Lecture 1:
Fisher's variance decomposition and the resemblance between relatives

Bruce Walsh
Second Bangalore School of
Population Genetics and Evolution
25 Jan- 5 Feb 2016

Measures of Association and variation

- The variance
- The covariance
  - Correlations
  - regressions

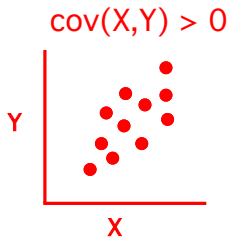
The variance

- The variance of a random variable x measures its spread about its mean, μ_x
- $Var(x) = E[(x-\mu_x)^2]$
  - Average of the squared deviations about the mean
  - Also denoted as V_x, σ_x^2
  - If $\mu_x = 0$, then $Var(x) = E[x^2]$

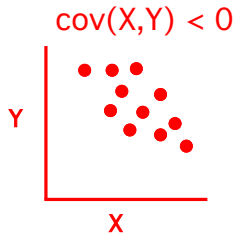
Covariances

- $Cov(x,y) = E [(x-\mu_x)(y-\mu_y)]$
- $= E [x*y] - E[x]*E[y]$

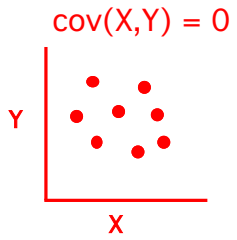
Cov(x,y) > 0, positive (linear) association between x & y



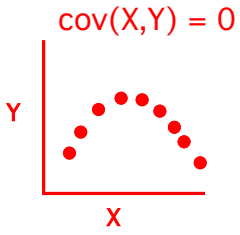
Cov(x,y) < 0, negative (linear) association between x & y



Cov(x,y) = 0, no *linear* association between x & y



Cov(x,y) = 0 DOES NOT imply no association



If x and y are independent, then cov(x,y) = 0

However, cov(x,y) = 0 DOES NOT imply that x and y are independent.

Correlation

Cov = 10 tells us nothing about the strength of an association

What is needed is an absolute measure of association

This is provided by the correlation, r(x,y)

$$r(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)\,Var(y)}}$$

r = 1 implies a perfect (positive) linear association

r = - 1 implies a perfect (negative) linear association

Useful Properties of Variances and Covariances

- Symmetry, Cov(x,y) = Cov(y,x)
- The covariance of a variable with itself is the variance, Cov(x,x) = Var(x)
- If a is a constant, then
  - Cov(ax,y) = a Cov(x,y)
- Var(a x) = a^2 Var(x).
  - Var(ax) = Cov(ax,ax) = a^2 Cov(x,x) = a^2Var(x)
- Cov(x+y,z) = Cov(x,z) + Cov(y,z)

Regressions

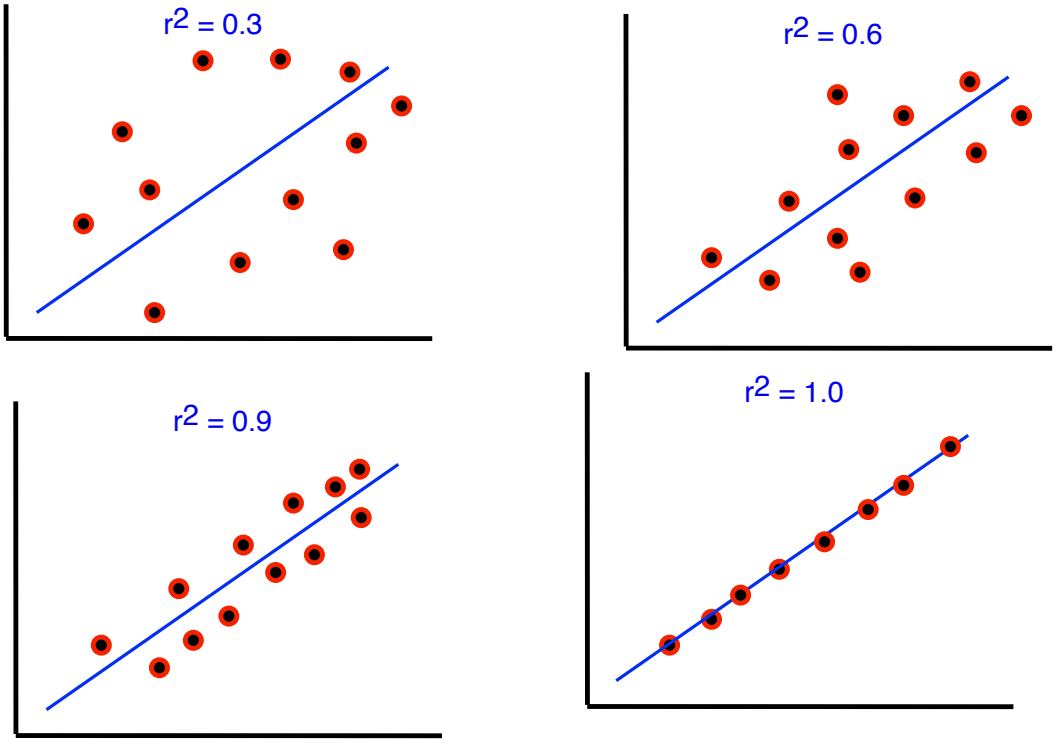
Consider the best (linear) predictor of y given we know x

$$\hat{y} = \bar{y} + b_{y|x}(x - \bar{x})$$

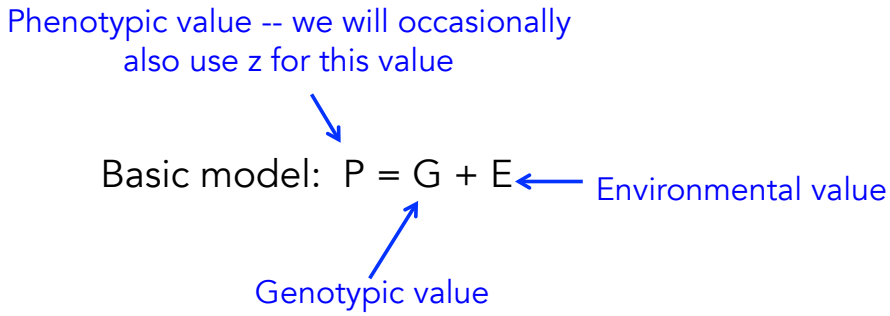
The slope of this linear regression is a function of Cov,

$$b_{y|x} = \frac{Cov(x,y)}{Var(x)}$$

The fraction of the variation in y accounted for by knowing x, i.e,Var(yhat - y), is Var(y) [1-r^2] . Hence, r^2 is the fraction of the total variation in y given by knowing the value of x



Basic model of Quantitative Genetics



G = average phenotypic value for that genotype if we are able to replicate it over the universe of environmental values, G = E[P]

Hence, genotypic values are functions of the environments experienced.

Basic model of Quantitative Genetics

Basic model: P = G + E

G = average phenotypic value for that genotype if we are able to replicate it over the universe of environmental values, G = E[P]

G = average value of an inbred line over a series of environments

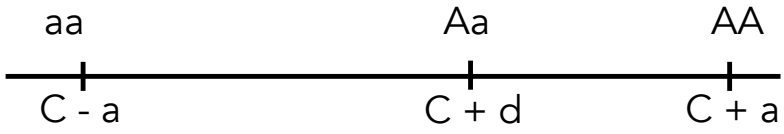
G x E interaction --- The performance of a particular genotype in a particular environment differs from the sum of the average performance of that genotype over all environments and the average performance of that environment over all genotypes. Basic model now becomes P = G + E + GE

The transmission of genotypes versus alleles

- With fully inbred lines, offspring have the same genotype as their parent, and hence the entire parental genotypic value G is passed along
  - Hence, favorable interactions between alleles (such as with dominance) are not lost by randomization under random mating but rather passed along.
- When offspring are generated by crossing (or random mating), each parent contributes a single allele at each locus to its offspring, and hence only passes along a PART of its genotypic value
- This part is determined by the average effect of the allele
  - Downside is that favorable interaction between alleles are NOT passed along to their offspring in a diploid (but, as we will see, are in an autoteraploid)

Genotypic values

It will prove very useful to decompose the genotypic value into the difference between homozygotes (2a) and a measure of dominance (d or k = d/a)



Note that the constant C is the average value of the two homozygotes.

If no dominance, d = 0, as heterozygote value equals the average of the two parents. Can also write d = ka, so that G(Aa) = C + a(1 + k)

Computing a and d

Suppose a major locus influences plant height, with the following values

Genotype	aa	Aa	AA
Trait value	10	15	16

C = [G(AA) + G(aa)]/2 = (16+10)/2 = 13
a = [G(AA) - G(aa)]/2 = (16-10)/2 = 3
d = G(Aa)] - [G(AA) + G(aa)]/2
= G(Aa)] - C = 15 - 13 = 2

Population means: Random mating

Let p = freq(A), q = 1-p = freq(a). Assuming random-mating (Hardy-Weinberg frequencies),

Genotype	aa	Aa	AA
Value	C - a	C + d	C + a
Frequency	q^2	2pq	p^2

Mean = $q^2(C - a) + 2pq(C + d) + p^2(C + a)$

$\mu_{RM} = C + a(p^2-q^2) + d(2pq)$

Contribution from homozygotes

Contribution from heterozygotes

The average effect of an allele

- The average effect α_A of an allele A is defined by the difference between offspring that gets that allele and a random offspring.
  - α_A = mean(offspring value given parent transmits A) - mean(all offspring)
  - Similar definition for α_a.
- Note that while C, a and d (the genotypic parameters) do not change with allele frequency, α_x is clearly a function of the frequencies of alleles with which allele x combines.

Random mating

Consider the average effect of allele A when a parent is randomly-mated to another individual from its population

Suppose parent contributes A

Allele from other parent	Probability	Genotype	Value
A	p	AA	C + a
a	q	Aa	C + d

Mean(A transmitted) = p(C + a) + q(C + d) = C + pa + qd

α_A = Mean(A transmitted) - μ = q[a + d(q-p)]

Random mating

Now suppose parent contributes a

Allele from other parent	Probability	Genotype	Value
A	p	Aa	C + d
a	q	aa	C - a

Mean(a transmitted) = p(C + d) + q(C - a) = C - qa + pd

α_a = Mean(a transmitted) - μ = -p[a + d(q-p)]

Dominance deviations

- Fisher (1918) decomposed the contribution to the genotypic value from a single locus as $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$
  - Here, μ is the mean (a function of p)
  - α_i are the average effects
  - Hence, $\mu + \alpha_i + \alpha_j$ is the predicted genotypic value given the average effect (over all genotypes) of alleles i and j.
  - The dominance deviation associated with genotype G_{ij} is the difference between its true value and its value predicted from the sum of average effects (essentially a residual)

Fisher's (1918) Decomposition of G

One of Fisher's key insights was that the genotypic value consists of a fraction that can be passed from parent to offspring and a fraction that cannot.

In particular, under sexual reproduction, parents only pass along SINGLE ALLELES to their offspring

Consider the genotypic value G_{ij} resulting from an A_iA_j individual

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Average contribution to genotypic value for allele i

Mean value $\mu_G = \Sigma\, G_{ij}\, \text{Freq}(A_iA_j)$

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Since parents pass along single alleles to their offspring, the α_i (the average effect of allele i) represent these contributions

The average effect for an allele is POPULATION-SPECIFIC, as it depends on the types and frequencies of alleles that it pairs with

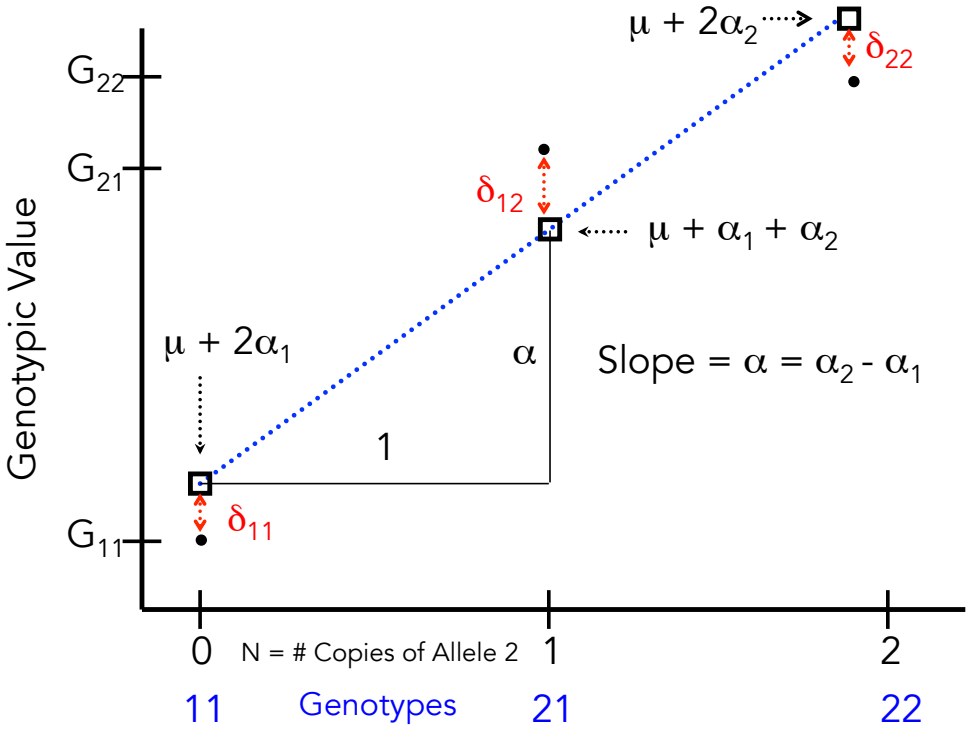
The genotypic value predicted from the individual allelic effects is thus $\hat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

The genotypic value predicted from the individual allelic effects is thus $\hat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

Dominance deviations --- the difference (for genotype A_iA_j) between the genotypic value predicted from the two single alleles and the actual genotypic value,

$$G_{ij} - \hat{G}_{ij} = \delta_{ij}$$



Average Effects and Additive Genetic Values

The α values are the average effects of an allele

A key concept is the Additive Genetic Value (A) of an individual

$$A(G_{ij}) = \alpha_i + \alpha_j$$

$$A = \sum_{k=1}^{n} \left(\alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

$\alpha_i^{(k)}$ = effect of allele i at locus k

A is called the Breeding value or the Additive genetic value

$$A = \sum_{k=1}^{n} \left(\alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

Why all the fuss over A?

Suppose pollen parent has A = 10 and seed parent has A = -2 for plant height

Expected average offspring height is (10-2)/2 = 4 units above the population mean. Offspring A = average of parental A's

KEY: parents only pass single alleles to their offspring. Hence, they only pass along the A part of their genotypic value G

Genetic Variances

Writing the genotypic value as

$$G_{ij} = \mu_G + (\alpha_i + \alpha_j) + \delta_{ij}$$

The genetic variance can be written as

$$\sigma^2(G) = \sum_{k=1}^{n} \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^{n} \sigma^2(\delta_{ij}^{(k)})$$

This follows since

$$\sigma^2(G) = \sigma^2(\mu_g + (\alpha_i + \alpha_j) + \delta_{ij}) = \sigma^2(\alpha_i + \alpha_j) + \sigma^2(\delta_{ij})$$

As Cov(α,δ) = 0

Genetic Variances

$$\sigma^2(G) = \sum_{k=1}^{n} \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^{n} \sigma^2(\delta_{ij}^{(k)})$$

Additive Genetic Variance
(or simply Additive Variance)

Dominance Genetic Variance
(or simply dominance variance)

Hence, total genetic variance = additive + dominance variances,

$$\sigma^2_G = \sigma^2_A + \sigma^2_D$$

Key concepts (so far)

- α_i = average effect of allele i
  - Property of a single allele in a particular population (depends on genetic background)
- A = Additive Genetic Value (A)
  - A = sum (over all loci) of average effects
  - Fraction of G that parents pass along to their offspring
  - Property of an Individual in a particular population
- Var(A) = additive genetic variance
  - Variance in additive genetic values
  - Property of a population
- Can estimate A or Var(A) without knowing any of the underlying genetical detail (forthcoming)

$$\sigma_A^2 = 2E[\alpha^2] = 2\sum_{i=1}^{m} \alpha_i^2 p_i$$

Q_1Q_1	Q_1Q_2	Q_2Q_2
0	a(1+k)	2a

Since $E[\alpha] = 0$,
$Var(\alpha) = E[(\alpha - \mu_a)^2] = E[\alpha^2]$

One locus, 2 alleles: $\sigma_A^2 = 2p_1 p_2 a^2[1 + k(p_1 - p_2)]^2$

Dominance alters additive variance

When dominance present, Additive variance is an asymmetric function of allele frequencies

Dominance variance

Q_1Q_1	Q_1Q_2	Q_2Q_2
0	a(1+k)	2a

$$\sigma_D^2 = E[\delta^2] = \sum_{i=1}^{m}\sum_{j=1}^{m} \delta_{ij}^2 p_i p_j$$

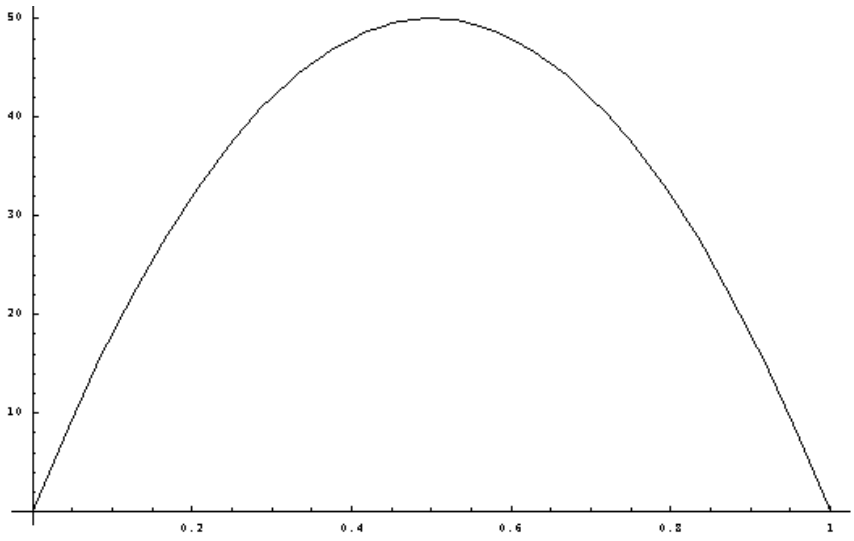
Equals zero if k = 0

One locus, 2 alleles: $\sigma_D^2 = (2p_1 p_2 ak)^2$

This is a symmetric function of allele frequencies

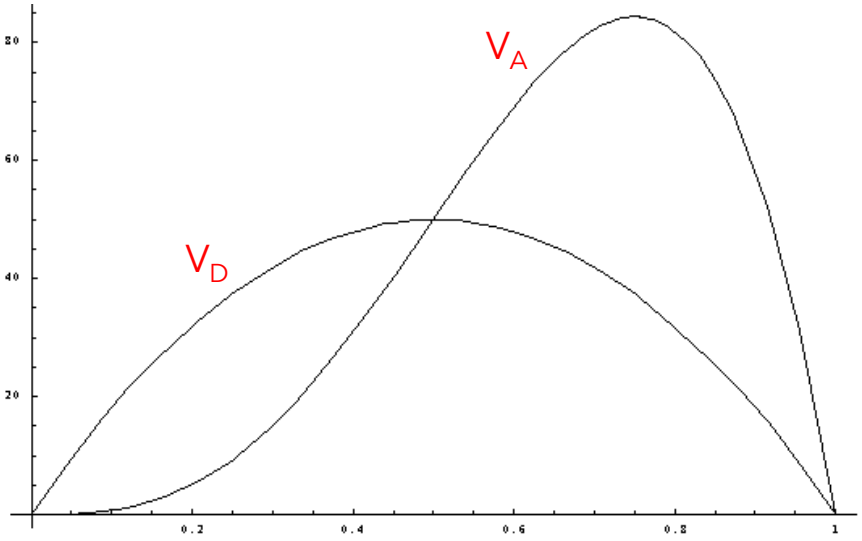
Can also be expressed in terms of d = ak

Additive variance, V_A, with no dominance (k = 0)



Allele frequency, p

Complete dominance (k = 1)



Allele frequency, p

Resemblance between relatives

Heritability

- Central concept in quantitative genetics
- Fraction of phenotypic variance due to additive genetic values (Breeding values)
  - $h^2 = V_A/V_P$
  - This is called the narrow-sense heritability
  - Phenotypes (and hence V_P) can be directly measured
  - Breeding values (and hence V_A) must be estimated
- Estimates of V_A require known collections of relatives

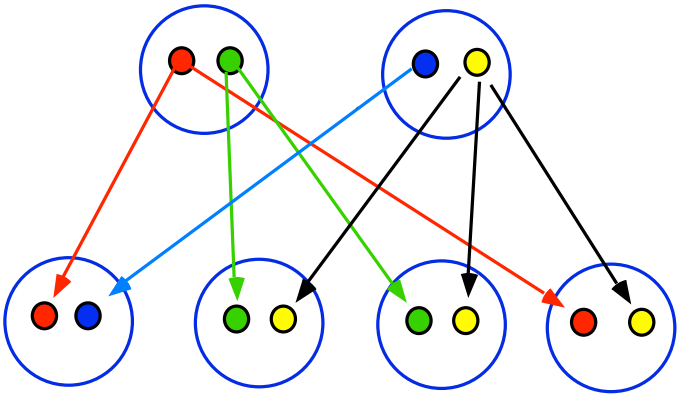
Key observations

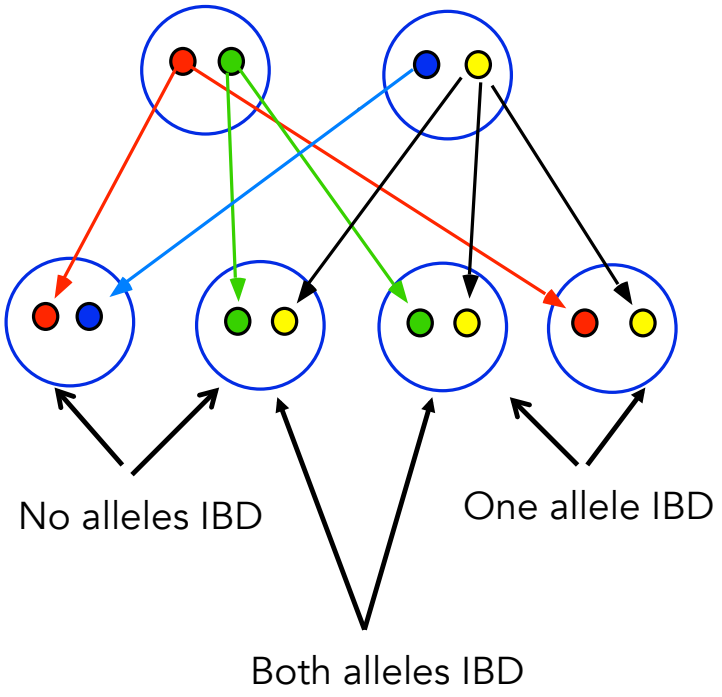
- The amount of phenotypic resemblance among relatives for the trait provides an indication of the amount of genetic variation for the trait.
- If trait variation has a significant genetic basis, the closer the relatives, the more similar their appearance
- The covariance between the phenotypic value of relatives measures the strength of this similarity, with larger Cov = more similarity

Genetic Covariance between relatives

Sharing alleles means having alleles that are identical by descent (IBD): both copies can be traced back to a single copy in a recent common ancestor.

Genetic covariances arise because two related individuals are more likely to share alleles than are two unrelated individuals.





Resemblance between relatives and variance components

- The phenotypic variance between relatives can be expressed in terms of genetic variance components
  - $Cov(z_x,z_y) = a_{xy}V_A + b_{xy}V_D$.
  - The weights a and b depend on the nature of the relatives x and y, and are measures of how often they are expected to share alleles identical by descent
  - These are critical in predicting selection response

Parent-offspring genetic covariance

$Cov(G_p, G_o)$ --- Parents and offspring share EXACTLY one allele IBD

Denote this common allele by A_1

$$G_p = A_p + D_p = \alpha_1 + \alpha_x + D_{1x}$$

$$G_o = A_o + D_o = \alpha_1 + \alpha_y + D_{1y}$$

IBD allele

Non-IBD alleles

$$
\begin{aligned}
Cov(G_o, G_p) &= Cov(\alpha_1 + \alpha_x + D_{1x}, \alpha_1 + \alpha_y + D_{1y} \\
&= Cov(\alpha_1, \alpha_1) + Cov(\alpha_1, \alpha_y) + Cov(\alpha_1, D_{1y}) \\
&+ Cov(\alpha_x, \alpha_1) + Cov(\alpha_x, \alpha_y) + Cov(\alpha_x, D_{1y}) \\
&+ Cov(D_{1x}, \alpha_1) + Cov(D_{1x}, \alpha_y) + Cov(D_{1x}, D_{1y})
\end{aligned}
$$

All blue covariance terms are zero.

- By construction, α and D are uncorrelated
  - By construction, α from non-IBD alleles are uncorrelated
  - By construction, D values are uncorrelated unless both alleles are IBD

14

$$
Cov(\alpha_x, \alpha_y) = \begin{cases} 0 & \text{if } x \neq y, \quad \text{i.e., not IBD} \\ Var(A)/2 & \text{if } x = y, \quad \text{i.e., IBD} \end{cases}
$$

$$
Var(A) = Var(\alpha_1 + \alpha_2) = 2Var(\alpha_1)
$$

so that

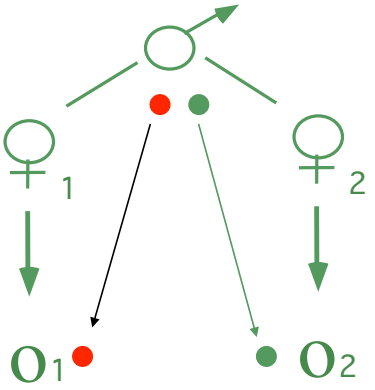
$$
Var(\alpha_1) = Cov(\alpha_1, \alpha_1) = Var(A)/2
$$

Hence, relatives sharing one allele IBD have a genetic covariance of Var(A)/2

The resulting parent-offspring genetic covariance becomes $Cov(G_p, G_o) = Var(A)/2$

Half-sibs

Each sib gets exactly one allele from common father, different alleles from the different mothers



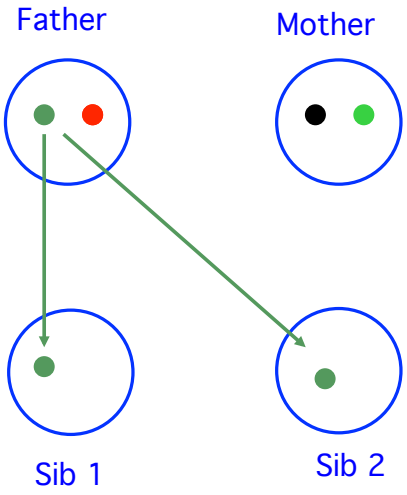
The half-sibs share no alleles IBD

- occurs with probability 1/2

Hence, the genetic covariance of half-sibs is just (1/2)Var(A)/2 = Var(A)/4

Full-sibs

Each sib gets exact one allele from each parent

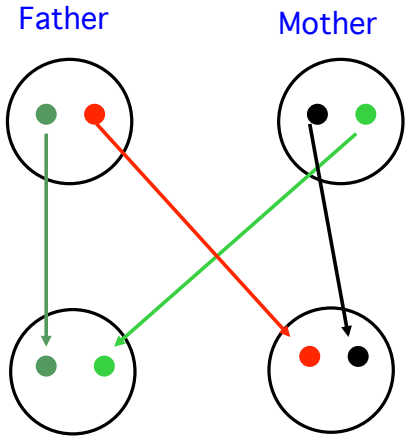


Prob(Allele from father IBD) = 1/2. Given the allele in parent one, prob = 1/2 that sib 2 gets same allele

Prob(Allele from father not IBD) = 1/2. Given the allele in parent one, prob = 1/2 that sib 2 gets different allele

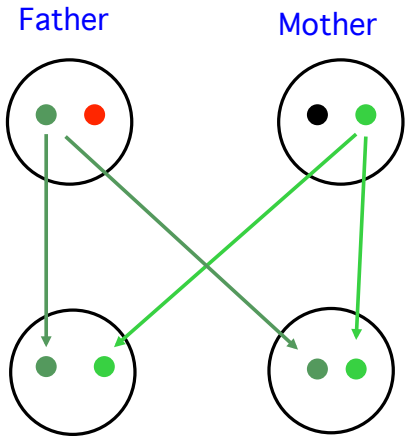
Full-sibs

Each sib gets exact one allele from each parent



Paternal allele not IBD [Prob = 1/2]
Maternal allele not IBD [Prob = 1/2]
Prob(sibs share 0 alleles IBD) = 1/2*1/2 = 1/4

Each sib gets exact one allele from each parent



Paternal allele IBD [Prob = 1/2]
Maternal allele IBD [Prob = 1/2]
Prob(sibs share 2 alleles IBD) = 1/2*1/2 = 1/4

Prob(share 1 allele IBD) = 1-Pr(0) - Pr(2) = 1/2

Resulting Genetic Covariance between full-sibs

IBD alleles	Probability	Contribution
0	1/4	0
1	1/2	Var(A)/2
2	1/4	Var(A) + Var(D)

Cov(Full-sibs) = Var(A)/2 + Var(D)/4

Genetic Covariances for General Relatives

Let r = (1/2)Prob(1 allele IBD) + Prob(2 alleles IBD)

Let u = Prob(both alleles IBD)

General genetic covariance between relatives
Cov(G) = rVar(A) + uVar(D)

When epistasis is present, additional terms appear
r^2Var(AA) + ruVar(AD) + u^2Var(DD) + r^3Var(AAA) +

More general relationships

- To obtain the expected covariance for any set of relatives, we normally need only compute r and u for that set of relatives
- With general inbreeding, becomes more complex (as three other terms, in addition to V_A and V_D arise --- not discussed here, see WL chapter 11 for details)
- With crosses involving inbred and/or related parents, values for r and u are different from those presented above.