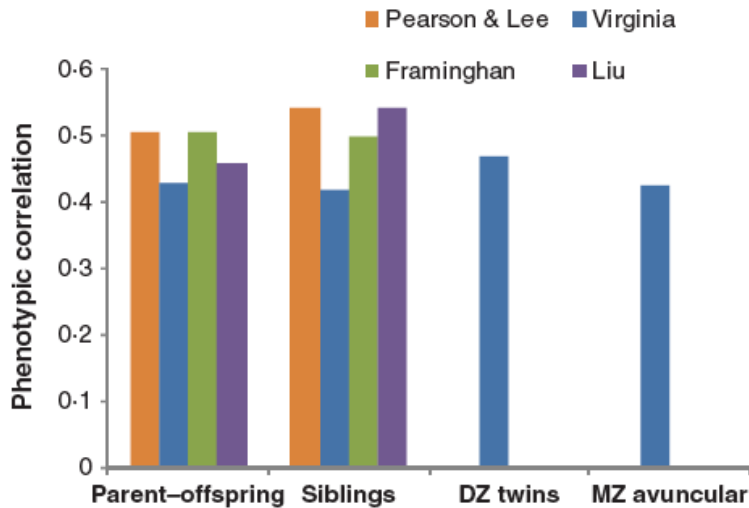
DIAGRAM I. *Probable Stature of Son for given Father's Stature.*

Regression Line:  $S = 33.73 + .516F$ . 1078 Cases.



PAIR	CORRELATION	SE
Spouse	0.28	0.02
Son-Father	0.51	0.02
Daughter-Father	0.51	0.01
Son-Mother	0.49	0.02
Daughter-Mother	0.51	0.01
Brother-brother	0.51	0.03
Sister-sister	0.54	0.02
Brother-sister	0.55	0.01

# Resemblance between relatives (height)



# More data on height

Data from ~172,000 18-year old brother pairs

