

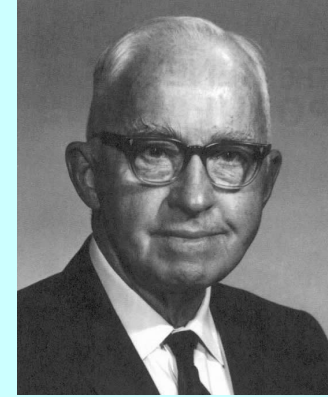
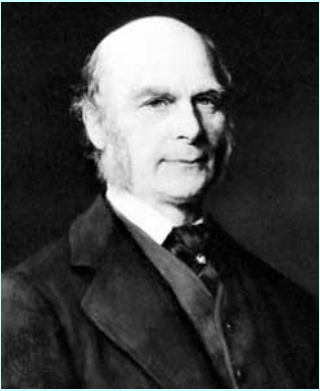
Introduction; variance components and heritability

Friday Harbor Laboratories, 10 June 2019

Joe Felsenstein

Evolutionary Quantitative Genetics Workshop

Development of quantitative genetics theory



Empirical statistical models (their “Law of Ancestral Heredity”) by Galton and Pearson from 1889 on ...

... were overwhelmed by R. A. Fisher’s 1918 Mendelian tour-de-force

... which Jay Lush and his colleagues made the center of modern animal and plant breeding

We use the standard quantitative genetic model

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

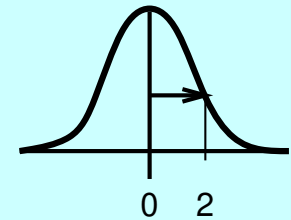
AA Bb Cc dd Ee

Aa bb cc DD ee

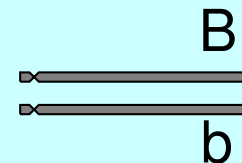
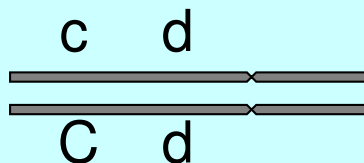
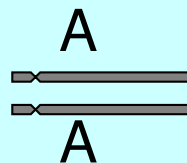
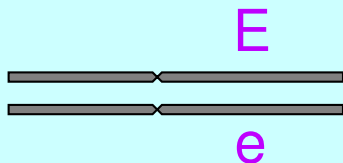
aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee



4 + 0.3



We use the standard quantitative genetic model

$$P = \mu + \begin{Bmatrix} \text{AA} & -2 \\ \text{Aa} & 0 \\ \text{aa} & 3 \end{Bmatrix} + \begin{Bmatrix} \text{BB} & 0.6 \\ \text{Bb} & 0.1 \\ \text{bb} & -0.2 \end{Bmatrix} + \begin{Bmatrix} \text{CC} & -1 \\ \text{Cc} & 6 \\ \text{cc} & 6 \end{Bmatrix} + \begin{Bmatrix} \text{DD} & 0.3 \\ \text{Dd} & 0.3 \\ \text{dd} & 0.7 \end{Bmatrix} + \begin{Bmatrix} \text{EE} & -0.4 \\ \text{Ee} & 0.3 \\ \text{ee} & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

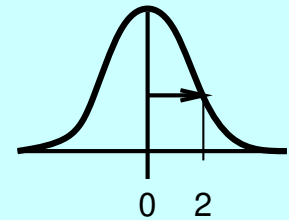
AA Bb Cc dd Ee

Aa bb cc DD ee

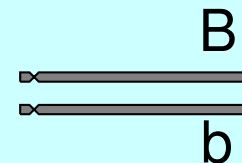
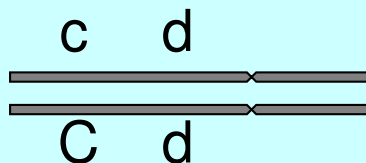
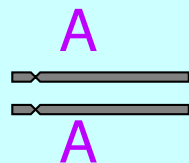
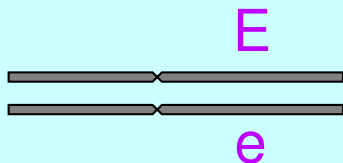
aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee



$$4 + 0.3 - 2$$



We use the standard quantitative genetic model

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

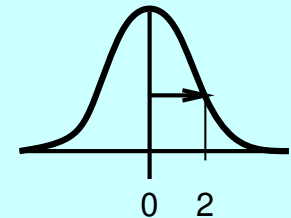
AA Bb Cc dd Ee

Aa bb cc DD ee

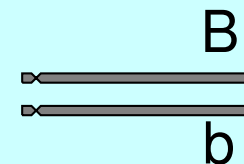
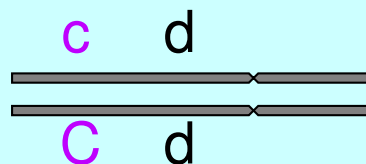
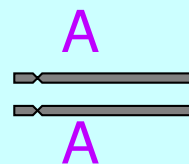
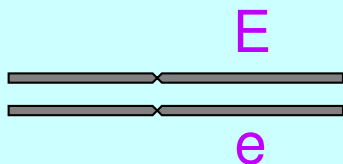
aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee



$$4 + 0.3 - 2 + 6$$



We use the standard quantitative genetic model

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

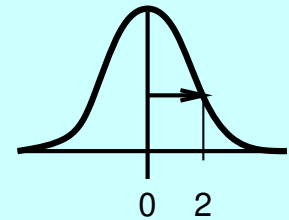
AA Bb Cc dd Ee

Aa bb cc DD ee

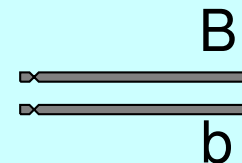
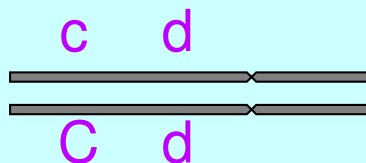
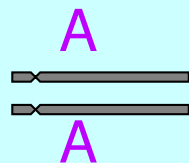
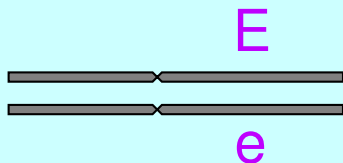
aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee



$$4 + 0.3 - 2 + 6 + 0.7$$



We use the standard quantitative genetic model

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

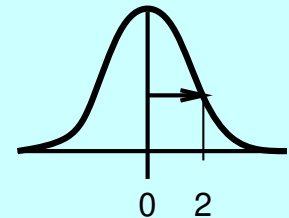
AA Bb Cc dd Ee

Aa bb cc DD ee

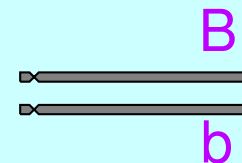
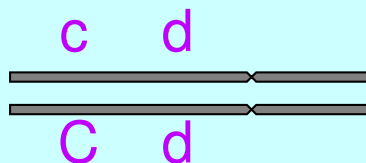
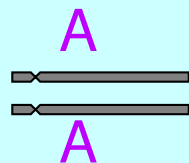
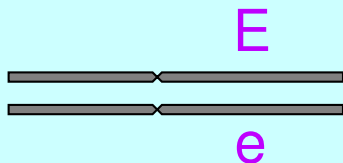
aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee

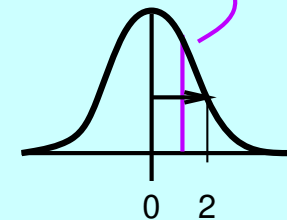


$$4 + 0.3 - 2 + 6 + 0.7 + 0.1$$



... which leads to various character values

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$



arbitrary starting point (4)

AA Bb Cc dd Ee 10

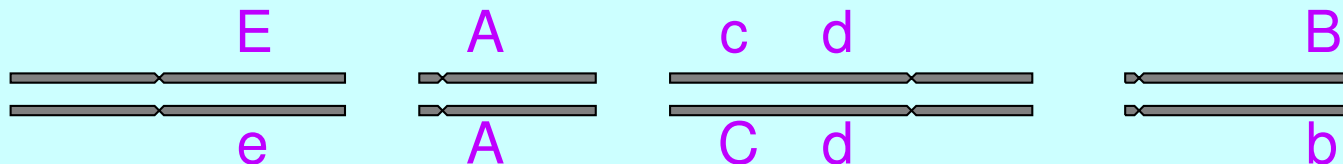
Aa bb cc DD ee

aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee

$$4 + 0.3 - 2 + 6 + 0.7 + 0.1 + 0.9$$



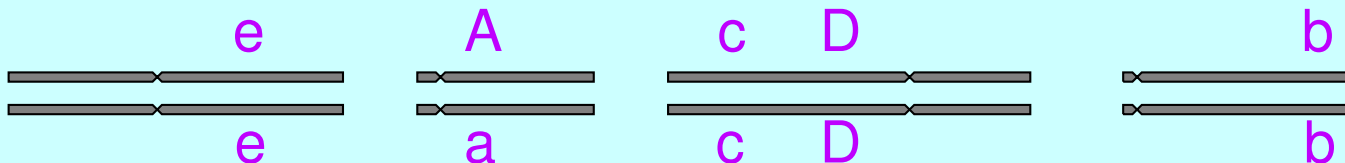
... which leads to various character values

$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

AA	Bb	Cc	dd	Ee	10
Aa	bb	cc	DD	ee	8.8
aa	bb	CC	DD	Ee	
aa	bb	Cc	DD	EE	
Aa	Bb	Cc	DD	Ee	

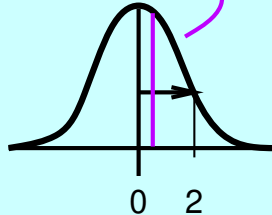
$$4 - 0.3 + 0 + 6 + 0.3 - 0.2 - 1.0$$



... which leads to various character values

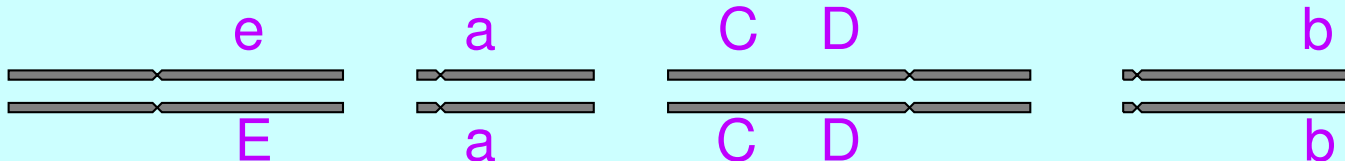
$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)



AA	Bb	Cc	dd	Ee	10
Aa	bb	cc	DD	ee	8.8
aa	bb	CC	DD	Ee	6.5
aa	bb	Cc	DD	EE	
Aa	Bb	Cc	DD	Ee	

$$4 + 0.3 + 3 - 1 + 0.3 - 0.2 + 0.1$$

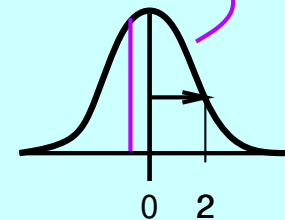


... which leads to various character values

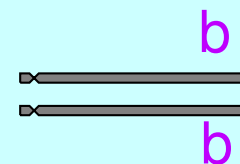
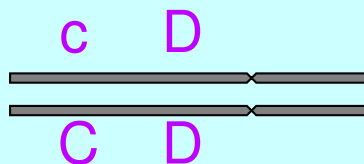
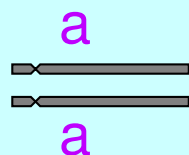
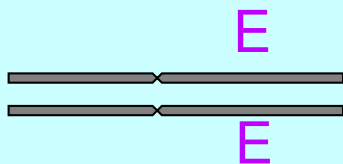
$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

AA	Bb	Cc	dd	Ee	10
Aa	bb	cc	DD	ee	8.8
aa	bb	CC	DD	Ee	6.5
aa	bb	Cc	DD	EE	12.2
Aa	Bb	Cc	DD	Ee	



$$4 - 0.4 + 3 + 6 + 0.3 - 0.2 - 0.5$$

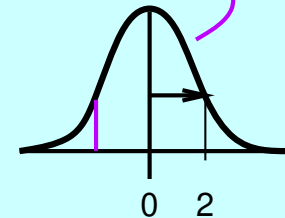


... which leads to various character values

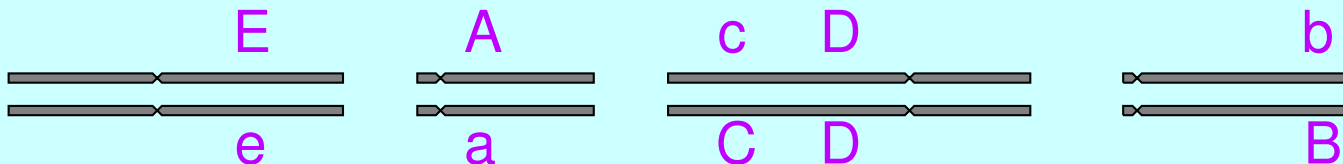
$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

arbitrary starting point (4)

AA	Bb	Cc	dd	Ee	10
Aa	bb	cc	DD	ee	8.8
aa	bb	CC	DD	Ee	6.5
aa	bb	Cc	DD	EE	12.2
Aa	Bb	Cc	DD	Ee	8.9



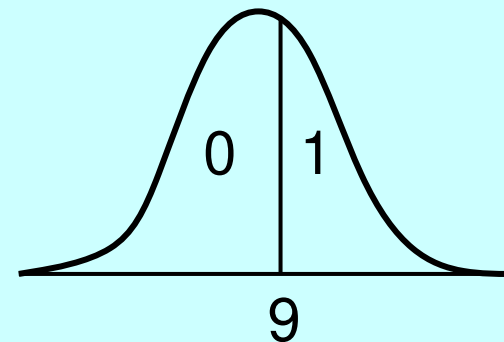
$$4 + 0.3 + 0 + 6 + 0.3 + 0.1 - 1.8$$



... to make a 0/1 character use a threshold of, for example, 9

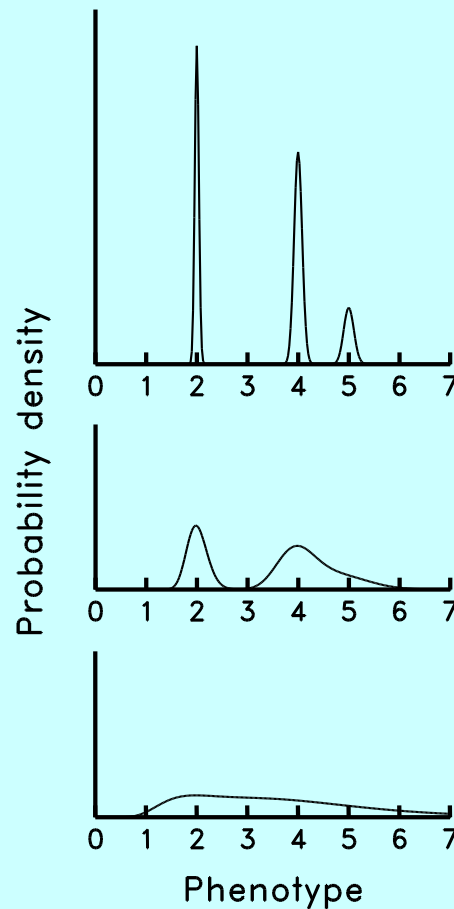
$$P = \mu + \begin{Bmatrix} AA & -2 \\ Aa & 0 \\ aa & 3 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$

AA	Bb	Cc	dd	Ee	10	→	1
Aa	bb	cc	DD	ee	8.8	→	0
aa	bb	CC	DD	Ee	6.5	→	0
aa	bb	Cc	DD	EE	12.2	→	1
Aa	Bb	Cc	DD	Ee	8.9	→	0



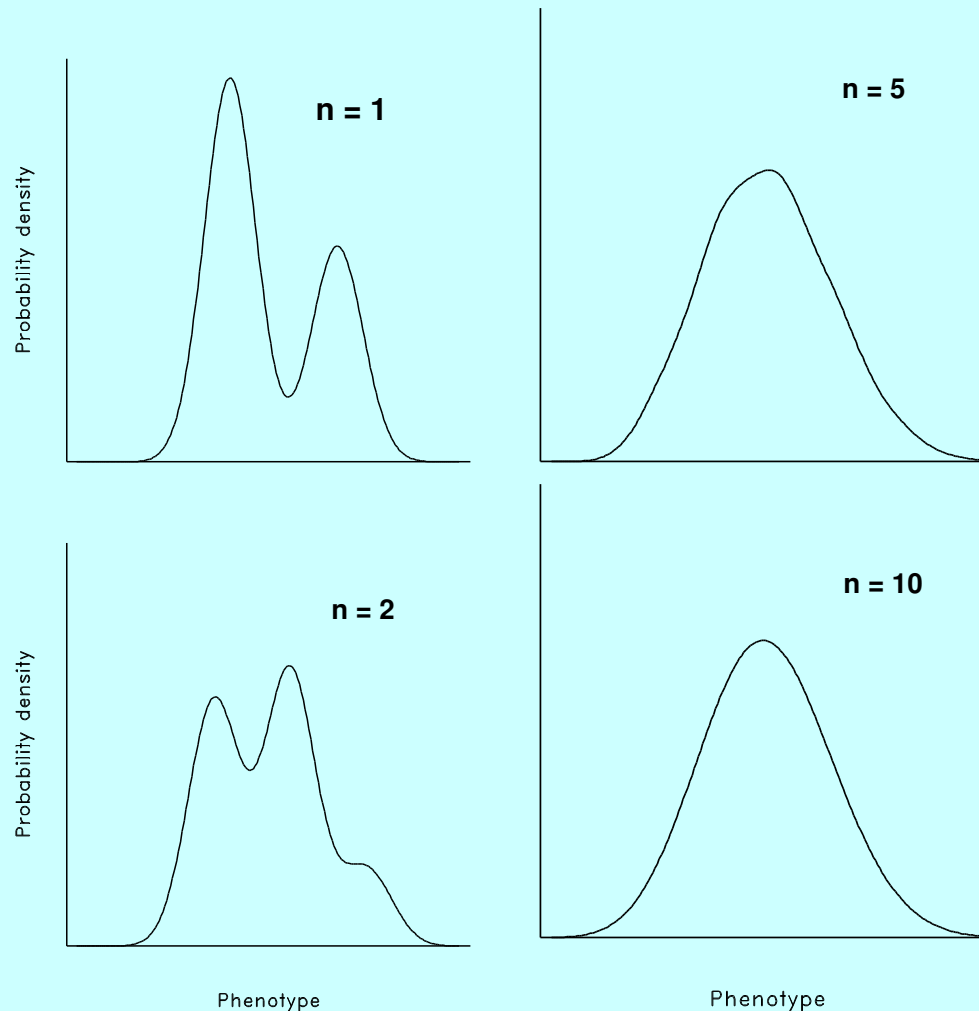
(This is Sewall Wright's threshold model, of which more later).

A single mendelian locus with environmental effects



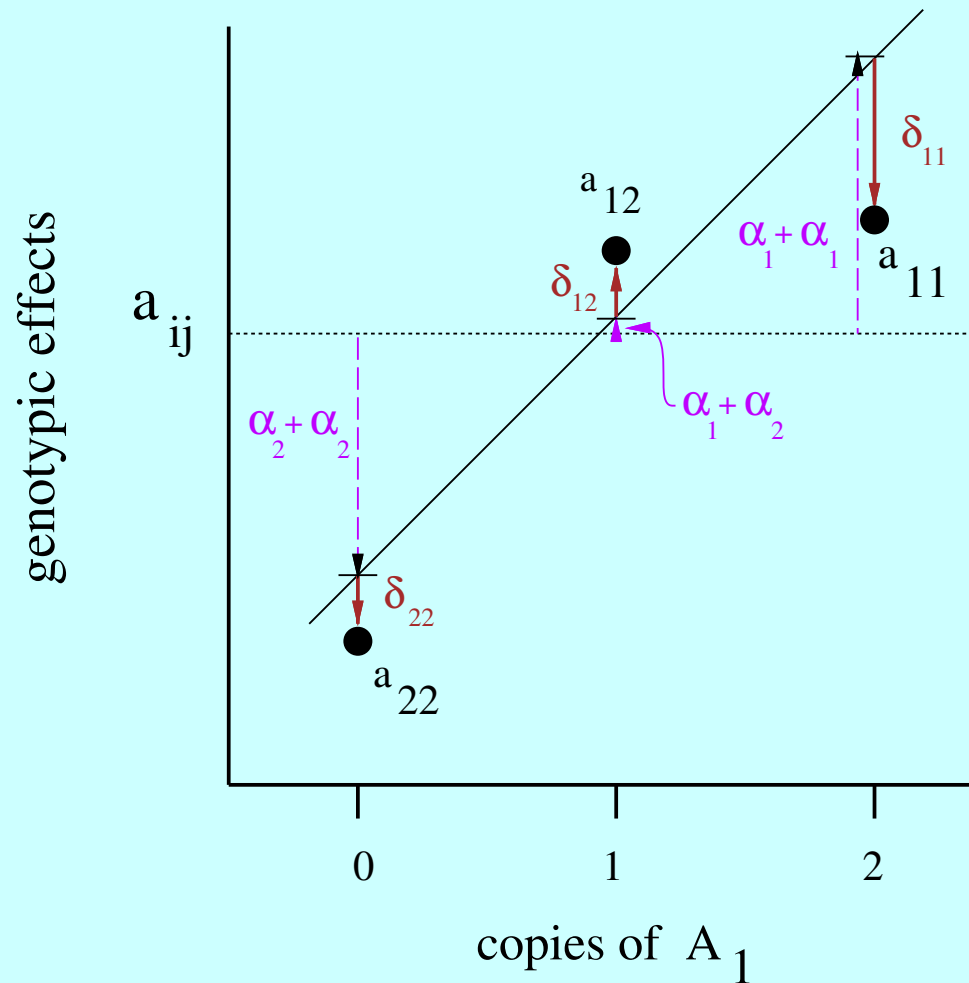
In this arbitrary example, each genotype has background environmental variance distributed in a lognormal distribution. Those environmental variables get larger as we move from the top figure downwards.

Or many mendelian loci (and environmental effect)



(Each locus is dominant and all have the same effect and same gene frequency, but with the effect sizes scaled in each case so that the cases to have the same total genetic variance and the same environmental variance, with 30% of the phenotypic variance being genetic).

Additive effects and dominance deviations



R. A. Fisher (1918) defined additive effects α_i and dominance deviations δ_{ij} by fitting a straight line through the genotypic mean phenotypes, with points weighted by their Hardy-Weinberg population genotype frequency.

Fisher's regression partitions the variance

Fisher's model assumes

- Random mating, so Hardy-Weinberg proportions

Fisher's regression partitions the variance

Fisher's model assumes

- Random mating, so Hardy-Weinberg proportions
- Linkage equilibrium between all loci

Fisher's regression partitions the variance

Fisher's model assumes

- Random mating, so Hardy-Weinberg proportions
- Linkage equilibrium between all loci
- Environmental effects independent of the genotype

Fisher's regression partitions the variance

Fisher's model assumes

- Random mating, so Hardy-Weinberg proportions
- Linkage equilibrium between all loci
- Environmental effects independent of the genotype
- Environmental effects independent in different individuals

Fisher's regression partitions the variance

Fisher's model assumes

- Random mating, so Hardy-Weinberg proportions
- Linkage equilibrium between all loci
- Environmental effects independent of the genotype
- Environmental effects independent in different individuals
- No interaction between loci (effects are additive between loci)

As a result, the following are uncorrelated

- The two α 's in a locus (by Hardy-Weinberg)

As a result, the following are uncorrelated

- The two α 's in a locus (by Hardy-Weinberg)
- The α 's and the δ in a locus, as a result of having defined them by the weighted least squares regression (this is OK, trust me)

As a result, the following are uncorrelated

- The two α 's in a locus (by Hardy-Weinberg)
- The α 's and the δ in a locus, as a result of having defined them by the weighted least squares regression (this is OK, trust me)
- All the effects (α 's and δ 's) at different loci (by linkage equilibrium)

As a result, the following are uncorrelated

- The two α 's in a locus (by Hardy-Weinberg)
- The α 's and the δ in a locus, as a result of having defined them by the weighted least squares regression (this is OK, trust me)
- All the effects (α 's and δ 's) at different loci (by linkage equilibrium)
- The environmental effect (by independence of the environment from the genotype)

So the phenotype is a sum of terms, all of which are uncorrelated (have zero covariance):

$$\begin{aligned} P = & \mu + \alpha_i^{(1)} + \alpha_j^{(1)} + \delta_{ij}^{(1)} + \alpha_k^{(2)} + \alpha_\ell^{(2)} + \delta_{k\ell}^{(2)} \\ & + \alpha_m^{(3)} + \alpha_n^{(3)} + \delta_{mn}^{(3)} + \dots + \epsilon \end{aligned}$$

Rearranging and adding up variances

$$P = \mu + (\alpha^{(1)} + \alpha^{(1)} + \alpha^{(2)} + \alpha^{(2)} + \alpha^{(3)} + \alpha^{(3)} + \dots) \\ + (\delta^{(1)} + \delta^{(2)} + \delta^{(3)}) + \dots + \epsilon$$

(The $\mu +$ (all the α terms) is also called the “breeding value” of the individual).

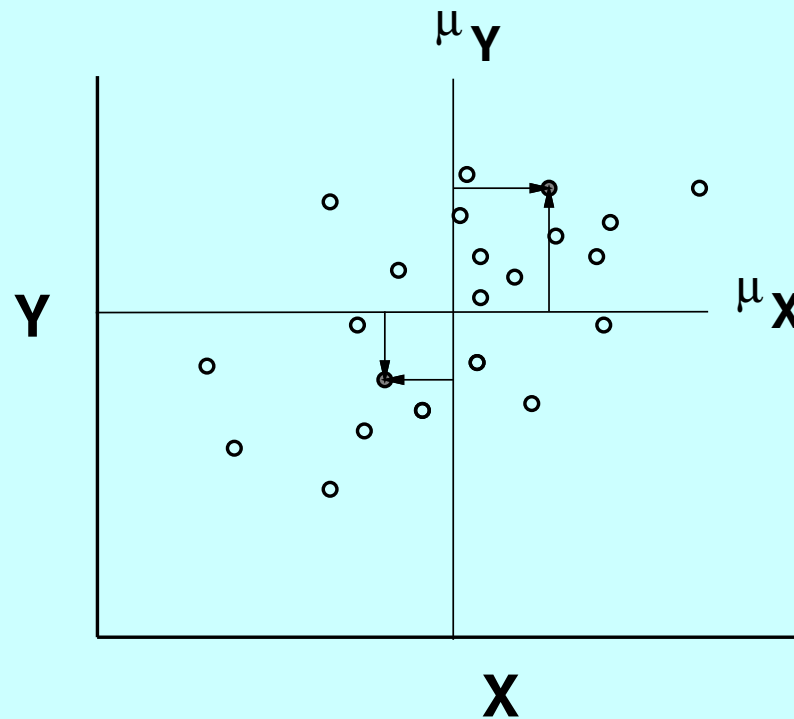
Since random variables that are uncorrelated have variances that sum up, we have partitioned the phenotype into four parts (one a nonvarying starting point μ),

$$P = \mu + A + D + E$$

and that means their variances add up too (the variance of μ being 0):

$$\text{Var}(P) = V_A + V_D + V_E$$

An aside – what are covariances?



They are analogous to variances, but for two traits. The variance of X is the expectation of the square of the difference of X from its mean, $E[X - \mu_X]^2$. For the two variables X and Y , their covariance is the expectation of the product of $X - \mu_X$ and $Y - \mu_Y$. The covariance of two variables is the numerator of their correlation. The correlation is also the covariance of the variables after they have each been standardized.

Covariances and fraction of shared effects

If X , Y , and Z are independent random variables,

$$\text{Cov}[X + Y, X + Z] = \text{Var}(X)$$

So the covariance is the variance of the part shared between the two sums.

$$\begin{array}{rcccl}
 P_x & = & \mu & + (\alpha + \alpha + \alpha + \alpha + \dots) & + (\delta + \delta + \delta + \dots) & + \varepsilon \\
 & & & \uparrow & & \uparrow \\
 & & & f_1 \text{ of these shared} & & f_2 \text{ of these shared} \\
 & & & \downarrow & & \downarrow \\
 P_y & = & \mu & + (\alpha + \alpha + \alpha + \alpha + \dots) & + (\delta + \delta + \delta + \dots) & + \varepsilon
 \end{array}$$

If a fraction f_1 of the α are shared, that fraction of the variance from them is shared, and similarly for the δ s, and environment is not shared,

$$\text{Cov}(P_x, P_y) = f_1 V_A + f_2 V_D$$

Variance components and ANOVA

Does all this look a lot like analysis of variance (ANOVA)?

It ought to: that was also developed by R. A. Fisher in the same era, and published in the early 1920s, very soon after his 1918 paper.

- Each additive effect is a single-factor (“row” or “column”) effect
- The dominance deviations are two-way interaction effects
- All interaction effects between loci, and between the environment and anything else, have been forced to be zero by the assumed model
- The least squares regression is the same one used to assign fixed effects in ANOVA

Variance components and covariances of relatives

This can be done by computing, from the kind of relationship:

- the probability that a single copy of a gene is also found in the relative, owing to identity by descent (f_1)
- the probability that both copies at a locus are found in the relative (f_2)

These give the fraction of V_A and of V_D that contribute to the covariance of pairs of relatives.

$$\text{Cov}(\text{this kind of relative}) = f_1 V_A + f_2 V_D$$

It establishes algebraic relationships between covariances of different kinds of relatives (and the associated correlation coefficients).

The important case for us is parents and offspring ($f_1 = 1/2$, $f_2 = 0$):

$$\text{Cov}(P, O) = \frac{1}{2} V_A$$

Coefficients f_1 and f_2 in relatives

Relationship	f_1	f_2	
Oneself	1	1	*
Identical twin	1	1	
Parent	1/2	0	
Offspring	1/2	0	
Full sib	1/2	1/4	
Half sib	1/4	0	
Aunt/Uncle	1/4	0	
Niece/Nephew	1/4	0	
First cousin	1/8	0	

* for oneself, the environmental variance V_E is also shared

Heritability

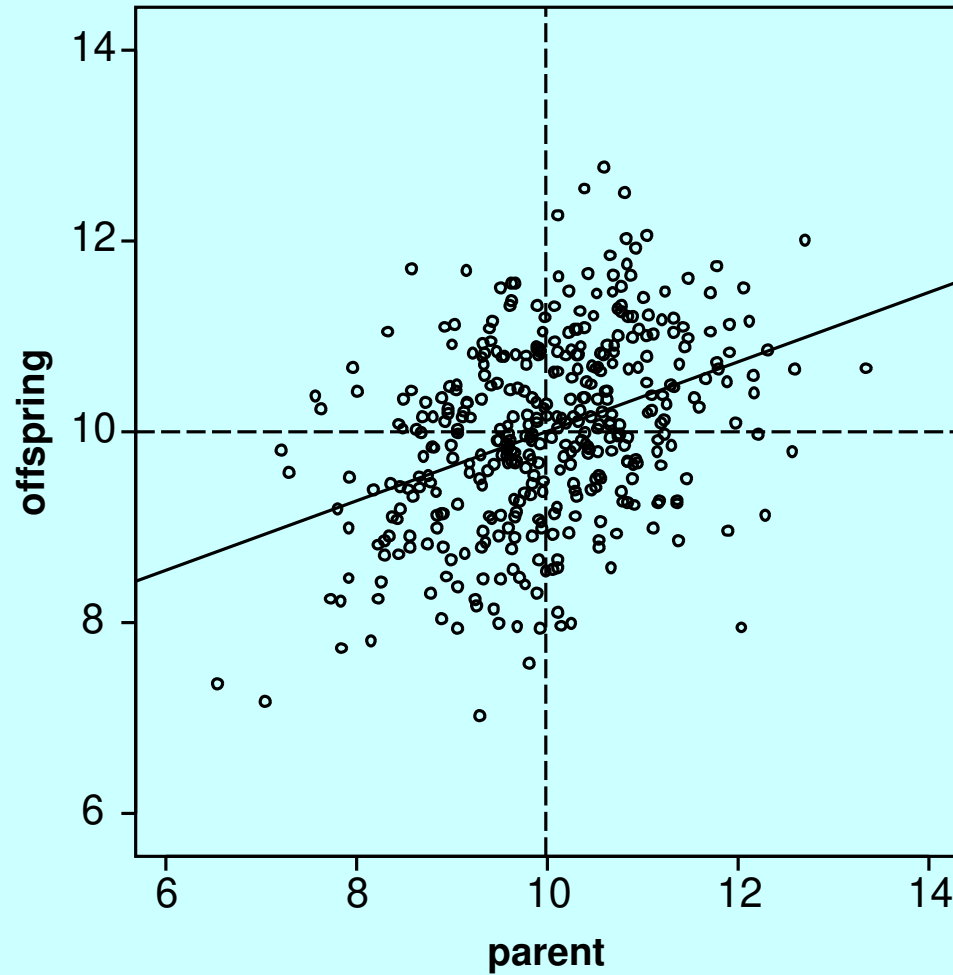
Heritability is the fraction of the total phenotypic variance that is from the additive genetic variance V_A . Note that

$$h^2 = \frac{V_A}{V_A + V_D + V_E}$$

- It is *not* the fraction of all variance that is genetic, as it does not include the dominance variance V_D
- It depends on the gene frequencies (the weights in the regression) so that it can be different from population to population
- For that matter, it can be different from generation to generation as gene frequencies change by genetic drift and selection
- Which is comforting, since when the population has fixed or lost all alleles, we'd think there would be no genetic variance at all

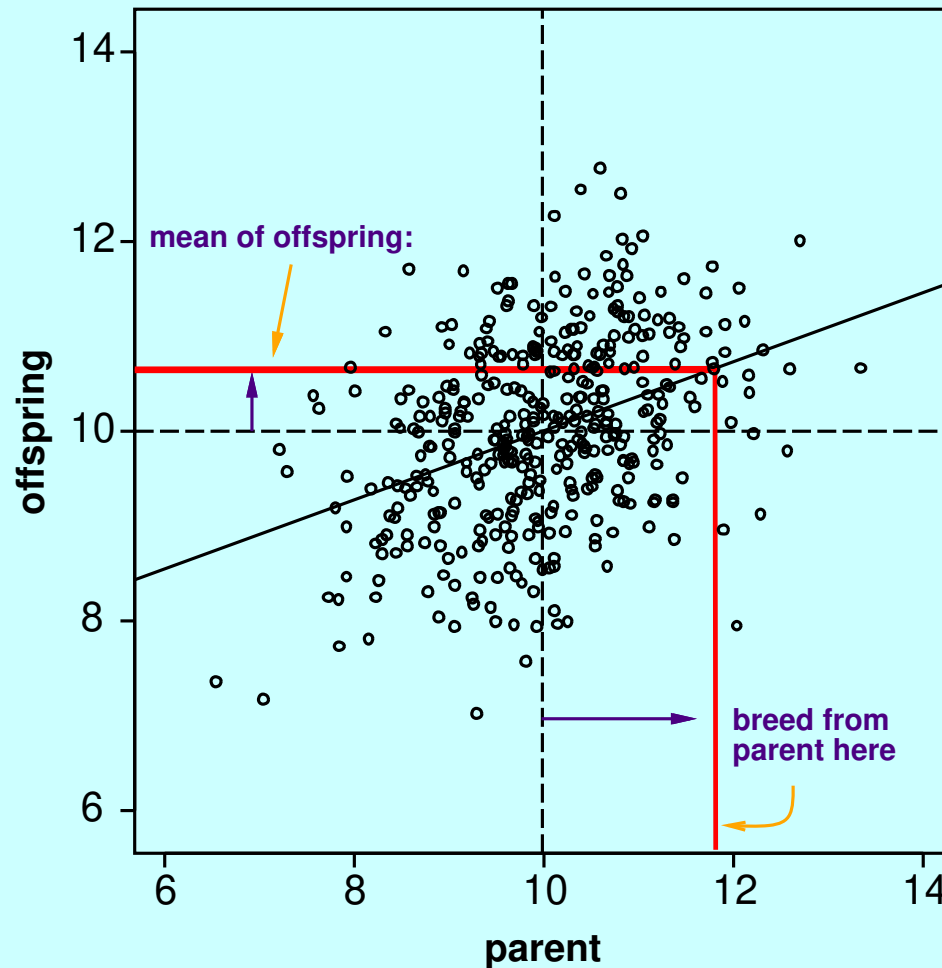
Heritability is denoted by h^2 for historical reasons (Sewall Wright in 1921) and no one ever uses unsquared h .

Slope of parent-offspring regression



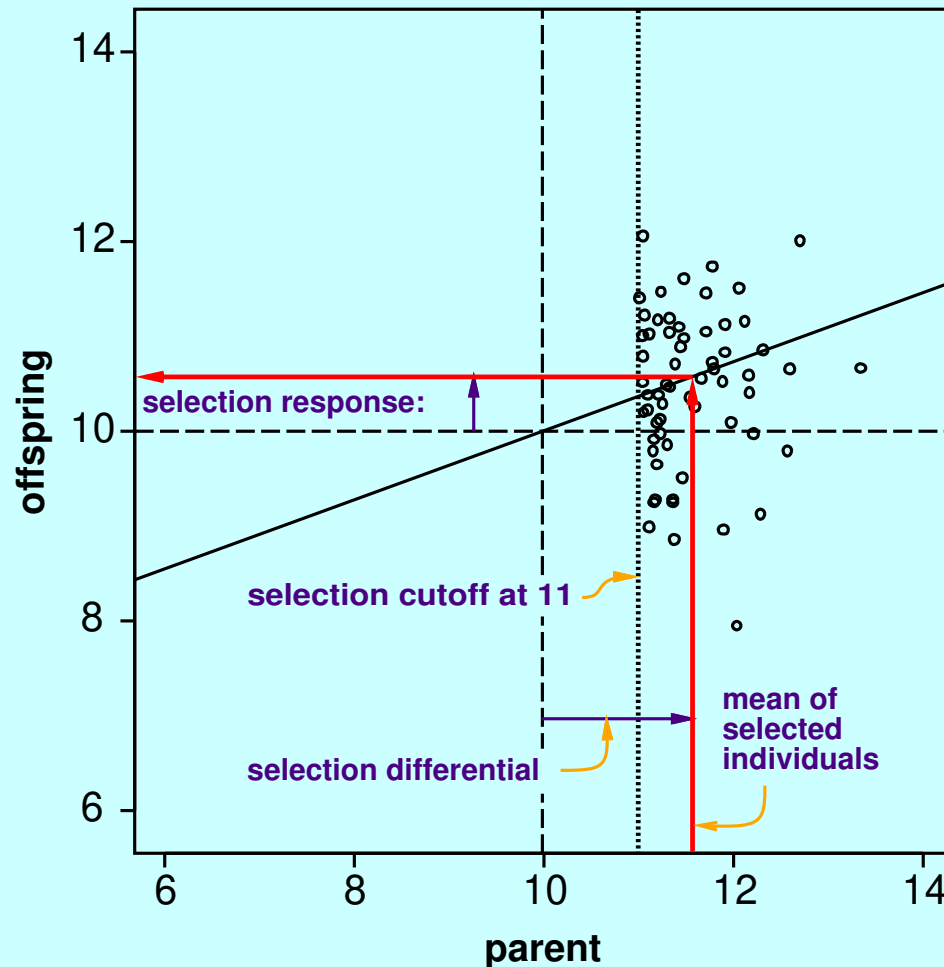
Expected slope is $\text{Cov}(O, P) / \text{Var}(P) = \frac{1}{2}h^2$

Mean of offspring of a selected parent



Expected slope is $\text{Cov}(O, P) / \text{Var}(P) = \frac{1}{2}h^2$ When lots of genes of small effect add up, get bivariate normality which implies this linear dependence.

Response to truncation selection on one parent



The bivariate normality of the joint distribution implied this linear dependence. The linearity of expectations of linear dependence guarantees this.

Response to truncation selection on both parents

This response, with response doubled for selection on both parents, is simply “The Breeder’s Equation”:

$$\text{Response} = \text{heritability} \times \text{selection differential}$$

(if selection is on both sexes of parent)

Multiple characters

All of these results can be extended to multiple characters:

- The character value P is replaced by a column vector p
- The additive, dominance, and environmental variances V_A , V_D , and V_E are replaced by covariance matrices of additive effects, of dominance deviations, and of environmental effects for all the characters A , D , and E .
- The covariances of the characters in an individual are in the phenotypic covariance matrix P
- The covariances of characters between a pair of relatives are then a covariance matrix

$$f_1 A + f_2 D$$

- The selection differential is then a vector s
- The Breeder's Equation is then the matrix equation

$$\Delta z = A P^{-1} s$$

Steve will be covering this in his lecture on Multivariate Inheritance & Response to Selection.

References on variance components and heritability

Fisher, R. A. 1918. On the correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**: 399-433. [The great founding paper of modern quantitative genetics. An impossible read.]

Lush, J. L. 1937. *Animal Breeding Plans*. Iowa State College Press, Ames, Iowa. [The book that brought quantitative genetics to animal breeding] (Also available online in its 1943 edition at <http://archive.org/details/animalbreedingpl032391mbp>)

Falconer, D. S. and T. F. C. MacKay. 1996. *An Introduction to Quantitative Genetics*. 4th Edition. Benjamin-Cummings, San Francisco. [Falconer's classic 1960 textbook, updated in recent editions by Trudy MacKay]

Lynch, M. and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Oxford University Press, Oxford. [A recent and more intensive and comprehensive text]

Walsh, B. and M. Lynch. 2018. *Evolution and Selection of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusetts. [The second volume of Lynch and Walsh]

References on variance components and heritability

Felsenstein, J. 2016. *Theoretical Evolutionary Genetics*. Available as a free download from the Department of Genome Sciences, University of Washington, Seattle at <http://evolution.gs.washington.edu/pgbook/pgbook.html> [My population genetics theory class text; chapter IX covers quantitative genetics theory, and chapter XI covers quantitative characters under optimizing selection.]