Walsh Lecture 3
Univariate and Multivariate Selection

Bruce Walsh. 25 Jan - 5 Feb 2016

Second Bangalore School on Population Genetics and Evolution

Quantitative geneticists usually distinguish between the short- and long-term response to selection. If we are only trying to predict a few generations of selection response, knowledge of the base population genetic variances (and in particular the heritability) is usually sufficient to make a satisfactory prediction. However, as selection proceeds and allele frequencies change significantly, the initial genetic variances essentially lack any predictive power for the long-term response. The focus in this lecture is on just what predictions we can make for short-term response .

I: Univariate Selection: the Breeder's Equation and its Extensions

It is critical to distinguish between the **within-** and **between-generation** changes induced by selection. The within-generation change is the difference in a population before and after an episode of selection, while the between-generation change (the **response to selection**) is the difference between the population distribution before selection and the distribution of the trait in the next generation (measured at the suitable stage). The response to selection depends not only on the strength of within-generation selection, but also on the fraction of offspring trait value that can be predicted from parental value. If the latter is zero, no matter how strong the within-generation selection is, there will be no response to selection.

The Selection Differential S and Response R

The within-generation change in the mean due to selection is

$$S = \mu_* - \mu \tag{3.1}$$

where μ is the population mean before selection and μ_* the mean of the parents that reproduce (the population mean after selection). S is called the **selection differential**.

The between-generation change, (the response to selection) R, is the change in means between the population before selection and the population in the next generation,

$$R = \mu_o - \mu \tag{3.2}$$

where μ_o is the character mean in the offspring (measured at the same stage as in their parents).

Another useful way to think about the response is in terms of *breeding values*, as the average deviation of offspring from the population mean is just the ***mean breeding value of their parents***. Hence, the ***response can simply thought of as the net change in breeding value among of the selected parents***.

The Selection Intensity i

Much akin to the covariance being a poor indicator of the strength of an association, the selection differential S is not particularly informative when trying to compare the strength of selection on different traits and/or in different populations. A much more useful measure is the **selection intensity** i,

$$i = \frac{S}{\sigma_z} \tag{3.3}$$

which is the differential expressed in phenotypic standard deviations.

The Breeder's Equation: Translating S into R

The parent-offspring regression allows us to translate the within-generation change S into the between-generation change R. Recall (Lecture 1) that the predicted value $\widehat{y}$ given we know x is

$$\widehat{y} = \mu_y + b_{y\,|\,x}(\,x - \mu_x\,)$$

Here we are trying to predict the offspring value y_O given $x = (z_f + z_m)/2$, the midparent value. Hence, $b_{y\,|\,x} = b_{O\,|\,MP} = h^2$ is the slope of the midparent-offspring regression, while $\mu_y = \mu_x = \mu$, the mean trait value in the population, giving

$$y_O = \mu + h^2 \left(\frac{z_f + z_m}{2} - \mu \right) \tag{3.4}$$

This regression holds for each midparent-offspring pair. Averaging over all parents, the average difference between the mean (μ_*) of selected parents and the (before selection) population mean is

$$E[\,(z_f + z_m)/2 - \mu] = \mu_* - \mu = S$$

Likewise, the average value over all the offspring of these selected parents is $E[\,y_O\,] = \mu_O$. Thus, averaging over all the midparents gives

$$\mu_O = \mu + h^2\, S$$

since $R = \mu_O - \mu$, we have

$$R = h^2 S \tag{3.5}$$

This relationship is often called the **breeder's equation**, and shows that the heritability of a character is the link between the within-generation change S and the between-generation response R. If $h^2 \simeq 0$, then $R \simeq 0$ no matter how strong much selection is applied.

In some situations, males and females are subjected to different amounts of selection. The regression of offspring value on the value of its sire and dam can also be written as

$$\mu_o = \mu_z + \frac{h^2}{2}\,(z_s - \mu_s) + \frac{h^2}{2}\,(z_d - \mu_d) + e$$

giving the expected response as

$$R = \frac{h^2}{2}\,S_s + \frac{h^2}{2}\,S_d \tag{3.6}$$

In this case, the breeders equation holds with the selection differential is simply the average differential of both sexes,

$$S = \frac{S_s + S_d}{2}$$

There are several equivalent expressions for the breeder's equation. First,

$$R = \frac{\sigma_A^2}{\sigma_z^2} S = \sigma_A \frac{\sigma_A}{\sigma_z} \frac{S}{\sigma_z} = \sigma_A h i \tag{3.7}$$

Alternatively,

$$R = h^2 S \frac{\sigma_z}{\sigma_z} = h^2 \sigma_z i \tag{3.8}$$

While the breeder's equation holds for a single generation of selection from an unselected base population, its validity in predicting response over several generations depends on:

- The reliability of the h^2 estimate
- Absence of environmental change between generations
- The absence of genetic change between the generation in which h^2 was estimated and the generation in which selection is applied.

The later point is critical, as strictly speaking, the prediction equation is true for one generation only, since selection changes gene frequencies and thus h^2 (through changes in the genetic variances). In practice, the breeder's equation is generally valid over several generations.

The Generalized Breeder's Equation: Accuracy

We can extent the breeder's equation to apply to much more general selection schemes beyond simply choosing an individual solely on the basis of its phenotype. To obtain this extension, first note that

$$h^2\sigma_z = \left(\frac{\sigma_A^2}{\sigma_z^2}\right)\sigma_z = \left(\frac{\sigma_A}{\sigma_z}\right)\sigma_A = h\sigma_A$$

Hence, we can rewrite Equation 3.8 as

$$R = ih\sigma_A \tag{3.9a}$$

Note that h is simply the correlation between an individual's breeding (A) and phenotypic (z) values, $h = \rho_{zA}$. This follows since

$$h = \frac{\sigma_A}{\sigma_z} = \frac{\sigma_A^2}{\sigma_A\sigma_z} = \frac{\sigma(A,z)}{\sigma_A\sigma_z} = \rho_{zA}$$

This correlation quantifies the ability to predict the breeding value of an individual from some measure (here that individual's phenotype) and is called the **accuracy** of the selection scheme used to chose parents. We can thus express the breeders equation in terms of the accuracy of selection as

$$R = i\rho_{zA}\sigma_A \tag{3.9b}$$

Hence, the breeder's equation can be consider as the following product:

Response = (Intensity)*(Accuracy in Predicting Breeding Value)*($\sqrt{\text{Usable Variance}}$)

More generally, if we use some measure u for predicting the breeding value of an individual, then the breeder's equation can be expressed in terms of the accuracy ρ_{uA} of that measure in predicting breeding value,

$$R = i\rho_{uA}\sigma_A \tag{3.10}$$

Example 3.1. **Progeny testing**, using the mean of a parent's offspring to predict the parent's breeding value, is an alternative predictor of an individual's breeding value. In this case, the correlation between the mean of n offspring and the breeding value of the parent is

$$\rho_{uA} = \sqrt{\frac{n}{n+a}}, \quad \text{where} \quad a = \frac{4-h^2}{h^2}$$

From Equation 3.10, the response to selection under progeny testing is

$$R = i\sigma_A\sqrt{\frac{n}{n+a}} = i\sigma_A\sqrt{\frac{h^2 n}{4+h^2(n-1)}}$$

Note that for very large n that the accuracy approaches one. Progeny testing gives a larger response than simple selection on the phenotypes of the parents (**mass selection**) when

$$\sqrt{\frac{n}{4+h^2(n-1)}} > 1, \quad \text{or} \quad n > \frac{4-h^2}{1-h^2}$$

In particular, $n > 4$, 5, and 7, for h^2 = 0.1, 0.25, and 0.5. Also note that the ratio of response for progeny testing (R_{pt}) to mass selection (R_{ms}) is just

$$\frac{R_{pt}}{R_{ms}} = \frac{1}{h}\sqrt{\frac{h^2 n}{4 + h^2(n-1)}} = \sqrt{\frac{n}{4 + h^2(n-1)}}$$

Note for large n that $R_{pt}/R_{ms} \to 1/h \geq 1$.

The Generalized Breeder's Equation: Generation Intervals

So far, we have been assuming non-overlapping generations — all parents only reproduce in one generation interval. In most settings, domesticated animals live multiple years and can have progeny over different years. In such cases, the response should be expressed in terms of response per year. To express the breeder's equation in terms of *response per year*, we first need to compute the **generation intervals** L_x (the average age of parents when progeny are born) for both sexes.

Example 3.2. Compute L_s and L_d for the following age structure, where number denotes the number of offspring produced in the particular year by each sex:

Age of Birth at Progeny

Sires	year 2	year 3	year 4	year 5	Total
Number	60	30	0	0	90

Dams	year 2	year 3	year 4	year 5	Total
Number	400	600	100	40	1140

$$L_s = \frac{2 \cdot 60 + 3 \cdot 30}{60 + 30} = 2.33, \quad L_d = \frac{2 \cdot 400 + 3 \cdot 600 + 4 \cdot 100 + 5 \cdot 40}{400 + 600 + 100 + 40} = 2.81$$

Incorporating the generation intervals, the yearly rate of response can be expressed as

$$R_y = \left(\frac{i_s + i_d}{L_s + L_d}\right) h^2 \sigma_p = \left(\frac{i_s + i_d}{L_s + L_d}\right) h \sigma_A \tag{3.11a}$$

Thus, one way to increase response is to *reduce the generation intervals*, for example by using younger parents. The problem is that there is a tradeoff between generation interval and selection intensity. In species that are reproductively-limited (few offspring per dam), using younger dams means that a higher fraction of the dams must be chosen to replace the population (i.e., to keep the same number of animals in a herd). As a consequence, the selection intensity on these parents (which increases as fewer parents are chosen) is reduced.

More generally, we can combine both selection accuracy (Equation 3.10) and generation interval (Equation 3.11a) to give a more general version of the breeder's equation:

$$R_y = \left(\frac{i_s + i_d}{L_s + L_d}\right) \rho_{uA} \sigma_A \tag{3.11b}$$

Expressed this way, there are three components of response that the breeder has some control over:

- (i) selection intensity (i)
- (ii) generation interval (L)
- (iii) selection accuracy (ρ)

Note that not much can be done with increasing σ^2_A. Response is increased by decreasing L and/or increasing ρ and/or i. We have already discussed tradeoffs between L and i, and there are similar tradeoffs between L and ρ. Clearly, the longer we wait before choosing parents to form the next selected generation, the more accurate we can predict their breeding value, as information from other relatives and from progeny-testing accumulates over time. However, these increases in ρ also result in increases in L. The optimal selection program must balance all of these competing interests.

Equation 3.11b also highlights the importance to animal breeding of advances in reproductive technology such as **artificial insemination** (AI) and **multiple ovulation embryo transplant** (MOET) schemes. The more offspring a parent can produce, the stronger a selection intensity we can apply (and still keep a required fixed number of animals in our herd). Hence, AI has resulted in the potential for far greater selection intensities (and unfortunately far more inbreeding) than would be possible under natural insemination. Likewise, MOET schemes to increase the number of offspring from females potentially allow for increases in the selection intensity on dams as well as decreases in the generation interval.

Truncation Selection

Truncation selection is by far the commonest form of artificial selection in plant and animal breeding and in laboratory experiments. Under truncation selection, only the largest (or smallest) individuals are allowed to reproduce (Figure 3.1). Truncation selection is usually described by either the percent p of the population saved or the threshold phenotypic value T below (above) which individuals are **culled** (removed by selection). The investigator usually sets these in advance of the actual selection. Hence, while S is trivially computed *after* the parents are chosen, we would like to *predict* the expected selection differential given either T or p. Specifically, given either T or p, what is the expected mean of the selected parents?

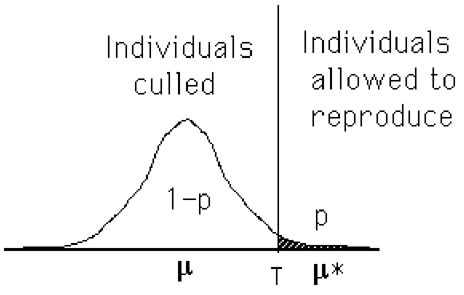


Figure 3.1. Under truncation selection, the uppermost (or lowermost) fraction p of a population is selected to reproduce. Alternatively, one could set a threshold level in advance, above (below) which individuals are allow reproduce. To predict response given either p or T, we need to know the mean of the selected tail (μ^*), from which we can compute $S = \mu^* - \mu$ and then apply the breeder's equation.

Selection Intensities and Differentials Under Truncation Selection

Given the threshold cutoff T, the expected mean of the selected adults is given by the conditional mean, $E(\, z \,|\, z \geq T)$. Generally it is assumed that phenotypes are normally distributed, and we use this throughout. With initial mean μ and variance σ^2, the expected selection differential is

$$S = \varphi\left(\frac{T-\mu}{\sigma}\right)\,\frac{\sigma}{p} \tag{3.12a}$$

where p is the fraction saved and $\varphi(x) = (2\pi)^{-1/2}\, e^{-x^2/2}$ is the unit normal density function evaluated at x. We can show that the (within-generation) mean after selection is

$$\mu_* = E(\, z \,|\, z \geq T) = \mu + S = \mu + \varphi\left(\frac{T-\mu}{\sigma}\right)\,\frac{\sigma}{p} \tag{3.12b}$$

Generally, it is the fraction to be saved p (rather than T) that is preset by the investigator (for example, if we save the uppermost 5 percent, what is the expected S?). Given p, to apply Equation 3.12a, we must first find the threshold value T_p satisfying $\Pr(z > T_p) = p$. Notice that T in Equation 3.12a enters only as $(T - \mu)/\sigma$, which transforms T_p to a scale with mean zero and unit variance. Hence,

$$\Pr\left(\frac{z-\mu}{\sigma} > \frac{T_p - \mu}{\sigma}\right) = \Pr\left(U > \frac{T_p-\mu}{\sigma}\right) = p$$

where $U \sim \mathrm{N}(0,1)$ denotes a unit normal random variable. Define $z_{[p]}$, the **probit transformation** of p, by

$$\Pr(\, U \leq z_{[p]}\,) = p \tag{3.13a}$$

Hence

$$\Pr(\, U > z_{[1-p]}\,) = p \tag{3.13b}$$

It immediately follows that $z_{[1-p]} = (T_p - \mu)/\sigma$, and Equation 3.12a gives the expected selection intensity as

$$i = \frac{S}{\sigma} = \frac{\varphi(z_{[1-p]})}{p} \tag{3.14a}$$

One can obtain $z_{[1-p]}$ from normal distribution tables. Alternatively, a number of approximations have been suggested for Equation 3.14a. Assuming normality, Smith (1969) suggests

$$i \simeq 0.8 + 0.41 \,\ln\!\left(\frac{1}{p} - 1\right) \tag{3.14b}$$

Example 3.3. Consider selection on a normally distributed character in which the upper 5% of the population is saved (p= 0.05). From unit normal tables (e.g., Table 11.1 of LW), $z_{[1-0.05]} = 1.645$ as $\Pr[U \geq 1.645] = 0.05$. Hence,

$$i = \frac{\varphi(1.645)}{0.05} = \frac{0.103}{0.05} \simeq 2.06$$

Applying Equation 3.8 gives the expected response to this amount of selection as $R = h^2\,\sigma\,2.06$. Smith's approximation gives the selection intensity as

$$i \;\simeq\; 0.8 + 0.41 \,\ln\!\left(\frac{1}{0.5} - 1\right) \simeq\; 2.01$$

which is quite reasonable. The **`R`** code, given p, for computing i is just **`dnorm(qnorm(1-p))/p`**.

Gene Frequency Changes Under Selection

Next, we turn to population-genetic considerations of the expected allele frequency changes at loci underlying a quantitative trait.

How quickly does selection change the frequency of alleles at loci contributing to a trait under selection? We start by reviewing a few results from population genetics. Consider a diallelic locus, with alleles A_1 and A_2, whose genotypes have the following relative fitnesses:

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	$1+s$	$1+2s$

This is an example of **additive fitness**. With these fitnesses, for every offspring left by an individual with an A_1A_1 genotype, $1+2s$ offspring are left (on average) by individuals with an A_2A_2 genotype. If q represents the frequency of allele A_2 before selection, then the change in the frequency of q after selection is given by

$$\Delta q = \frac{sq(1-q)}{1+2sq} \simeq sq(1-q) \quad \text{when} \quad |\,2sq\,| << 1 \tag{3.15}$$

Thus, under these fitnesses, the change in the frequency of the favorable allele is proportional to s. In finite populations, genetic drift can overpower the effects of selection. In particular, when

$$4N_e\,|\,s\,| << 1$$

the fate of an allele is largely determined by gene drift, rather than selection. In such cases, favorable alleles can easily be lost by drift.

Now consider a locus contributing to a character z under selection. Suppose the genotypes at this locus make the following contribution to the character:

Genotype	A_1A_1	A_1A_2	A_2A_2
Contribution	0	a	$2a$

For a trait with phenotypic variation σ^2_z under selection intensity i, this induces additive fitnesses on these genotypes, with

$$s \simeq \frac{a}{\sigma_z}\, i \tag{3.16}$$

Hence, the change in allele frequency depends on both the strength of selection i and the relative contribution a/σ_z of the character to the overall trait value. As expected, loci with larger contributions are under stronger selection than loci with minor contributions and hence have faster allele frequency changes (which scales as $1/s$). Further note that if

$$4N_e|\,s\,| = \frac{4N_e|\,a\,i\,|}{\sigma_z} << 1 \tag{3.17}$$

then the effect of selection on this locus is weaker than the effects of drift. Thus, many favorable QTL alleles can be lost by drift if either their effects (a/σ_z), the strength of selection on the character (i), or the effective population size (N_e) are sufficiently small.

More generally, if the locus shows dominance towards the character, the fitnesses become

Genotype	A_1A_1	A_1A_2	A_2A_2
Contribution	0	$a(1+k)$	$2a$
Induced fitness	1	$1+s(1+h)$	$1+2s$

where for the induced fitnesses $s = ai/\sigma_z$(as above) and $h = k$.

II: The Bulmer Equation: Short-term Changes in the Variance

Selection has two routes by which to change the genetic variances, and hence the heritability and selection response. First, it can change the frequencies at individual alleles. When the contribution to a trait from any locus is very small, these selection-induced changes in allele frequencies over a few generations are also very small (scaling, from Equation 3.16, as $\sigma_z/[\,ai\,]$). However, selection also creates correlations between alleles at different loci (**linkage disequilibrium**), and this can result in an immediate change in the variance, ***even among unlinked loci***.

Consider the within-generation change in the variance, $\delta\sigma^2_z = \sigma^2_{z^*} - \sigma^2_z$. Using regression arguments similar to those leading to the breeder's equation, the expected response in the variance to a single generation of selection is

$$d = \sigma^2_O - \sigma^2_z = \frac{h^4}{2}\,\delta\sigma^2_z \tag{3.18a}$$

where σ^2_O is the variance in the offspring and σ^2_z the variance in the unselected population. This the variance response analog to the response in mean (the breeder's equation), with $h^4/2$ replacing h^2 and $\delta\sigma^2_z$ replacing S. In many situations (such as truncation selection), we can write

$$\sigma^2_{z^*} = (1-k)\sigma^2_z, \quad \text{so that} \quad \delta\sigma^2_z = -k\sigma^2_z \tag{3.18b}$$

so that the result of selection is a proportional change (the constant k) in the variance.

It turns out that all the change in the variance is due to a change in the additive genetic variance, so that if σ^2_a denotes the additive variance before selection (i.e., the additive variance in the absence of linkage disequilibrium), then after one generation of selection

$$\sigma^2_A(1) = \sigma^2_a + d, \quad \sigma^2_z(1) = \sigma^2_A(1) + \sigma^2_D + \sigma^2_E = \sigma^2_z + d \tag{3.19}$$

where σ^2_z is the phenotypic variance in the base (pre-selection) population. The heritability thus becomes

$$h^2(1) = \frac{\sigma^2_A(1)}{\sigma^2_z(1)} = \frac{\sigma^2_a + d}{\sigma^2_z + d}$$

Truncation selection reduces the variance ($\delta\sigma_z < 0$), which results in reduced additive genetic variance and heritability in the next generation, slowing response. This reduction in variance due to selection creating linkage disequilibrium is referred to as the **Bulmer effect**, after Michael Bulmer's pioneering work on this subject in the 1970's.

One subtle feature of changes in the variance is that recombination breaks down the selection-induced correlations, so that in the absence of selection, $d(t+1) = d(t)/2$ (for unlinked loci). Hence, one must iterate to obtain the value of the variance in generation t. Starting with an unselected base population, $d(0) = 0$, we obtain the value for $d(t+1)$ by iterating

$$\begin{aligned} d(t+1) &= \frac{d(t)}{2} + \frac{h^4(t)}{2}\,\delta\sigma_{z(t)} \\ &= \frac{d(t)}{2} - k\frac{h^4(t)}{2}\,\sigma^2_z(t) \end{aligned} \tag{3.20}$$

The first term ($d/2$) is the decay in linkage disequilibrium from recombination while the second term is the amount of new disequilibrium created by selection. Note for above that

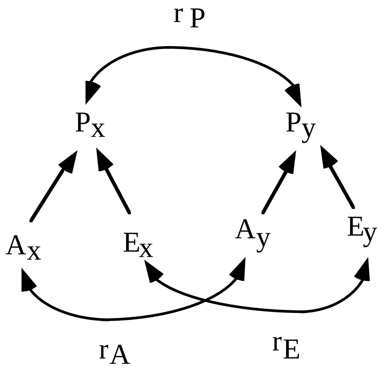
$$\sigma^2_z(t) = \sigma^2_z(0) + d(t), \qquad \text{and} \qquad h^2(t) = \frac{\sigma^2_A(t)}{\sigma^2_z(t)} = \frac{\sigma^2_a + d(t)}{\sigma^2_z + d(t)} \tag{3.21}$$

While all this looks rather complicated at first glance, its really a very straight forward series of substitutions. The net result for directional selection is that most of the reduction in variance occurs over the first few generations, which rapidly approaches an equilibrium value (the equilibrium reduction in the additive variance). However, under disruptive selection (selection to increase the variance, for example by selecting both the largest and smallest parents), the variance may continue to increase substantially over many generations before settling on its equilibrium value.

III: Multivariate Triats: Genetic and Environmental Correlations

Many characters are positively or negatively correlated at the level of phenotype (e.g., height and weight, forearm length and digit span) and we can directly measure the phenotypic correlation, r_P, between two traits X and Y.

As the following (path) diagram indicates, the phenotypic correlation r_P between two traits is generated by correlations between the genetic (r_A) and/or environmental (r_E) values of X and Y. In the figure, double-headed arrows imply possible correlations between variables.



r_A = correlation of breeding values arises from two sources

- **pleiotropic effects** of loci on both traits; correlation from pleiotropy indicates the extent to which the character is controlled by the same genes
- linkage disequilibrium, which will decay over time

r_E = correlation of environmental deviations

- includes non-additive genetic effects
- arises from exposure of the two traits to the same individual environment

Recall (Lecture 1) that the correlation between X and Y equals $r = cov(X,Y)/\sigma_X\sigma_Y$. Rearranging, we can express the covariance as

$$cov(X,Y) = r\,\sigma_X\,\sigma_Y$$

Therefore

$$cov_P = r_P\,\sigma_{XP}\,\sigma_{YP}, \quad cov_A = r_A\,\sigma_{XA}\,\sigma_{YA}, \quad cov_E = r_E\,\sigma_{XE}\,\sigma_{YE} \tag{3.22}$$

The covariance of phenotypic values = the sum of the genetic and environmental covariances:

$$cov_P = cov_A + cov_E$$

so that

$$r_P\,\sigma_{XP}\,\sigma_{YP} = r_A\,\sigma_{XA}\,\sigma_{YA} + r_E\,\sigma_{XE}\,\sigma_{YE} \tag{3.23}$$

describes the relationship between phenotypic, genetic and environmental correlations.

This expression can be simplified somewhat. First, define h_X and h_Y as the square roots of the heritabilities of characters X and Y, with

$$h = \frac{\sigma_A}{\sigma_P}, \quad \sigma_A = h\sigma_P$$

Likewise define $e^2 = 1 - h^2$

$$e^2 = \frac{\sigma_E^2}{\sigma_P^2}, \quad e = \sqrt{1-h^2} = \frac{\sigma_E}{\sigma_P}, \quad \sigma_E = e\sigma_P \tag{3.24}$$

Substituting $\sigma_A = h\sigma_P$ and $\sigma_E = e\sigma_P$ into Equation 3.23 gives

$$\begin{aligned} r_P &= h_X\,h_Y\,r_A + e_X\,e_Y\,r_E \\ &= r_A\,h_X\,h_Y + r_E\,\sqrt{(1-h_x^2)(1-h_Y^2)} \end{aligned} \tag{3.25}$$

Hence, the phenotypic correlation is a function of the heritabilities of the traits and the genetic and environmental correlations. If heritabilities are high, the genetic correlation is more important; if heritabilities are low, the environmental correlation is more important. In practice, phenotypic and genetic correlations often have the same sign and are of similar magnitude, but this is not always the case.

Estimating the Genetic Correlation

Methods for estimating r_A and r_E are analogous to estimating heritabilities from resemblance between relatives. The difference is that with a single trait, the covariance between the trait value in two relatives provides an estimate of the additive genetic variance of that trait, while with two traits, the covariance of trait X in one relative and trait Y in the other provides an estimate of the additive genetic covariance between the two traits.

Offspring-parent

The covariance of trait X in the offspring with trait Y of the parents

$$cov(X_O, Y_P) = (1/2) cov_{A_X, A_Y} \tag{3.26a}$$

The covariance of trait X in the offspring with trait X of the parents

$$cov(X_O, X_P) = (1/2) \sigma^2_{A_X} \tag{3.26b}$$

The covariance of trait Y in the offspring with trait Y of the parents

$$cov(Y_O, Y_P) = (1/2) \sigma^2_{A_Y} \tag{3.26c}$$

Hence,

$$r_A = \frac{cov(X_O, Y_P)}{\sqrt{cov(X_O, X_P) \cdot cov(Y_O, Y_P)}} \tag{3.26d}$$

Estimates of genetic correlation have very large sampling errors, typically requiring extremely large experiments for precise estimates.

Half sibs

The covariance of traits X and Y between sires = $cov_{XY} = (1/4) cov_A = (1/4) cov_{A_X, A_Y}$

The variance between sires of trait $X = \sigma^2_{S_X} = (1/4)\sigma^2_{A_X}$

The variance between sire of trait $Y = \sigma^2_{S_Y} = (1/4)\sigma^2_{A_Y}$

Therefore, $r_A = cov_{XY} / (\sigma_{S_X} \sigma_{S_Y})$

Genotype-environment interaction and the cross-environment genetic correlation

If a character is measured in two environments, it can be considered as two characters (e.g., X in environment one, Y in environment two). By rearing families in both environments, it is possible to estimate the genetic correlation between these ”characters”. The magnitude of this cross-environment correlation reflects the extent to which the same genes control the character in each environment.

IV: Correlated Response to Selection

When characters are genetically correlated, selection solely on one will result in a correlated change in the second. Such a change in the unselected character is called a **correlated response**.

We compute the expected correlated response in Y given selection on X as follows. The response to selection of character X (the mean value of offspring of selected parents) is (by definition) the mean breeding value of the selected group. Thus the change in character Y in response to selection on X is the regression of the breeding value of Y on the breeding value of X. The slope of this regression is given by

$$b_{A_Y | A_X} = \frac{cov_A}{\sigma^2_{A_X}} = \frac{r_A \, \sigma_{A_X} \, \sigma_{A_Y}}{\sigma^2_{A_X}} = r_A \, \frac{\sigma_{A_Y}}{\sigma_{A_X}} \tag{3.27}$$

Recalling first that a regression passes through the mean of both variables (with $y - \mu_y = b_{y\,|\,x}[x - \mu_x]$) and second that the breeding values have mean zero ($\mu_{A_X} = \mu_{A_Y} = 0$), the regression of the breeding values of Y on the breeding values of X is just

$$Y = b_{A_Y | A_X} \, X = r_A \, \frac{\sigma_{A_Y}}{\sigma_{A_X}} \, X \tag{3.28}$$

The response of the directly selected character X is

$$R_X = i_X\, h^2_X\, \sigma_{PX} = i\,(\sigma^2_{A_X}/\sigma^2_{P_X})\,\sigma_{P_X} = i_X\,\sigma^2_{A_X}/\sigma_{P_X} = i_X\, h_x\, \sigma_{AX} \tag{3.29a}$$

where

$$i_X = S_x/\sigma_X \tag{3.29b}$$

Recalling our comment above that R_X is the change in the breeding value of X, the correlated response of character Y is

$$\begin{aligned} CR_Y &= b_{A_Y|A_X} R_X \\ &= (r_A\, \sigma_{A_Y}/\sigma_{A_X})\,(i_X\, h_X\, \sigma_{A_X}) \\ &= r_A\, \sigma_{A_Y}\, i_X\, h_X \end{aligned}$$

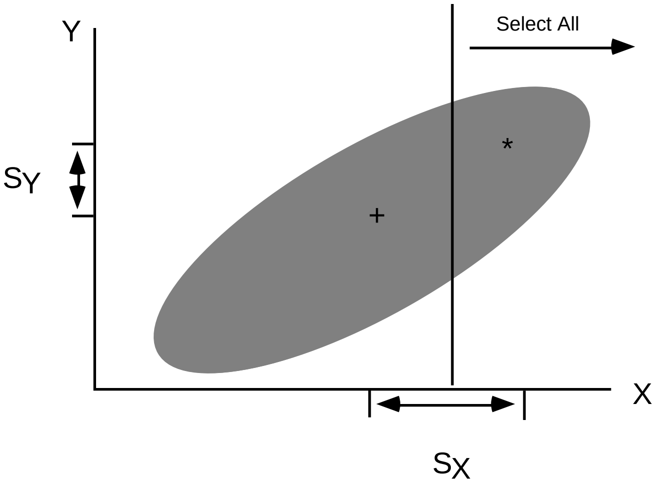
Substituting $\sigma_{A_Y} = h_Y\, \sigma_{PY}$ gives

$$CR_Y = i_X\, h_X\, h_Y\, r_A\, \sigma_{P_Y} \tag{3.30}$$

Noting that the direct response on X is $R_X = i_X\, h^2_X\, \sigma_{P_X}$, we see that h^2_X and $h_X\, h_Y\, r_A$ play similar roles, resulting in the later being called the **co-heritability**.

Correlated Selection Differentials

When selection is applied to character X, character Y will show a correlated selection differential when Y is phenotypically correlated with X. The figure below shows the change in the bivariate mean before $(+)$ and after $(*)$ selection. In this case, there is truncation selection only on trait X, but there is a correlated within-generation change in Y, with the mean of Y in the selected parents differing from the mean before selection.



Note that the correlated selection differential simply measures the *within-generation* change. Whether this translates into a *between-generation* change (i.e., a response to selection) depends on whether there is any genetic correlation between the characters. With only phenotypic correlations, none of the correlated within-generation change is passed on to the offspring.

The reason for this can be illustrated by a hypothetical example, with truncation selection only on X. Suppose that $r_A = 0$ and $r_E > 0$; therefore r_P and S_Y will be > 0. The selected individuals, however, will not have higher than average breeding values for Y, in spite of their higher than average phenotypic values. Therefore the correlated response will be zero, regardless of the magnitude of h^2_Y and S_Y.

Estimating the Genetic Correlation from Selection Response

Another method for estimating the genetic correlation is analogous to the realized heritability. Recall in the latter case, heritability is estimated by $h^2 = R/S$.

One procedure to obtain a realized estimate of the additive genetic correlation is as follows: From the same base population, in different lines:

- select individuals on the basis of character X. From these lines you can measure the direct response to character $X(R_X)$ and the correlated response of character $Y(CR_Y)$.
- select individuals on the basis of character Y. From these lines you can measure the direct response to character $Y(R_Y)$ and the correlated response of character $X(CR_X)$.

From the identities above, note that

$$r_A^2 = \frac{CR_X}{R_X}\frac{CR_Y}{R_Y} \tag{3.31}$$

Similarly, to obtain a realized estimate of the genetic covariance between a trait in two different environments,

- select individuals for character X in environment 1, and measure response in environment 1 (R_1) and environment 2 (CR_2)
- select individuals for character X in environment 2, and measure response in environment 2 (R_2) and environment 1 (CR_1)

The again provides an estimate of the genetic covariance of the trait in the two environments as

$$r_{GE}^2 = \frac{CR_1}{R_1}\frac{CR_2}{R_2} \tag{3.32}$$

Example: Computing the genetic correlation from a double selection experiment

The experiment is as follows: Select for increased and decreased abdominal bristle number, and increased and decreased sternopleural bristle number, from the same base population of Drosophila melanogaster.

- Large base population (the Raleigh population of the realized h^2 example); select for high and low values of each trait; 25 selected out of 100 scored per sex; selection continued for 25 generations.

Results: At generation 25, the mean abdominal (AB) and sternopleural (ST) bristle numbers in the four selection lines were:

	Mean Bristle Number	
Selection Line	AB	ST
High AB	33.4	26.4
Low AB	2.4	12.8
High ST	22.2	45.0
Low ST	11.1	9.5

Hence, selection to increase AB gives a direct response in AB of 33.4 and a correlated response in ST of 26.4. Expressing these responses in terms of divergence selection, we have

- Response in AB = $R_{AB} = 33.4 - 2.4 = 31.0$
- Response in ST = $R_{ST} = 45.0 - 9.5 = 35.5$
- Correlated response in ST = $CR_{ST} = 26.4 - 12.8 = 13.6$
- Correlated response in AB is $CR_{AB} = 22.2 - 11.1 = 11.1$
- The estimated genetic correlation of abdominal and sternopleural bristle number is

$$r_A = \sqrt{\frac{CR_{AB}}{R_{AB}}\frac{CR_{ST}}{R_{ST}}} = \sqrt{\frac{11.1}{31}\frac{13.6}{35.3}} = 0.37$$

The positive genetic correlation between the two bristle traits may be due to linkage disequilibrium or pleiotropy. Not all loci affecting the trait necessarily have the same pleiotropic effects; one could conceive of a situation in which all loci were pleiotropic but r_A is zero, if pleiotropic effects are not directional across loci.

Indirect Selection

There are two ways the mean of a character X can change by selection:

- as a direct response to selection for trait $X(R_X)$
- as a correlated, or indirect, response to direct selection for trait $Y(CR_X)$

The relative magnitudes of the change in mean is given by the ratio

$$\frac{CR_X}{R_X} = \frac{i_Y\, r_A\, \sigma_{AX} h_Y}{i_X\, h_X\, \sigma_{AX}} = \frac{i_Y\, r_A\, h_Y}{i_X\, h_X} \tag{3.33}$$

Therefore the correlated response of X to selection for Y will be greater than direct response to selection for X when $i_Y\, r_A\, h_Y > i_X\, h_X$, or when

- character Y has a greater heritability than X, and the genetic correlation between X and Y is high. This could occur if X is difficult to measure with precision but Y is not.
- the selection intensity is much greater for Y than X. This would be true if Y were measurable in both sexes but X measurable in only one sex.

V: Multitrait Selection Response in Matrix Form

The response to selection of several traits is best handled using matrix notation, which we briefly introduce here. A **matrix** is an array of elements, e.g.

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} i \\ j \end{pmatrix}$$

When the matrix consists of only a single row or a single column, it is called a **vector**, whereas **A** and **B** are **square matrices**. **C** above is a column vector (consisting of a single column of entries).

Matrix multiplication is defined in such a fashion that the order of multiplication of the matrices is critical. For the above matrices,

$$\mathbf{AB} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}, \qquad \mathbf{BA} = \begin{pmatrix} ae+cf & eb+df \\ ga+ch & gd+dh \end{pmatrix}$$

and

$$\mathbf{AC} = \begin{pmatrix} ai+bj \\ ci+dj \end{pmatrix}, \qquad \mathbf{BC} = \begin{pmatrix} ei+fj \\ gi+hj \end{pmatrix}$$

while the matrix products **CA** and **CB** are not defined. The **identity matrix, I**, which serves a role similar to 1 in scaler multiplication/division, is given by (for the 2 x 2 cases),

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that **AI** = **IA** = **A**. Finally, the **Inverse** $\mathbf{A}^{-1}$ of a square matrix **A** is defined as satisfying

$$\mathbf{A^{-1}A} = \mathbf{AA^{-1}} = \mathbf{I} \tag{3.34a}$$

For a 2 x 2 matrix,

$$\mathbf{A}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{3.34b}$$

The role of the inverse and identity matrix is in solving systems of equations. Suppose we are trying to solve for **x** in **AX** = **C**. Premultiplying both sides by the inverse of **A** gives

$$\mathbf{A^{-1}Ax} = \mathbf{A^{-1}C}$$

noting that $\mathbf{A^{-1}Ax} = \mathrm{Ix} = \mathbf{x}$ gives $\mathbf{x} = \mathbf{A^{-1}C}$

We now have all the pieces in place to express the response in matrix form. Suppose there are n traits under selection and we place the n selection differentials for each trait into a column vector **S**,

$$\mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix} \tag{3.35a}$$

Likewise, define the **phenotypic** and **additive genetic covariance matrices**, **P** and **G**, respectively, as matrices whose element in the ith row and j column is the covariance (phenotype or additive genetic) between traits i and j. Note that the diagonal elements are the variances. For example, for two characters

$$\mathbf{P} = \begin{pmatrix} \sigma^2(z_1) & \sigma(z_1,z_2) \\ \sigma(z_1,z_2) & \sigma^2(z_2) \end{pmatrix}, \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \sigma^2(A_1) & \sigma(A_1,A_2) \\ \sigma(A_1,A_2) & \sigma^2(A_2) \end{pmatrix} \tag{3.35b}$$

Let **R** denote the column vector of selection responses, so that the ith element in the list is R_i, the change in the mean of character i following one generation of selection. The response to selection becomes

$$\mathbf{R} = \mathbf{GP}^{-1}\mathbf{S} \tag{3.36}$$

This equation is often referred to as the **multidimensional breeder's equation**. Recall that the response for a single character under selection is $R = h^2 S = \sigma^2_A(\sigma^2_P)^{-1}S$. In the multidimensional case, the genetic and phenotypic variance are replaced by variance-covariance matrices and we use matrix inversion and multiplication.

The Directional Selection Gradient

The multivariate breeder's equation can also be written as

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} \tag{3.37a}$$

where

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S} \tag{3.37b}$$

is called the **directional selection gradient**. The ith element of $\boldsymbol{\beta}$, β_i, measures the amount of direct selection on trait X_i (i.e., the effects of correlated selection differentials are removed). From the rules of matrix multiplication, the response in trait j can be written as

$$R_j = \sigma^2(A_j)\,\beta_j + \sum_{i\neq j}\sigma(A_j,A_i)\,\beta_i \tag{3.38a}$$

where the first term is the change due to direct selection on trait j and the sum is the indirect contribution to the response due to the correlated response of selection on other traits. Likewise note that $\mathbf{P}\boldsymbol{\beta} = \mathbf{S}$, so that the selection differential on trait j can be written as

$$S_j = \sigma^2(P_j)\,\beta_j + \sum_{i \neq j} \sigma(P_j, P_i)\,\beta_i \tag{3.38b}$$

where the first term represents the contribution from direct selection on trait j and the sum term the contribution to the within-generation change due to direct selection on phenotypically correlated traits.

Effect of Selection on Genetic Correlation

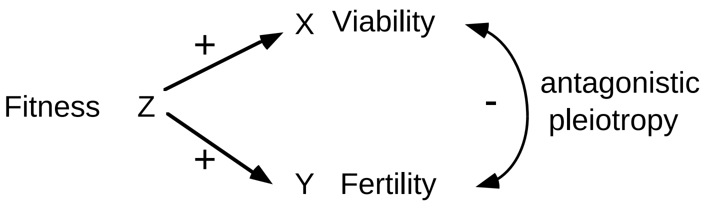
Sometimes two or more correlated characters are selected jointly — e.g. selection for increased body length and weight of mice. This is also true of natural selection, for which all traits contributing positively to the composite trait, fitness, are jointly selected.

One consequence of simultaneous selection for correlated characters is that the genetic correlation between them can becomes negative as alleles become fixed.

Consider a simple example of a composite trait Z controlled by multiple genes, which have different pleiotropic effects on two component traits, X and Y:

Locus	Effects of allele 1 relative to allele 2:	
	on X	on Y
A	+	+
B	+	-
C	-	+
D	-	-

The effect of selection on Z will be to fix allele 1 at locus A and allele 2 at locus D. However alleles at loci B and C will remain segregating at intermediate frequencies. The genetic correlation of traits X and Y tend to become negative because the only genes that remain affect X and Y in opposite directions. The heritability of trait Z will tend to 0 and it will exhibit no further selection response. The heritabilities of traits X and Y will be greater than 0, and there would be a response to selection for either trait separately. In this case the mean value of the correlated trait would decline.



Lecture 3 Problems

1. Taking the selection differential as the difference between the means of selected parents and the mean before selection makes the assumption that each selected parent contributes equally to the next generation. Biases introduced by differential fertility can be removed by using **effective selection differentials**, S_e,

$$S_e = \frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{n_i}{\overline{n}} \right) (z_i - \mu_z) = \left(\frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{n_i}{\overline{n}} \right) z_i \right) - \mu_z$$

where z_i and n_i are the phenotypic value and total number of offspring of the ith parent, n_p the number of parents selected to reproduce, $\overline{n}$ the average number of offspring for selected parents, and μ_z is the mean before selection. If all selected parents have the same number of offspring ($n_i = \overline{n}$ for all i), then S_e reduces to S. However, if there is variation in n_i among selected parents, S_e can be considerably different from S. This corrected differential is occasionally referred to as the **realized selection differential**.

Suppose 5 parents are selected, with the following trait values and offspring number:

Parent	phenotypic value	number of offspring
1	45	1
2	40	2
3	35	3
4	33	5
5	32	5

If the mean before selection is 30, compute the S and S_e. If $h^2 = 0.3$, what is the expected response that would be estimated under the two differentials?

2. Consider a population not currently under selection, with $\sigma^2_z = 100$ and $h^2 = 0.5$ and $d(0) = 0$ (no disequilibrium). Consider two types of selection (i) stabilizing where $\sigma^2_{z*} = 0.5\sigma^2_z$ (i.e., $k = 1/2$ in Equation 3.18b) and (ii) disruptive selection $\sigma^2_{z*} = 1.5\sigma^2_z$ ($k = -1/2$). For both types of selection compute $d(1)$ and $d(2)$, $\sigma^2_A(1)$ and $\sigma^2_A(2)$, $\sigma^2_z(1)$ and $\sigma^2_z(2)$, and $h^2(1)$ and $h^2(2)$.

3. Consider selection acting on two traits (1 and 2). The phenotypic variances and covariances are $\sigma^2(z_1) = \sigma^2(z_2) = 10, \sigma(z_1, z_2) = -5$, while the additive genetic variances and covariances are $\sigma^2(A_1) = 4, \sigma^2(A_2) = 9, \sigma(A_1, A_2) = 3$. Compute the response on both characters when:

a: We select directly on trait 2, with $S_2 = 10$ (use Equations 3.28 and 3.30)

b: We select directly on trait 1, with $S_1 = 10$ (use Equations 3.28 and 3.30).

c: $S_1 = 5, S_2 = 5$ (use the multivariate breeders equation)

4. The following covariances of performance in randomly chosen dam and daughter pairs of dairy cattle were obtained from an analysis. Estimate the heritabilities of milk yield and fat % and the genetic, phenotypic, and environmental correlations between them.

	Dam's yield	Dam's fat %
Dam's yield (in 100 kg units)	68	-0.55
Dam's fat %	-0.55	0.11
Daughter's yield	7.8	-.20
Daughter's fat	-0.18	0.035

5. The heritability of growth rate in pine seedlings is 0.5 when they are grown in the greenhouse, and 0.2 when grown in the field. The genetic correlation between growth rate in the two environments is 0.8. Suppose you wish to select for increased growth in the field. In which environment should you do the selection? How would your conclusions change if the genetic correlations between growth rate in the two environments was 0.5?

Solutions to Lecture 3 Problems

1. Here $\mu_* = 37$, giving $S = 7$, while $\overline{n} = 3.2$ and

i	z_i	n_i	$n_i/\overline{n}$	$z_i \cdot n_i/\overline{n}$
1	45	1	0.3125	5.06
2	40	2	0.6250	25.00
3	35	3	0.9375	32.81
4	33	5	1.563	51.56
5	32	5	1.563	50.0

$$\frac{1}{n_p}\sum_{i=1}^{n_p}\left(\frac{n_i}{\overline{n}}\right) z_i = 34.69$$

Giving $S_e = 4.69$. Assuming $h^2 = 0.3$, using the uncorrected S gives a response of $R = 0.3 \cdot 7 = 2.1$, while the true expected response if $R = 0.3 \cdot 4.69 = 1.4$

2. Here $\sigma_a^2 = h^2\sigma_z^2 = 50$, and $d(0) = 0$

$$d(1) = d(0) - k(h^4/2)\sigma_z^2(0) = \begin{cases} 0 - 0.5 * 0.125 * 100 = -6.25 & \text{for stabilizing}, k = 0.5 \\ 0 + 0.5 * 0.125 * 100 = 6.25 & \text{for disruptive}, k = -0.5 \end{cases}$$

$$\sigma_A^2(1) = \sigma_a^2 + d(1) = \begin{cases} 43.75 & \text{for stabilizing} \\ 56.25 & \text{for disruptive} \end{cases}, \quad \sigma_z^2(1) = \sigma_z^2 + d(1) = \begin{cases} 93.75 & \text{for stabilizing} \\ 106.25 & \text{for disruptive} \end{cases}$$

$$h^2(1) = \sigma_A^2(1)/\sigma_z^2(1) = \begin{cases} 0.467 & \text{for stabilizing} \\ 0.529 & \text{for disruptive} \end{cases}$$

$$d(2) = d(1)/2 - k(h^4(1)/2)\sigma_z^2(1) = \begin{cases} -6.25/2 - 0.5(0.46^2/2) * 93.75 = -8.02 & \text{for stabilizing} \\ 6.25/2 + 0.5(0.53^2/2) * 106.25 = 10.59 & \text{for disruptive} \end{cases}$$

$$\sigma_A^2(2) = \sigma_a^2 + d(2) = \begin{cases} 41.77 & \text{for stabilizing} \\ 60.57 & \text{for disruptive} \end{cases}, \quad \sigma_z^2(2) = \sigma_z^2 + d(2) = \begin{cases} 91.77 & \text{for stabilizing} \\ 110.6 & \text{for disruptive} \end{cases}$$

$$h^2(2) = \sigma_A^2(2)/\sigma_z^2(2) = \begin{cases} 0.455 & \text{for stabilizing} \\ 0.548 & \text{for disruptive} \end{cases}$$

3. Here

$$h_X = \sqrt{\frac{4}{10}} = 0.63, \quad h_Y = \sqrt{\frac{9}{10}} = 0.95, \quad r_A = \frac{3}{\sqrt{4 \cdot 9}} = 0.5$$

a) $S_2 = 10$ implies $i = 10/\sqrt{10} = 3.16$

$$R_2 = i\, h_2^2\, \sigma_{z_w} = 3.16 \cdot 0.9 \cdot \sqrt{10} = 8.99$$
$$CR_1 = r_A\, \sigma_{A_1}\, i\, h_2 = 0.5 \cdot \sqrt{4} \cdot 3.16 \cdot 0.95 = 3.00$$

b) $S_1 = 10$ implies $i = 10/\sqrt{10} = 3.16$

$$R_1 = i\, h_1^2\, \sigma_{z_1} = 3.16 \cdot 0.4 \cdot \sqrt{10} = 4.00$$
$$CR_2 = r_A\, \sigma_{A_2}\, i\, h_1 = 0.5 \cdot \sqrt{9} \cdot 3.16 \cdot 0.63 = 2.99$$

c) $S_1 = S_2 = 5$. Using the multivariate breeder's equation,

$$\mathbf{S} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 10 & -5 \\ -5 & 10 \end{pmatrix},$$

First note that

$$\mathbf{P}^{-1} = \frac{1}{15} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \text{and} \quad \mathbf{P}^{-1}\mathbf{S} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Hence

$$\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S} = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 12 \end{pmatrix}$$

4. Let character 1 denote yield, 2 denote fat %. First, using the covariance of a trait with itself to estimate the phenotypic variances, we have

$$\sigma^2_{z_1} = 68, \quad \sigma^2_{z_2} = 0.11$$

Second, the additive genetic variance in the trait is twice the covariance (for the same character) between parent and offspring, giving

$$\sigma^2_{A_1} = 2 \cdot 7.8 = 15.6, \quad \sigma^2_{A_2} = 2 \cdot 0.035 = 0.07$$

The phenotypic covariance is the covariance between characters 1 and 2 in the parent,

$$\sigma(z_1, z_2) = -0.55$$

Finally, the additive genetic covariance is twice the covariance for trait 1 in the parent and trait 2 in the offspring. Likewise trait 2 in parent and 1 in offspring also estimates this covariance, so we take the average of these two,

$$\sigma(A_1, A_2) = 2(1/2)(-0.20 - 0.18) = -0.38$$

Thus

$$h_1^2 = \frac{15.6}{68} = 0.23, \quad \text{and} \quad h_2^2 = \frac{0.07}{0.11} = 0.64$$

$$r_p = \frac{-0.55}{\sqrt{68 \cdot 0.11}} = -0.20$$

$$r_A = \frac{-0.38}{\sqrt{15.6 \cdot 0.07}} = -0.36$$

Finally, since $V_E = V_P - V_A$, and $Cov(z_1, z_2) = Cov(A_1 + A_2) + Cov(E_1, E_2)$

$$V_{E_1} = 68 - 15.6 = 52.4, \quad V_{E_2} = 0.11 - 0.07 = 0.4, \quad Cov(E_1, E_2) = -0.55 - (-0.38) = -0.17$$

giving

$$r_e = \frac{-0.17}{\sqrt{52.4 \cdot 0.4}} = -0.04$$

5. Compare the ratio of the direct and correlated response. Let X denote the trait value in the field and Y the value in the lab. Assuming the same amount of selection in either setting, the ratio of the correlated field response CR_X (based on selection in the lab) to the direct response R_X if selection is in the field is

$$\frac{CR_X}{R_X} = \frac{i_Y r_A \sigma_{A_X} h_Y}{i_X h_X \sigma_{A_X}} = \frac{r_A h_Y}{h_X} = \frac{0.8\sqrt{0.5}}{\sqrt{0.2}} = 1.265$$

So that a larger response in the field is given by selecting in the lab than by selecting directly in the field.

If the correlation between environments is 0.5, then

$$\frac{CR_X}{R_X} = \frac{0.5\sqrt{0.5}}{\sqrt{0.2}} = 0.791,$$

in which case direct selection in the field is more efficient.