Quantitative characters II: heritability

The *variance* of a trait (x) is the average squared deviation of x from its mean:

$V_P = (1/n)\Sigma(x-m_x)^2$

This total *phenotypic variance* can be partitioned into components:

$V_P = V_G + V_E$ (genetic and environmental)

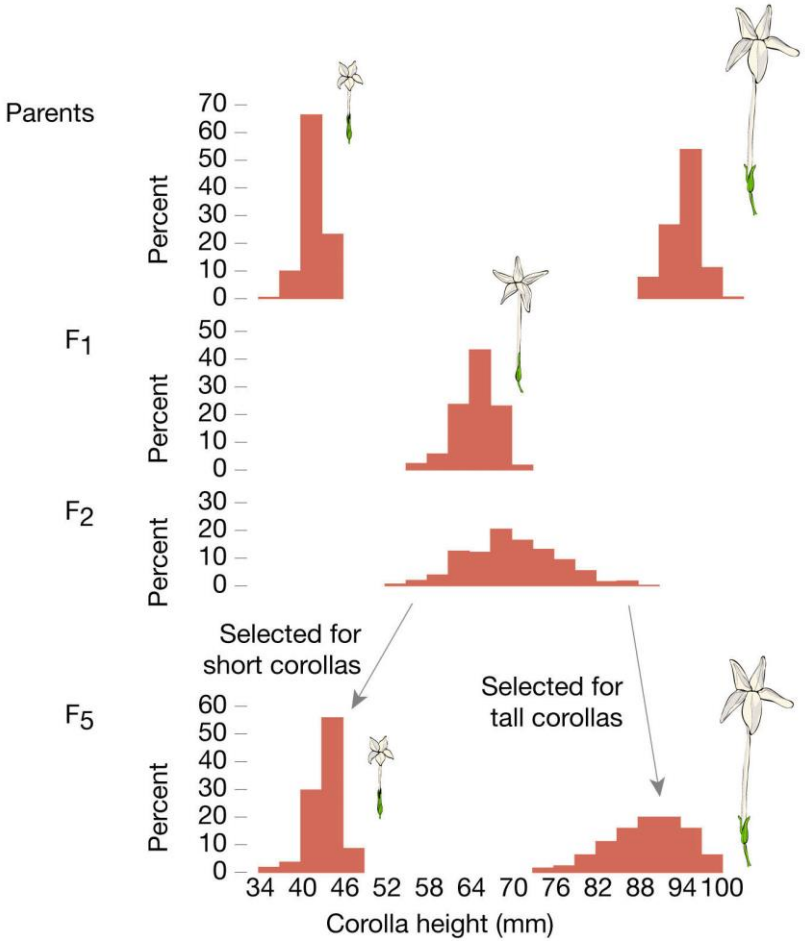
$V_G = V_A + V_D + V_I$ (additive, dom., interaction)

The broad-sense heritability is the fraction that's genetic:

$H^2 = V_G / V_P$

The narrow-sense heritability is the fraction that's *additive* genetic:

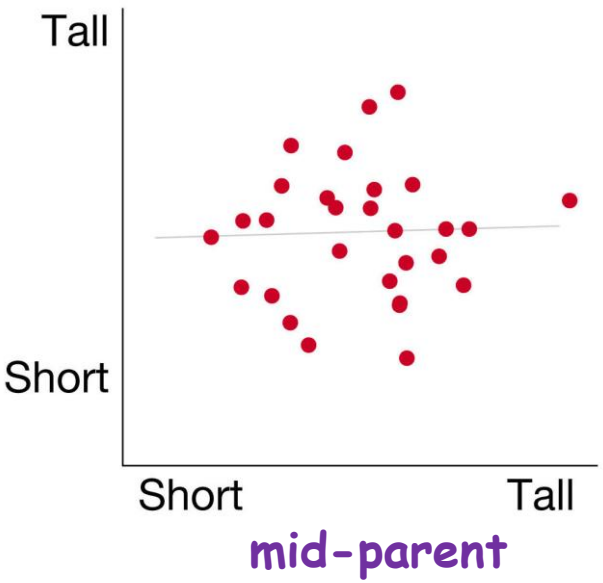
$h^2 = V_A / V_P$



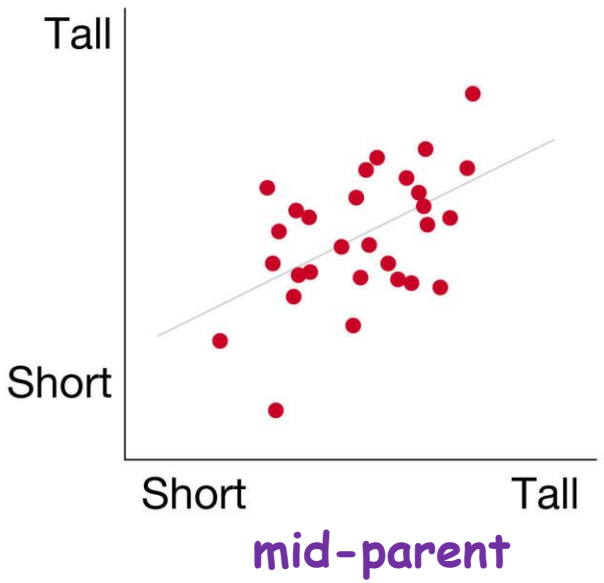
***h^2* determines (1) the *resemblance of offspring to their parents*, and (2) the *population's evolutionary response* to selection.**

h^2 is the regression (slope) of offspring on parents

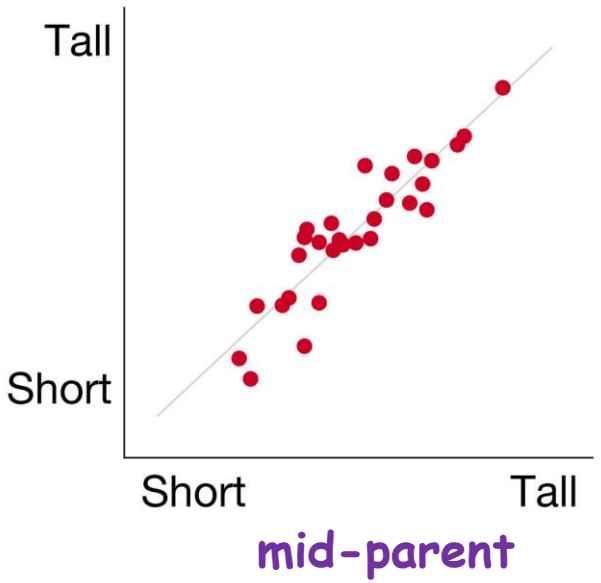
$h^2 \approx 0$



$h^2 \approx \frac{1}{2}$



$h^2 \approx 1$



Definition of the regression coefficient (slope): $b_{yx} = cov(x,y)/var(x)$

Here x is the *midparent* value (parental mean), y is the *offspring* value (see Gillespie, Table 6.2).

The higher the slope, the better offspring resemble their parents.

In other words, the higher the *heritability*, the better **offspring** trait values are *predicted* by **parental** trait values.

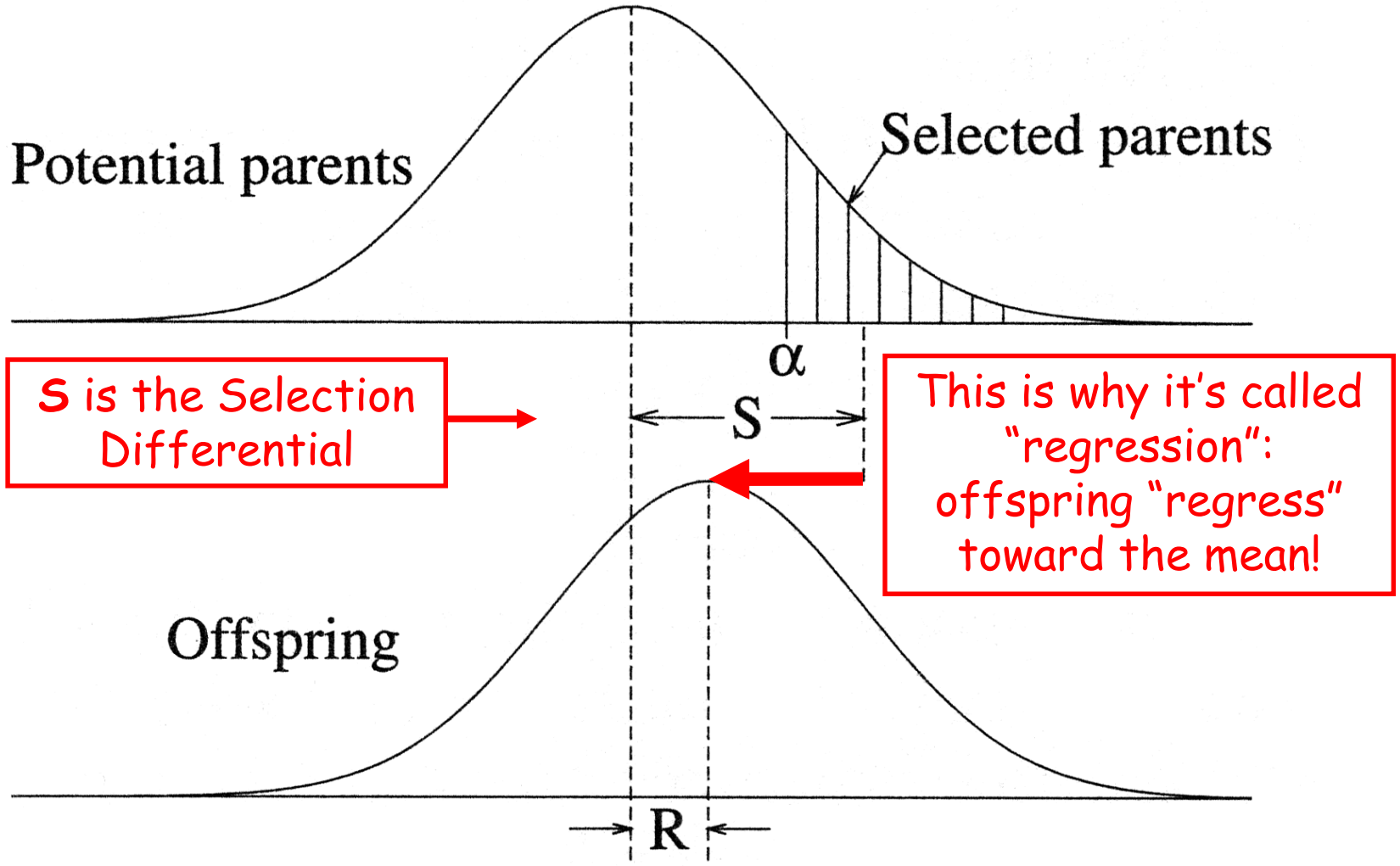
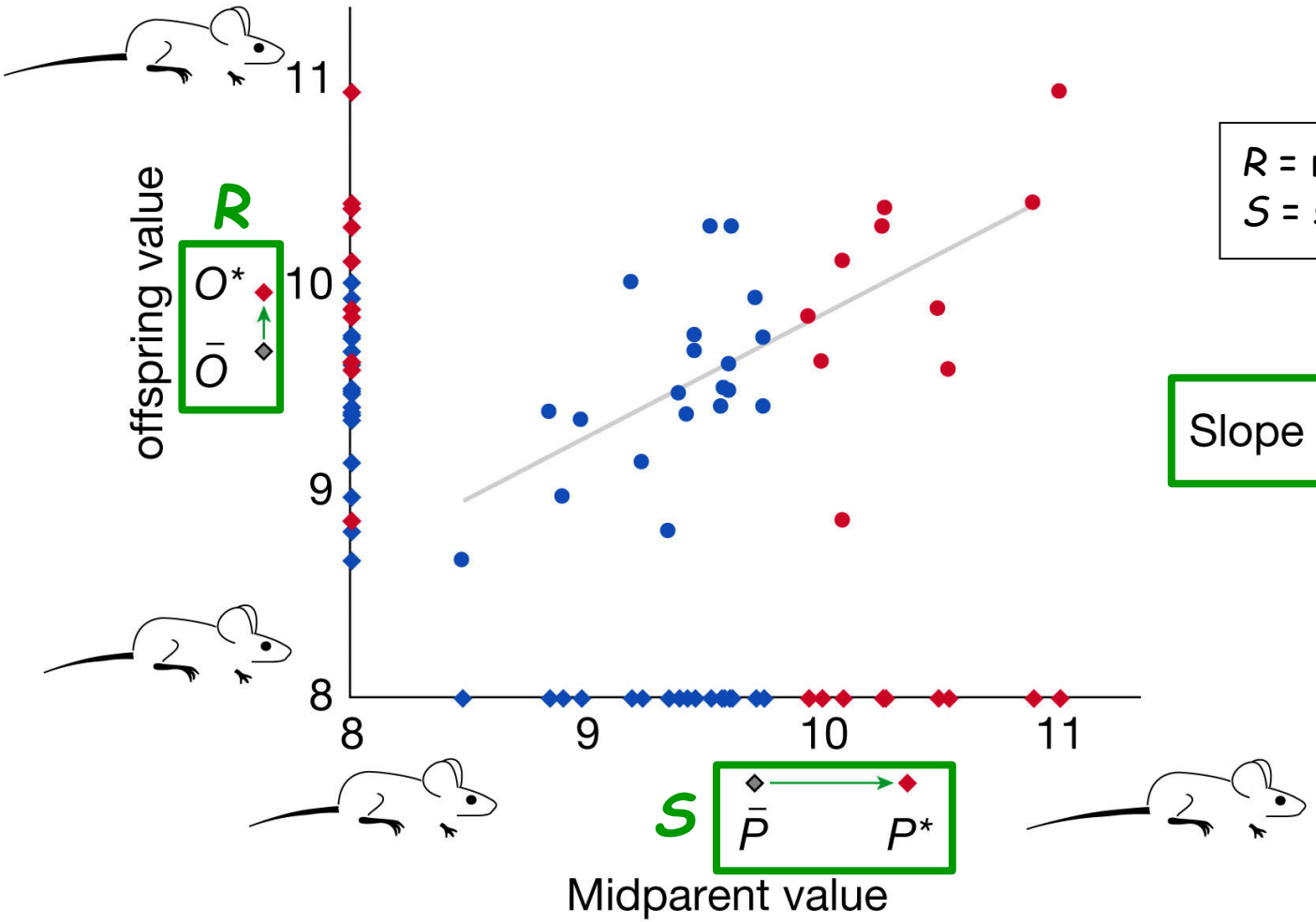


Figure 6.5: The response to selection.

The "geometric" interpretation of heritability shows why $R = h^2S$ ($h^2 = R/S$)



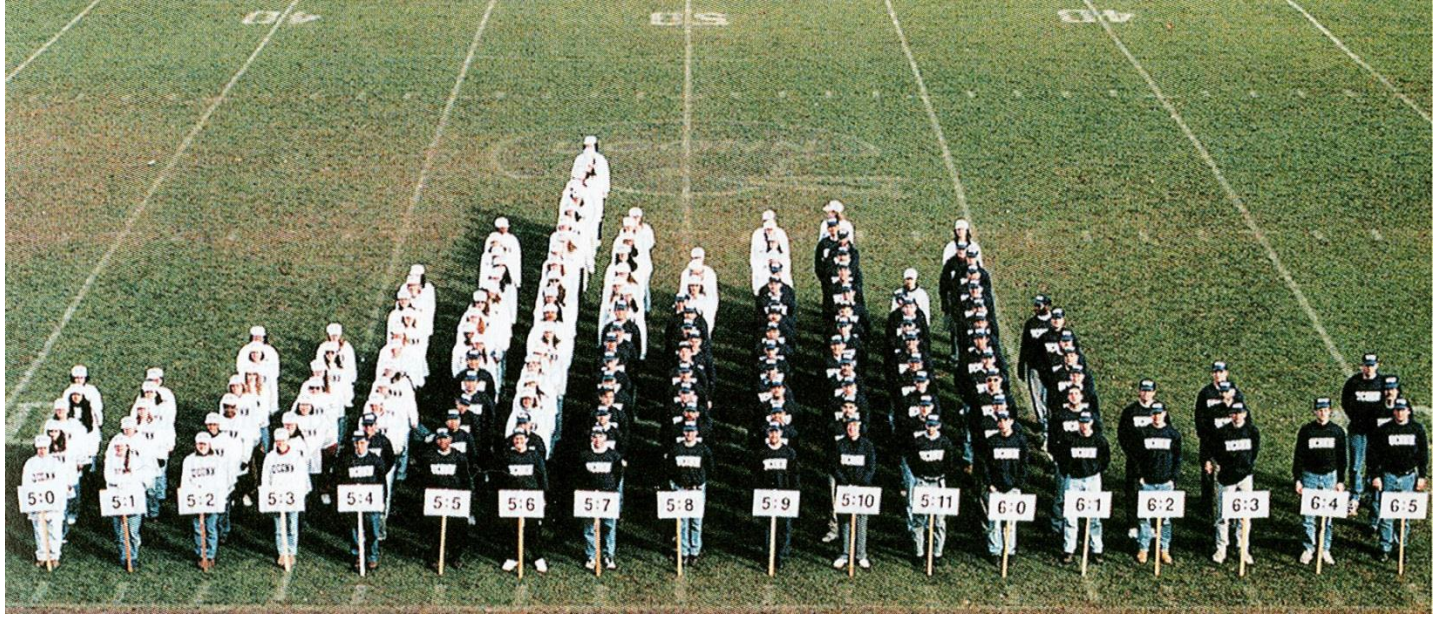
R = response to selection
S = selection differential

$$\text{Slope} = h^2 = \frac{[O^* - \bar{O}]}{[P^* - \bar{P}]} = \frac{R}{S}$$

$$R = h^2S$$

As it turns out, the additive genetic variance (V_A) is the part that makes offspring resemble their parents (*i.e.*, sets the slope).

What's the heritability of height in humans?

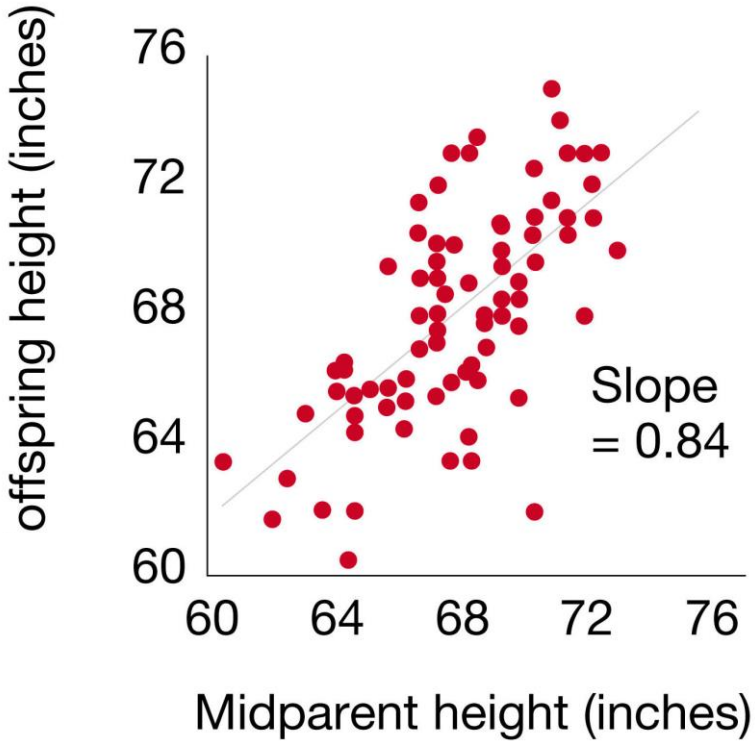


Scott Freeman and Jon Herron asked the students in their evolution course at the University of Washington to measure themselves and their parents.

Their regression plot is shown at the right.

The estimated heritability is 0.84.

That means 84% of the variance in height (V_P) is *additive genetic variance* (V_A).



The heritability of *genotypes* is 1.0 (illustrated by 30 Utah HapMap families)

One locus from each of eight xsomes.

Genotypes coded 0/1/2.

Each row is a family.

Mom's (m), dad's (f) and child's (c) genotypes are in columns.

The sum of genotype scores ("plus" alleles) is shown in the last set of columns.

	chr 1			chr 2			chr 7			chr 8			chr19			chr20			chr21			chr22			total		
fam	m	f	c	m	f	c	m	f	c	m	f	c	m	f	c	m	f	c	m	f	c	m	f	c	m	f	c
0:	1	1	2	0	1	0	2	2	2	1	1	0	0	0	0	0	1	0	2	1	1	1	2	1	7	9	6
1:	2	2	2	1	0	1	2	1	2	2	1	2	1	0	1	0	0	0	2	2	2	2	1	1	12	7	11
2:	2	1	2	1	0	1	1	2	2	1	2	1	1	1	1	1	1	2	1	0	0	2	2	2	10	9	11
3:	2	1	2	0	0	0	2	2	2	2	2	2	0	1	0	1	0	0	0	2	1	1	2	1	8	10	8
4:	1	1	2	1	1	0	2	1	1	1	2	1	1	1	2	0	0	0	2	1	2	2	1	1	10	8	9
5:	1	2	1	0	1	1	1	0	1	2	2	2	2	2	2	1	0	0	2	1	1	1	2	2	10	10	10
6:	1	2	1	0	0	0	1	2	2	2	2	2	1	0	0	0	0	0	1	1	1	2	1	2	8	8	8
7:	2	2	2	2	0	1	1	2	1	0	2	1	0	1	1	0	1	1	1	1	0	0	2	1	6	11	8
8:	1	1	1	0	0	0	2	2	2	2	2	2	0	1	1	1	1	2	1	2	1	2	2	2	9	11	11
9:	0	2	1	1	1	1	1	2	2	2	0	1	2	1	2	1	0	1	1	2	1	2	2	2	10	10	11
10:	2	2	2	1	1	1	2	1	1	1	2	1	0	0	0	1	1	1	2	2	2	2	2	2	11	11	10
11:	1	0	1	0	0	0	1	2	1	1	2	2	1	0	1	1	1	2	2	1	2	1	1	0	8	7	9
12:	1	1	0	0	0	0	2	2	2	2	1	2	1	0	1	0	1	0	1	2	1	2	1	1	9	8	7
13:	1	2	2	0	1	0	2	1	2	2	1	2	2	1	2	0	1	0	1	2	2	2	1	1	10	10	11
14:	1	1	2	2	1	2	2	2	2	2	2	2	0	1	1	0	0	0	1	1	0	2	0	1	10	8	10
15:	1	1	2	0	0	0	2	2	2	2	2	2	1	0	1	2	1	1	0	2	1	1	1	1	9	9	10
16:	0	1	1	0	1	0	2	1	1	1	1	2	1	1	1	1	0	1	2	2	2	2	2	2	9	9	10
17:	2	1	1	0	0	0	1	1	1	1	1	2	1	0	1	1	0	0	1	1	1	2	0	1	9	4	7
18:	1	2	2	1	1	1	1	2	2	2	0	1	1	1	1	0	0	0	2	2	2	1	2	1	9	10	10
19:	2	1	1	0	0	0	2	0	1	1	2	2	2	1	2	2	0	1	2	1	2	2	2	2	13	7	11
20:	2	2	2	0	1	0	0	1	1	1	2	1	1	1	2	1	0	1	1	1	1	2	1	2	8	9	10
21:	2	0	1	1	0	1	2	1	1	2	1	2	1	0	1	1	1	1	1	1	2	1	1	2	11	5	11
22:	2	1	1	0	1	1	2	2	2	1	1	1	0	0	0	0	1	1	2	1	1	2	2	2	9	9	9
23:	1	1	1	0	1	0	0	2	1	1	2	1	1	0	0	0	1	1	1	0	1	1	1	1	5	8	6
24:	1	1	1	1	0	1	2	2	2	1	2	2	0	0	0	0	0	0	1	2	2	2	2	2	8	9	10
25:	2	1	1	1	0	0	1	2	2	2	2	2	1	0	1	1	2	2	1	0	1	1	2	1	10	9	10
26:	2	1	2	0	0	0	1	2	1	2	1	1	0	1	0	0	0	0	1	2	1	2	1	2	8	8	7
27:	2	1	1	0	1	0	2	2	2	2	2	2	0	0	0	1	1	0	1	0	1	2	2	2	10	9	8
28:	1	2	1	1	1	2	2	2	2	0	2	1	1	0	0	1	2	1	1	1	1	0	2	1	7	12	9
29:	0	2	1	1	1	2	1	2	2	1	2	1	1	0	1	1	0	1	0	2	1	2	1	2	7	10	11

The average regression is close to 1, as is that of the sum over loci.

Offspring genotypes (0, 1, 2)

```
Chr 1 (A/C at 53433581)
2.0:            4   5   4
1.0:        2   6   8
0.0:            1
        0  0.5  1  1.5  2
r =  0.409   b =  0.569
```

```
Chr 2 (G/T at 224346716)
2.0:            2   1
1.0:        6   4
0.0:    9   7   1
        0  0.5  1  1.5  2
r =  0.731   b =  1.200
```

```
Chr 7 (A/T at 92947019)
2.0:                8  10
1.0:        2   3   7
0.0:
        0  0.5  1  1.5  2
r =  0.672   b =  0.792
```

```
Chr 8 (A/G at 122870354)
2.0:            2   7   8
1.0:            5   7
0.0:            1
        0  0.5  1  1.5  2
r =  0.569   b =  0.875
```

```
Chr 19 (A/G at 48986689)
2.0:            2   3   1
1.0:       11   3
0.0:    5   5
        0  0.5  1  1.5  2
r =  0.822   b =  1.216
```

```
Chr 20 (A/G at 48392908)
2.0:            3   1
1.0:        7   3   2
0.0:    7   6   1
        0  0.5  1  1.5  2
r =  0.682   b =  1.048
```

Midparent genotypes (0, 0.5, 1, 1.5, 2)

The regressions (b) are greater than the correlations (r) because $var(x)$ (*i.e.*, of midparent values) < $var(y)$ (*i.e.*, of offspring values). See Gillespie, table 6.2.

Chr 21 (C/T at 18426726)

```
2.0:           1   5   4
1.0:       3   7   7
0.0:       1   2
       0  0.5  1  1.5  2
```

r = 0.629 b = 0.875

Chr 22 (A/G at 16361045)

```
2.0:           1   5   9
1.0:           6   8
0.0:           1
       0  0.5  1  1.5  2
```

r = 0.673 b = 1.006

Sum of genotypic values for all 8 loci

```
11.0:               1   1       2   4
10.0:                   2   3   2   1       1
 9.0:           1           2   1
 8.0:               1   1   1   1
 7.0:       1       1   1
 6.0:       1       1
        6       7       8       9      10      11
```

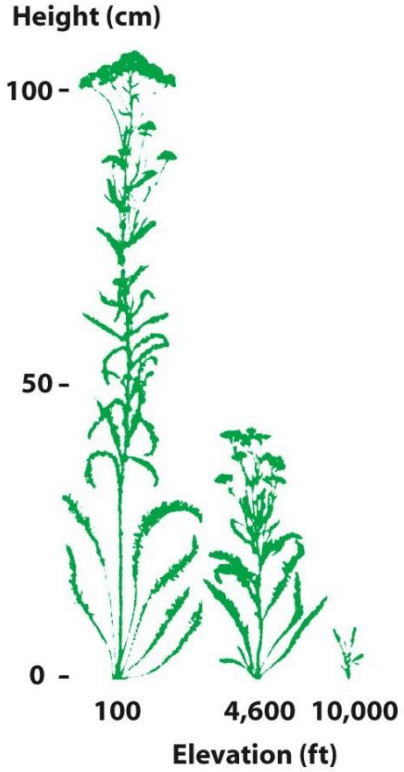
r = 0.651 b = 1.007

Expected: $b = h^2 = 1$

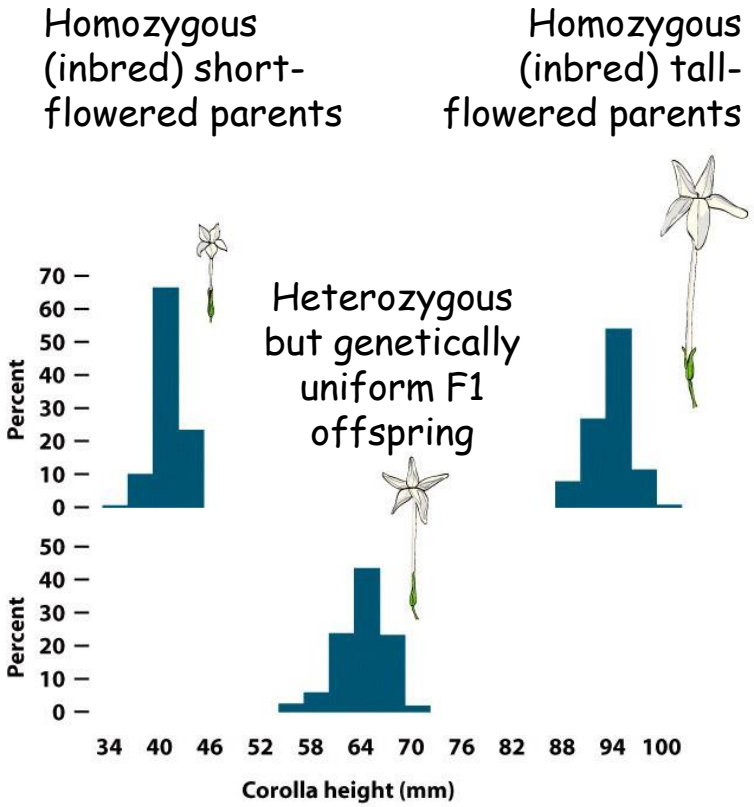
$r = h^2/\mathrm{sqrt}(2) = 1/1.414 = 0.71$

What about the variation induced by *environmental* factors?

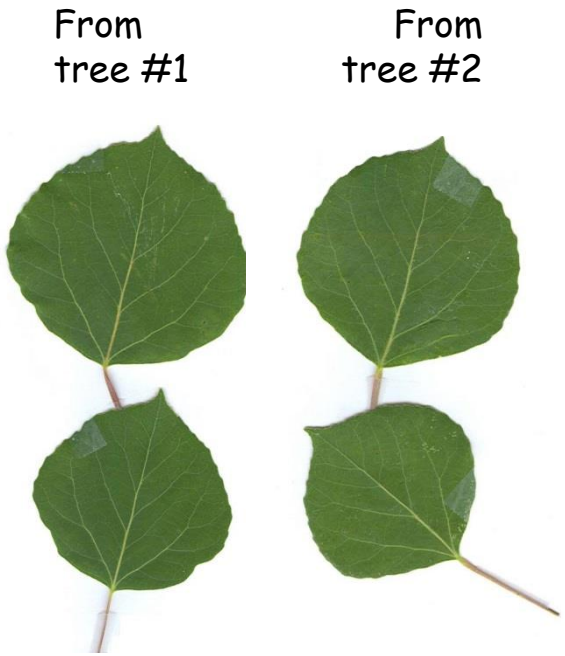
After all, even *clones* and *identical twins* differ from each other!



Clones (cuttings) of *Achillea* grown at three different elevations where the species normally occurs in California.



Edward East's *Nicotiana* plants growing in the *same* garden plots.



Leaves from a natural clone of quaking aspen (*Populus tremuloides*) growing at the top of Millcreek Canyon.

Total phenotypic variance = genetic variance + environmental variance

What you *see* is what you *get* from two distinct sources that can be separated.

1. ***Genetic variance*** is the variance among ***phenotypes*** caused by *genotypic* differences among individuals (holding their *environments* constant).
2. ***Environmental variance*** is the variance among ***phenotypes*** caused by differences in the *experiences* of individuals (holding *genotypes* constant).

Example: Suppose the *average* trait values of ***AA***, ***Aa*** and ***aa*** individuals are **-1**, **0**, and **+1** units, and $p = q = 0.5$.

Then the ***genetic variance*** (average squared deviation from the population mean) is **0.5**.

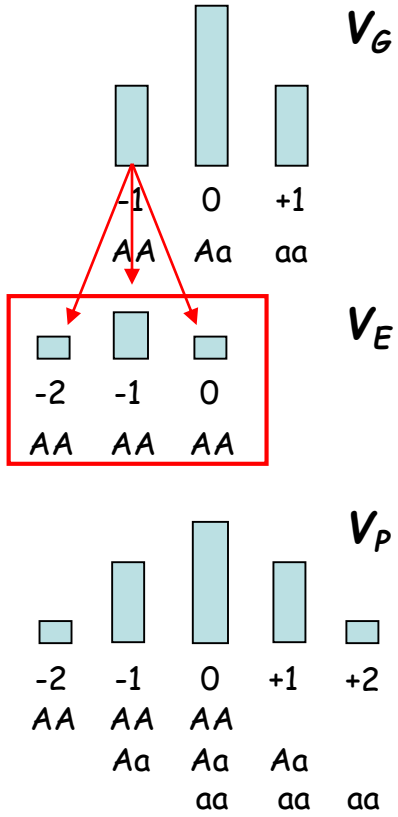
But suppose 25% of each genotype deviates one unit above or below its average trait value, because of the environment.

Then the ***environmental variance*** is also **0.5**.

The resulting ***phenotypic variance*** is **0.5** + **0.5** = **1.0**.

$$V_P = V_G + V_E = (V_A + V_D + V_I) + (V_E)$$

The trait's ***heritability*** is the ***fraction*** of V_P that is ***genetic*** (actually, ***additive genetic***, as we will see).



Not all genetic variance is additive! (A silly but instructive model.)

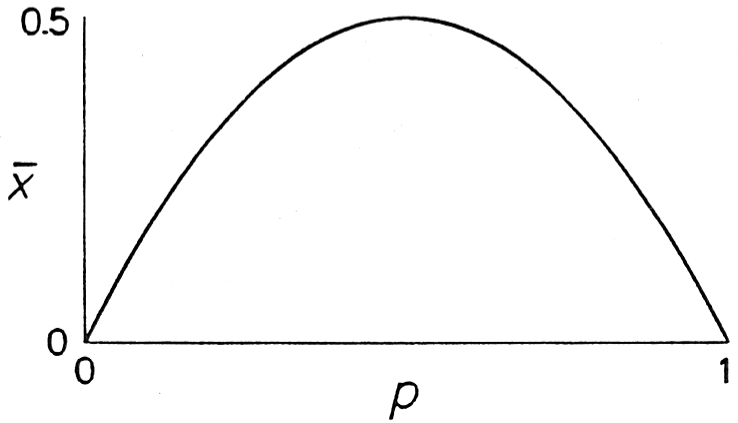
Consider a simple quantitative trait (x) controlled in a symmetrically overdominant manner by two alleles at one locus.

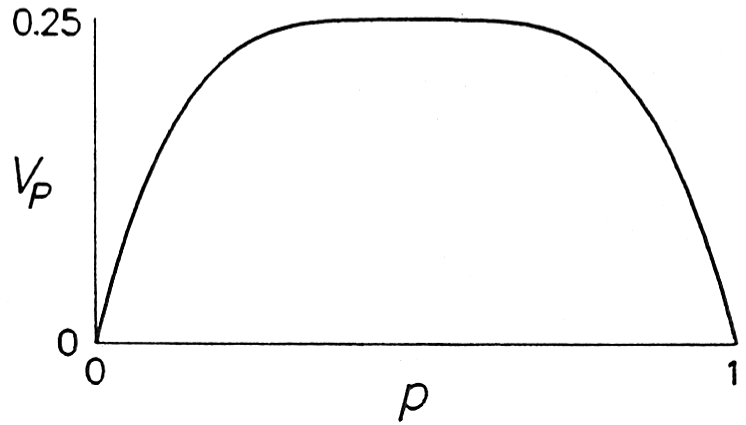
Assume that there is *no environmental variance*.

genotype	phenotype (x)	frequency
AA	0	p^2
Aa	1	$2p(1\text{-}p)$
aa	0	$(1\text{-}p)^2$

This model is fully developed in two handouts on the course web site, under "Readings".

Thus $\bar{x} = 2p(1\text{-}p)$, and $\mathrm{Var}(x) = V_P = V_G = E(x^2)\text{-}[E(x)]^2 = 2p(1\text{-}p)[1\text{-}2p(1\text{-}p)]$.



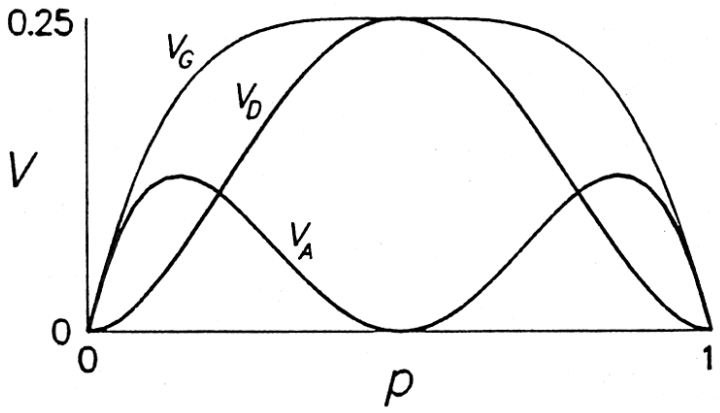


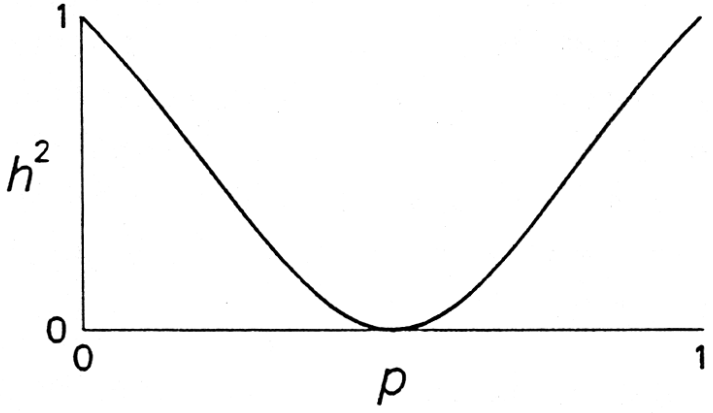
What happens if we *select* for *higher* values of *x*?

The heritability disappears when the genetic variance is greatest!

Now we can calculate the heritability, which is

$$h^2 = V_A/V_P = (1\text{-}2p)^2/[1\text{-}2p(1\text{-}p)] \ .$$

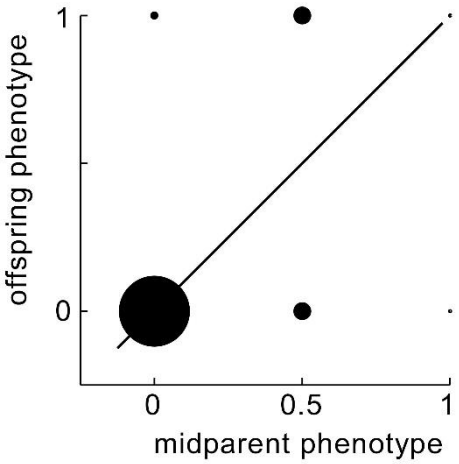




$p \approx 0$ or 1

Not much variance,

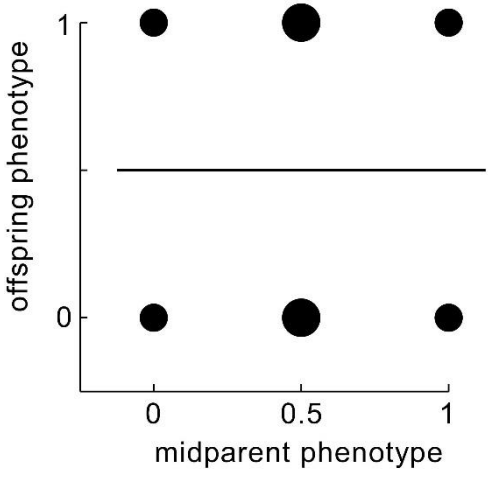
but $h^2 \approx 1$



$p = 0.5$

Lots of variance,

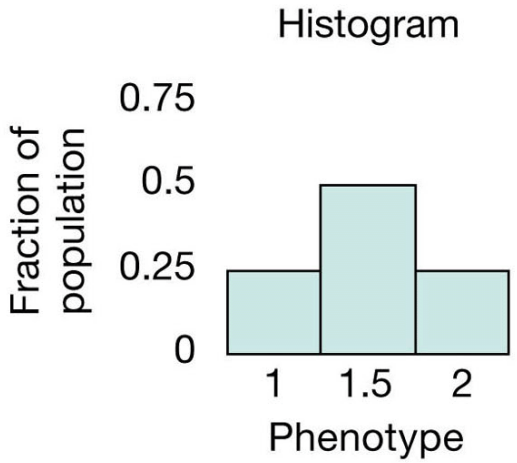
but $h^2 = 0$!

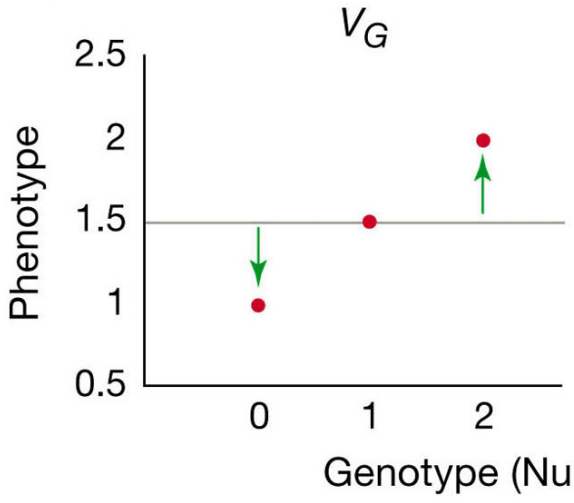


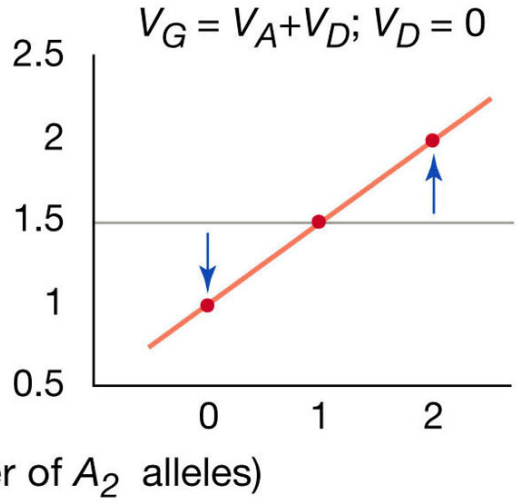
At $p = 0.5$, all of the genetic variance is *dominance variance*, not additive variance.

Dominance variance arises from non-additive relationships between the "dosage" of an allele (number carried) and the resulting phenotype

No dominance. Phenotypes: $A_1A_1 = 1$; $A_1A_2 = 1.5$; $A_2A_2 = 2$

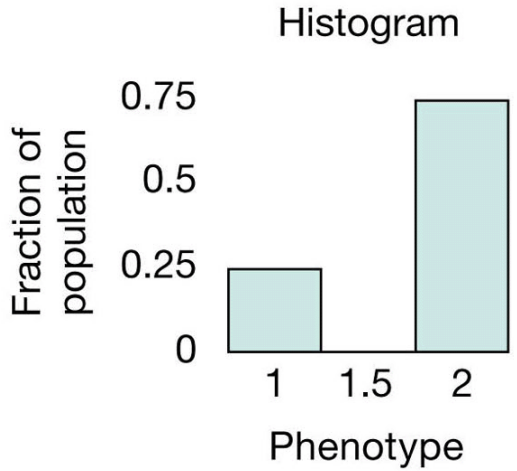


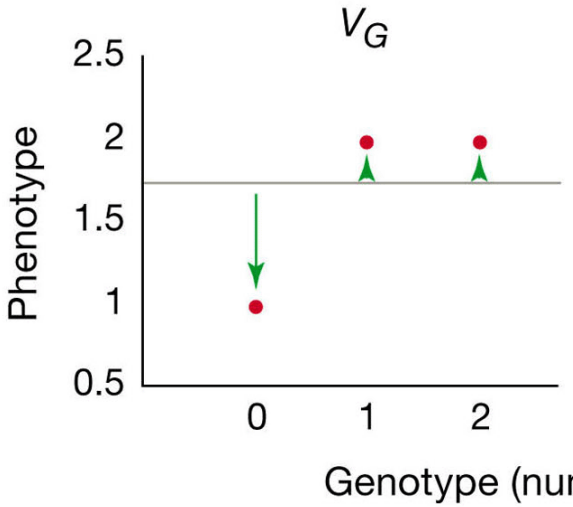


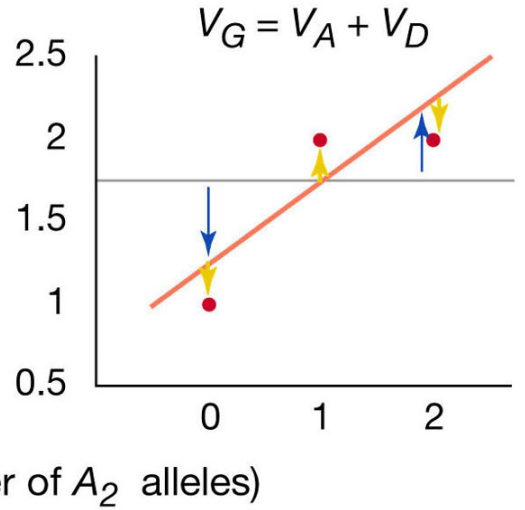


Genotype (Number of A_2 alleles)

Complete dominance. Phenotypes: $A_1A_1 = 1$; $A_1A_2 = 2$; $A_2A_2 = 2$







Genotype (number of A_2 alleles)

How to estimate the components of the phenotypic variance (V_P)

1. Measure phenotypes (trait values) in a large random sample of the population.

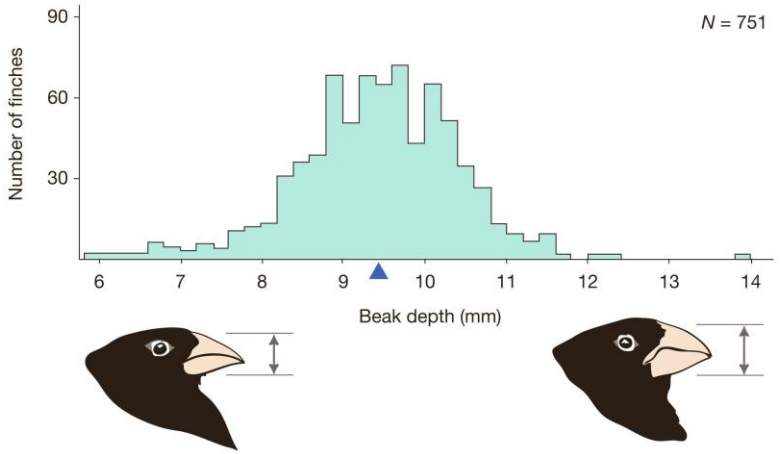
2. Calculate the mean and variance: the variance is V_P.

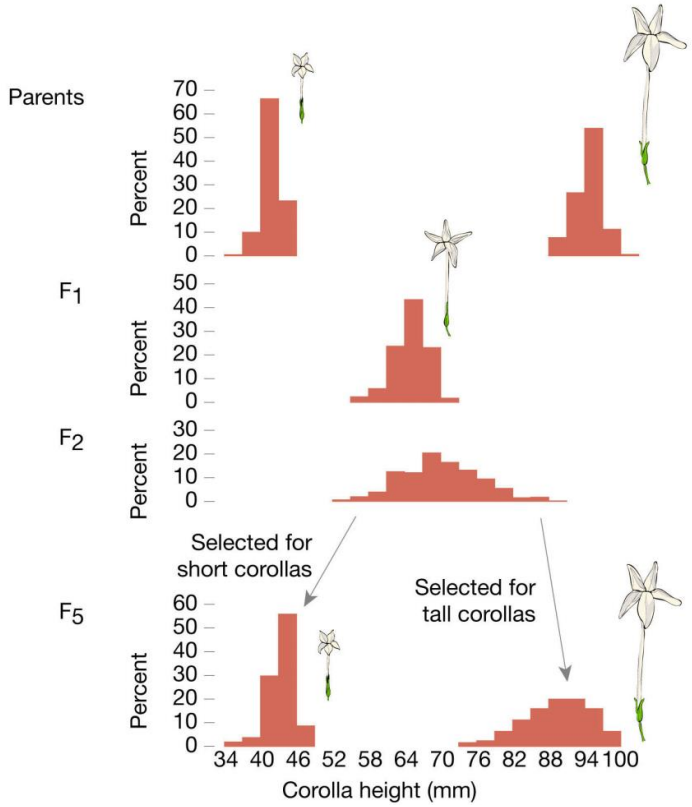
3. Estimate the heritability, either of two ways:
   (a) regress offspring on midparent values
   (b) measure the response to selection:
   $h^2 = R/S$

4. The additive variance (V_A) is the heritable fraction of the total: $V_A = h^2 V_P$.

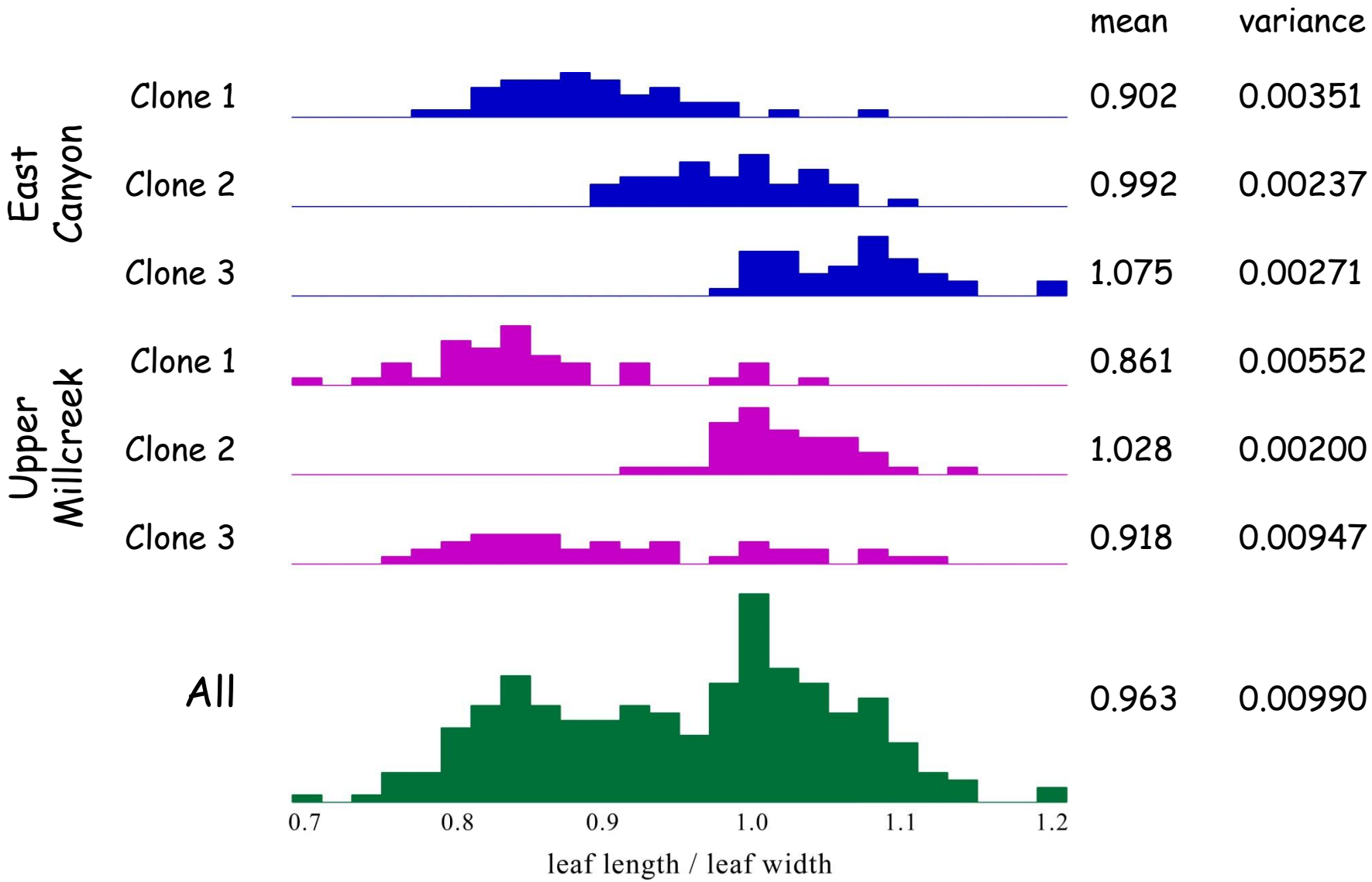
5. The remainder is environmental (V_E) and dominance variance (and other minor stuff).

6. If we can clone or closely inbreed members of the species, or find identical twins, then we can *directly* estimate the environmental variance.





Leaf shape within and among six quaking aspen clones



Analysis of variance (ANOVA)

		mean	variance
East Canyon	Clone 1	0.902	0.00351
	Clone 2	0.992	0.00237
	Clone 3	1.075	0.00271
Upper Millcreek	Clone 1	0.861	0.00552
	Clone 2	1.028	0.00200
	Clone 3	0.918	0.00947
All		0.963	0.00990

Variance **among** clones

= var(0.902, 0.992, … , 0.918)

= **0.00564**

Variance **within** clones

= mean(0.00351, … , 0.00947)

= **0.00426**

Total variance

= 0.00564 + 0.00426

= **0.00990**

Fraction **explained by** clones

= 0.00564 / 0.00990

= **0.57 = H^2**

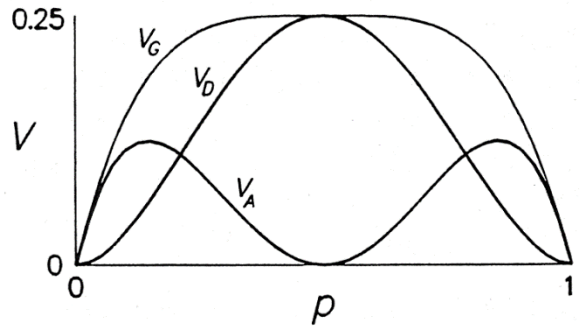
7. The dominance variance can be separated from the additive variance by exploiting the different ways these components appear in the covariances between different kinds of relatives. For example:

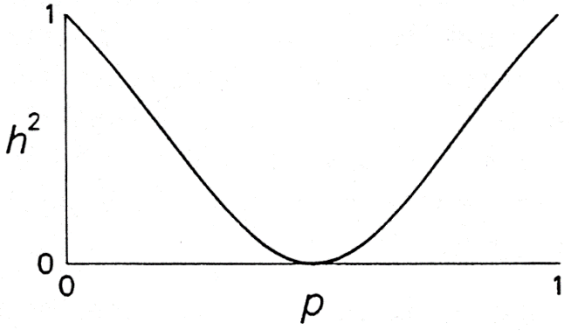
cov(parent-offspring)	$\approx \frac{1}{2} V_A$
cov(half sibs)	$\approx \frac{1}{4} V_A$
cov(full sibs)	$\approx \frac{1}{2} V_A + \frac{1}{4} V_D$

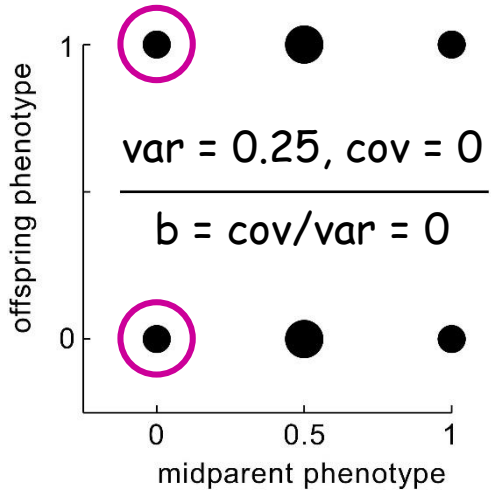
See the Quantitative Traits lecture notes, and Gillespie, for more on this.

And it works in the "pure overdominance" model developed above…

No heritability, but a *positive correlation between sibs*, in the pure-overdominance model with $p = 0.5$.



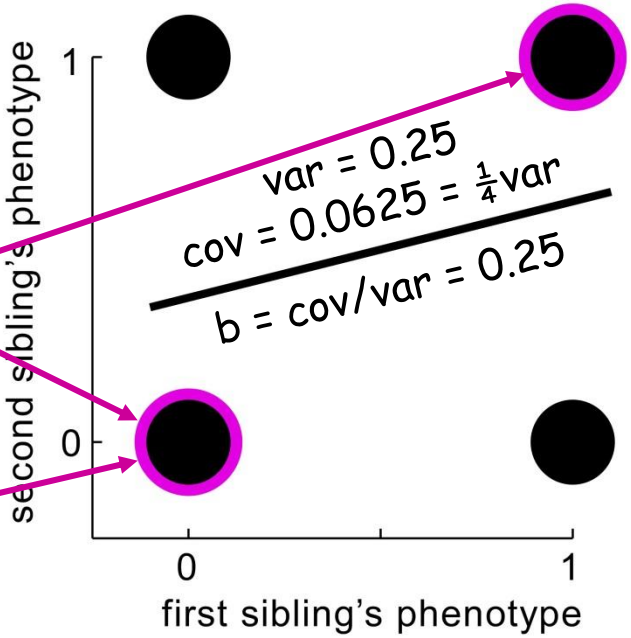




Parents both 0

Offspring Phenotypes

	parents	Freq	OP=0	OP=1
[0,0]	AA x AA	0.0625	1	
	AA x Aa	0.25	0.5	0.5
[0,0]	AA x aa	0.125		1
	Aa x Aa	0.25	0.5	0.5
	Aa x aa	0.25	0.5	0.5
[0,0]	aa x aa	0.0625	1	



7. The dominance variance can be separated from the additive variance by exploiting the different ways these components appear in the covariances between different kinds of relatives. For example:

cov(parent-offspring) $\approx \frac{1}{2} V_A$
cov(half sibs) $\approx \frac{1}{4} V_A$
cov(full sibs) $\approx \frac{1}{2} V_A + \frac{1}{4} V_D$ (see QT lecture notes)

An interesting and general finding is that traits closely related to *fitness* tend to have *little* additive variance but *more* dominance and interaction variance (epistasis) than typical morphological traits.

What might be the explanation?

Partitioning of the variance of four characters in *Drosophila melanogaster*. Components as percentages of the total, phenotypic, variance.

		Character (1) *Bristles*	(2) *Thorax*	(3) *Ovary*	(4) *Eggs*
Phenotypic	V_P	100	100	100	100
Additive genetic	V_A	52	43	30	18
Non-additive genetic	$V_D + V_I$	9	6	40	44
Environmental	V_E	39	51	30	38

Characters:

(1) Number of bristles on 4th + 5th abdominal segments (Clayton, Morris, and Robertson, 1957; Reeve and Robertson, 1954).
(2) Length of thorax (F. W. Robertson, 1957*b*).
(3) Size of ovaries, i.e. number of ovarioles in both ovaries. (F. W. Robertson, 1957*a*).
(4) Number of eggs laid in 4 days (4th to 8th after emergence) (F. W. Robertson, 1957*b*).

Summary

The narrow-sense *heritability* of a trait is the fraction of the total phenotypic variance that is caused by the *additive effects* of genes.

There can be considerable *non-additive genetic variance*, but this does not contribute to the resemblance between parents and offspring, or the response to selection.

(But the *dominance variance* increases the resemblance of full siblings.)

There can also be a lot of "environmental variance" (that is, variance of the *trait values* that is caused by *effects* of the environment).

These three components of the phenotypic variance literally add up to the total: $V_P = V_A + (V_D + V_I) + V_E$

The *analysis of variance* (ANOVA) was originally invented to allow these components to be estimated separately.