

19

Individual Fitness and the Measurement of Univariate Selection

Natural selection is not evolution — R. A. Fisher (1930)

Previous chapters examined the response to selection, assuming that the nature of selection is known. Here we are concerned with the complementary issue of measuring how selection acts on particular phenotypes (**phenotypic selection**). The estimation of selection involves two related issues: measuring individual fitness and measuring how the phenotype of a particular character influences individual fitness. The latter question is often phrased in terms of estimating $W(z)$, the expected fitness for an individual with character value z .

The first half of this chapter deals with various aspects of individual fitness, focusing on fitness components and measuring fitness over multiple episodes of selection. We conclude by examining the properties of an especially useful statistic, the population variance in relative fitness (the opportunity for selection), which bounds the maximum possible within-generation change in the mean and variance of any character.

The second half moves from individual fitness to the complementary problem of predicting the expected fitness of an individual given its phenotypic value. Our discussion here is concerned with selection acting exclusively on a single character. This admittedly uncommon situation offers the advantage of allowing basic methodological points to be stressed without the additional complications inherent in a multivariate analysis. In Chapter 31 we extend these univariate ideas to the situation where individual phenotypic value is a vector \mathbf{z} . The major complication with multiple characters is selection on phenotypically correlated characters. A within-generation change in the distribution of a trait may be due to **direct** selection on that character, **indirect** effects of selection on correlated (and often unmeasured) characters, or both. Methods for uncoupling these effects were mentioned briefly in Chapter 7 (Equation 7.10) and are developed much more fully in Chapters 31-33.

Excellent discussions on the detection of selection in natural populations can be found in Endler (1986) and Manly (1985). Endler reviews reports of natural selection in the wild, while Manly is an excellent source for statistical methods

for measuring selection on phenotypes. The use of individual fitness data is developed in Arnold and Wade (1984a,b) and Lande and Arnold (1983).

EPISODES OF SELECTION

Selection is often subdivided into components called **episodes of selection**. For example, a distinction is often made between **viability selection** (differences in survivorship) and **fertility selection** (differences in number of offspring per mating). **Tradeoffs** may be found, wherein a trait that does well in one episode does poorly in another. For example, large body size is usually favored in adults of Darwin's Medium Ground Finch *Geospiza fortis* (Boag and Grant 1981, Price et al. 1984), while small body size is apparently favored in juveniles (Price and Grant 1984). The possibility of tradeoffs between **natural** and **sexual selection** first suggested by Darwin (1859) has also received significant attention (Darwin 1859, 1871, Fisher 1930; recent ideas are reviewed in Arnold 1983, Bateson 1983, Bradbury and Andersson 1987). Sexual selection results from variance in male mating success due to male-male competition for females and/or female choice of particular males, while natural selection results from variance in all other fitness components, such as viability and fertility differences, differences in parental care, etc.

Fitness Components

Loosely stated, the **lifetime** (or **total**) *fitness* of an individual is the number of descendants it leaves at the start of the next generation. When measuring the total fitness of an individual, care must be taken not to cross generations or to overlook any stage of the life cycle in which selection acts. To accommodate these concerns, lifetime fitness is defined as the total number of zygotes (newly fertilized gametes) that an individual produces. Measuring total fitness from any other starting point in the life cycle (e.g., from adults in one generation to adults in the subsequent generation) can result in a very distorted picture of true fitness of particular phenotypes (Prout 1965, 1969). If generations are crossed, measures of selection on a particular parental phenotype in reality are averages over both parental and offspring phenotypes, which may differ considerably.

Systems for measuring lifetime fitness have been especially well developed for laboratory populations of *Drosophila* (reviewed by Sved 1989). Measurements of lifetime fitness in field situations are more difficult and (not surprisingly) are rarely made. Attention instead is usually focused on particular episodes of selection or particular phases of the life cycle. Fitness components for each episode of selection are defined to be multiplicative. For example, lifetime fitness can be partitioned as (probability of surviving to reproductive age)·(number of mates)·(number of zygotes per mating). Number of mates is a measure of sexual selection, while the viability and fertility components measure natural selection.

A commonly measured fitness component is **reproductive success**, the number of offspring per adult, which confounds natural (fertility) and sexual selection (in males, the number of matings per adult). Clutton-Brock (1988) reviews estimates of reproductive success from natural populations.

Fitness components can themselves be further decomposed. For example, fertility in plants might be decomposed as (seeds per plant) = (number of stems per plant) · (number of inflorescences per stem) · (average number of seed capsules per inflorescence) · (average number of seeds per capsule). This decomposition allows us to ask questions of the form: do plants differ in number of seeds mainly because some plants have more stems, or more flowers per stem, or are there tradeoffs between these?

Estimates of fitness can be obtained from either **longitudinal** or **cross-sectional** studies. A longitudinal study follows a cohort of individuals over time, while a cross-sectional study examines individuals at a single point in time. Cross-sectional studies typically generate only two fitness classes (e.g., dead versus living, mating versus unmated). Analysis of cross-sectional studies involves a considerable number of assumptions (Lande and Arnold 1983, Arnold and Wade 1984b), and longitudinal studies are preferred. Unfortunately, these usually require far more work and may be impossible to carry out in many field situations. Age-structured populations pose further complications in that proper fitness measures require knowledge of the population's demography. See Charlesworth (1980), Lande (1982), Lenski and Service (1982), and Travis and Henrich (1986) for details.

Assigning Fitness Components

We now turn to the task of partitioning measures of individual fitnesses in a longitudinal study into fitness components. A cohort of n individuals (indexed by $1 \leq r \leq n$) is followed through several episodes of selection. Let $W_j(r)$ be the fitness measure for the j th episode of selection for the r th individual. For example, if we are following viability W_j is either zero (dead) or one (alive) at the census period. Relative fitness components $w_j(r) = W_j(r)/\bar{W}_j$ will turn out to be especially useful. At the start of the study, the frequency of each individual is $1/n$, giving for the first (observed) episode of selection

$$\bar{W}_1 = \frac{1}{n} \sum_{r=1}^n W_1(r) \quad (19.1a)$$

We need to caution at this point that *considerable selection may have already occurred prior to the life cycle stages being examined*. Following the first episode of selection, the new fitness-weighted frequency of the r th individual is $w_1(r)/n$, implying

$$\bar{W}_2 = \sum_{r=1}^n W_2(r) \cdot w_1(r) \cdot \left(\frac{1}{n}\right) \quad (19.1b)$$

In general, for the j th episode of selection,

$$\bar{W}_j = \sum_{r=1}^n W_j(r) \cdot w_{j-1}(r) \cdot w_{j-2}(r) \cdots w_1(r) \cdot \left(\frac{1}{n}\right) \quad (19.1c)$$

Note that if $W_j(r) = 0$, further fitness components for r are unmeasured. Letting $p_j(r)$ be the fitness-weighted frequency of individual r after j episodes of selection, it follows that $p_0(r) = 1/n$ and

$$p_j(r) = w_j(r) \cdot p_{j-1}(r) = \frac{1}{n} \prod_{i=1}^j w_i(r) \quad (19.2a)$$

Thus, Equation 19.1c can also be expressed as $\bar{W}_j = \sum W_j(r) \cdot p_{j-1}(r)$. Using these weights allows fitness-weighted moments to be calculated, e.g., the mean of a particular character following the j th episode satisfies

$$\mu_{z(j)} = \sum z(r) \cdot p_j(r) \quad (19.2b)$$

where $z(r)$ is the value of the character of individual r .

Example 1. Total reproductive success and its components, mating success (W_1) and fertility (W_2 , eggs per successful mating), were measured in 38 male bullfrogs (*Rana catesbeiana*) in a longitudinal study by Howard (1979). For illustrative purposes, we use part of this data set to compute these fitness components for five males.

Male	W_1	W_2	w_1	p_1	w_2	p_2
1	1	25,820	0.714	0.143	1.628	0.233
2	1	22,670	0.714	0.143	1.429	0.204
3	0	0	0.000	0.000	0.000	0
4	2	7,230	1.429	0.286	0.456	0.130
5	3	15,986	2.143	0.429	1.008	0.432

Before selection, each male has frequency $1/5 = 0.20$, giving

$$\bar{W}_1 = \frac{1}{5} \left(1 + 1 + 0 + 2 + 3 \right) = \frac{7}{5} = 1.4, \quad \text{and } w_1 = \frac{W_1}{1.4}$$

While the *observed* frequencies of individuals have not changed after the first episode of selection (all are still present in the population), *fitness-weighted* frequencies change due to differences in acquiring mates. For male 2, $p_1(2) = 0.2 \cdot 0.714 = 0.143$ (e.g., 14.3% of all matings in the population involve male

2), with the values for the other adults being computed similarly. Hence, if we were to take offspring from these adults after this episode of selection, then for a randomly-drawn offspring, the probability that its parent is male 2 is 0.143. The mean fertility *per mating* is

$$\begin{aligned}\bar{W}_2 &= \sum W_2(r) \cdot p_1(r) \\ &= (25,820 \cdot 0.143) + (22,670 \cdot 0.143) + (7,230 \cdot 0.286) + (15,986 \cdot 0.429) \\ &= 15,860\end{aligned}$$

If each reproducing male were weighted equally, average fertility per individual (independent of the number of times each mates) is $(1/4) \cdot (25,820 + 22,670 + 7,230 + 15,986) = 17,927$. The actual mean fertility *per mating* \bar{W}_2 is lower because males 4 and 5 sired the most clutches, but had much lower fertility than the other (successful) males. Using $w_2(r) = W_2(r)/(15,860)$ and recalling Equation 19.2a, the final fitness weighting for male 1 is $(1/5) \cdot 0.714 \cdot 1.628 = 0.233$. The remaining p_2 values are computed similarly. Howard also measured body size (in mm). For males 1 through 5, body size was (respectively) 145, 128, 148, 138, and 141. Thus, before selection the mean and variance are $\bar{z}(0) = 140.0$ and $\text{Var}[z(0)] = 59.5$. From Equation 19.2b, the fitness-weighted mean following the first episode of selection is

$$\bar{z}(1) = 145 \cdot 0.143 + 128 \cdot 0.143 + 148 \cdot 0 + 138 \cdot 0.286 + 141 \cdot 0.429 = 138.996$$

Similarly,

$$\bar{z}^2(1) = 145^2 \cdot 0.143 + 128^2 \cdot 0.143 + 148^2 \cdot 0 + 138^2 \cdot 0.286 + 141^2 \cdot 0.429 = 19,325$$

Hence

$$\text{Var}[z(1)] = \frac{5}{4}(19,325 - 138.996^2) = 6.39$$

Thus, if we again take a hypothetical offspring following the first episode of selection, the mean character value of the adult that produced this randomly-chosen offspring is 138.996. Likewise, $\bar{z}(2) = 138.8$ and $\text{Var}[z(2)] = 67.7$.

VARIANCE IN INDIVIDUAL FITNESS

How do we compare the amount of selection acting on different populations? At first thought, one might consider using the standardized selection differential (the selection intensity) \bar{r} for comparing the relative strength of individual selection between populations. The drawback with \bar{r} as a measure of *overall* selection on individuals is that it is *character specific*. Hence, \bar{r} is appropriate if we are interesting in comparing the strength of selection on a particular *character*, but inappropriate

if we wish to compare the overall strength of selection on *individuals*. Two populations may have the same \bar{w} value for a given character, but if that character is tightly correlated with fitness in one population and only weakly correlated in the other, selection is much stronger in the latter population. Further, considerable selection can occur without changing the mean (e.g., stabilizing selection). Standardized differentials also exist for the variance, but the problem of character-specificity still remains.

A much cleaner measure (independent of the characters under selection), is I , the **opportunity for selection**, defined as the variance in *relative* fitness:

$$I = \sigma_w^2 = \frac{\sigma_W^2}{\bar{W}^2} \quad (19.3)$$

This measure was introduced by Crow (1958, reviewed in 1989), who referred to it as the **Index of Total Selection** and was independently developed by O'Donald (1970). I is estimated by

$$\hat{I} = \text{Var}(w) = \frac{n}{n-1} (\overline{w^2} - 1) \quad (19.4)$$

Crow noted that if fitness is perfectly heritable (e.g., $h^2(\text{fitness}) = 1$), then $I = \Delta\bar{w}$, the scaled change in fitness. Following Arnold and Wade (1984a,b) we call I the opportunity for selection, as any change in the distribution of fitness caused by selection represents an opportunity for within-generation change. The opportunity for selection bounds the maximum value of \bar{w} . This follows by using (respectively), the definition of a correlation, the covariance definition of s , and the fact that $|\rho| < 1$, to give

$$|\rho_{z,w}| = \frac{|\sigma_{z,w}|}{\sigma_z \sigma_w} = \frac{|s|}{\sigma_z \sqrt{I}} \leq 1,$$

implying

$$|\bar{w}| \leq \sqrt{I} \quad (19.5)$$

Thus, the most that any mean can be shifted within a generation is \sqrt{I} phenotypic standard deviations.

Example 2. To estimate I for the bullfrog data used in Example 1, we first compute the lifetime relative fitnesses as $w = w_1 \cdot w_2$, which gives relative lifetime fitnesses of 1.162, 1.020, 0, 0.652, and 2.160. Thus

$$\overline{w^2} = (1/5) [1.162^2 + 1.020^2 + 0^2 + 0.652^2 + 2.160^2] = 1.496$$

giving

$$\hat{I} = \frac{5}{4}(1.496 - 1) = 0.62$$

Hence, the most selection can change the mean of any character within a generation is $\sqrt{I} \simeq 0.79$ standard deviations. The observed change in male body size (in standard deviations) from Example 1 is $(138.8 - 140)/\sqrt{59.5} = -0.155$, less than one-fifth of the maximum absolute change of 0.79.

The usefulness of I as a bound of \bar{v} depends on the correlation between relative fitness and the character being considered. Figure 19.1 shows scatterplots of relative fitness versus two characters (z_1 and z_2) measured in the same set of individuals. The marginal distributions of fitness are identical for both characters (since the same set of individuals was measured), and hence both have the same opportunity for selection. The association between relative fitness and z_1 is fairly strong, while there is no relationship between relative fitness and z_2 , so that z_1 realizes much, while z_2 realizes none, of the opportunity for change.

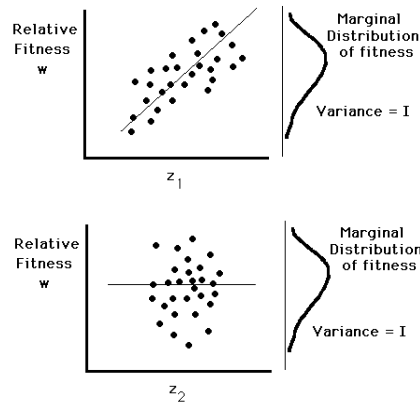


Figure 19.1. The fraction of the opportunity for selection I that is translated into a change in the mean depends on the correlation between relative fitness and the character. Characters z_1 and z_2 have the same marginal distribution of fitness, but only the regression of w on z_1 is significant. Thus (within a generation) selection changes the mean of z_1 but not z_2 .

In many cases individual fitnesses are not recorded. Rather, the average fitness for a number of phenotypic classes are estimated. In such cases, we can still obtain a lower bound for I , as the following example illustrates.

Example 3. O'Donald (1970, 1971b) analyzed the data of Dowdeswell (1961) who looked for selection on eyespot number on the hindwing of the European butterfly *Maniola jurtina* (see Brakefield 1984 for a recent review of the biology of this character). Dowdeswell compared the population distribution of eyespot number between a series of wild-collected females and a series of females reared from larvae. Presumably, the difference in distributions was due to selection on adults. By comparing the relative eyespot frequencies in reared and wild adults, fitnesses for each phenotypic class were estimated (see O'Donald 1971b for details). These are given in the table, along with the number of adults in each size class from the sample of 237 reared females (presumably representing the population before selection).

Eyespot number	Fitness	Number
0	1.000	124
1	0.699	67
2	0.657	34
3	0.548	10
4	0.000	2

$$\bar{W} = \frac{1}{237} [(1 \cdot 124) + (0.699 \cdot 67) + (0.675 \cdot 34) + (0.548 \cdot 10)] \simeq 0.838$$

$$\overline{W^2} = \frac{1}{237} [(1^2 \cdot 124) + (0.699^2 \cdot 67) + (0.675^2 \cdot 34) + (0.548^2 \cdot 10)] \simeq 0.736$$

Thus

$$\text{Var}(W) = \frac{237}{236} (0.736 - 0.838^2) \simeq 0.034$$

and

$$\hat{I} = \frac{0.034}{0.838^2} \simeq 0.048$$

This is an *underestimate*, as to properly estimate I the distribution of *individual* fitnesses, rather than mean fitness for each phenotypic class (*character* fitnesses), is required. These data only allow us to estimate the between-group variance in fitness (the variance in average fitness for the different eyespot classes) and hence this gives an *underestimate* of I because it neglects within-group variance (the variance in fitness among individuals with the same number of eyespots).

Partitioning I Across Episodes of Selection

The total opportunity for selection can be partitioned into opportunities associated with each episode. Such a partitioning allows the relative strength of selection to be compared across episodes as well as bounding the change in means and variances due to selection during an episode. Denote the opportunity of selection

associated with the j th episode by $I(j)$. By analogy with the definition of I , Arnold and Wade (1984a) suggest that the appropriate definition is the variance in the relative fitnesses of the j th fitness component:

$$I(j) = \sigma^2(w_j) = E(w_j^2) - 1 \tag{19.6}$$

which is estimated by

$$\begin{aligned} \widehat{I}(j) &= \text{Var}(w_j) = \frac{n}{n-1} (\overline{w_j^2} - 1) \\ &= \frac{n}{n-1} \left(\sum_r w_j^2(r) p_{j-1}(r) - 1 \right) \end{aligned} \tag{19.7}$$

Arnold and Wade show that the partition for I over k episodes of selection is given by

$$I = \sum_{j=1}^k I(j) + R \tag{19.8}$$

where the remainder term, R , represents a complex sum of covariances between fitness components (see Arnold and Wade 1984a for details).

Example 4. Compute the estimates of $I(1)$ and $I(2)$ using the data from Example 1. Using the relative fitnesses given in the table, $p_0(r) = 1/5$ giving

$$\overline{w_1^2} = \frac{1}{5} (0.714^2 + 0.714^2 + 0^2 + 1.429^2 + 2.143^2) \simeq 1.531$$

Likewise

$$\begin{aligned} \overline{w_2^2} &= \sum_r w_2^2(r) p_1(r) = 1.628^2 \cdot 0.143 + 1.429^2 \cdot 0.143 \\ &\quad + 0.456^2 \cdot 0.286 + 1.008^2 \cdot 0.429 \simeq 1.165 \end{aligned}$$

Hence

$$\widehat{I}(1) = \frac{5}{4} (1.531 - 1) \simeq 0.664$$

and

$$\widehat{I}(2) = \frac{5}{4} (1.165 - 1) \simeq 0.206$$

From Example 2, $\widehat{I} = 0.62$. Observe that $\widehat{I}(1) + \widehat{I}(2) = 0.87 \neq \widehat{I}$. From Equation 19.8, $\widehat{R} = 0.62 - 0.87 = -0.25$, reflecting the strong negative covariance within

individuals between the first and second fitness components (in this data set, individuals with high W_1 tend to have a low W_2 and vice versa).

Some Caveats in Using the Opportunity for Selection

There are some subtle issues in the interpretation of I . To begin with, even though I appears to remove scaling effects due to different types of fitnesses, for estimates of I to be truly comparable, they must be based on consistent measures of fitness (Triall 1985). Consider the opportunity for selection based on number of mates per male (sexual selection) I_s versus the opportunity for selection based on total male reproductive success (the number of offspring per male) I_{rs} . Total male reproductive success depends on both number of mates and fertility per mating. Recalling Equation 19.8, $I_{rs} = I_s + I_f + R$, where I_f is the opportunity for selection based on differences in male fertility per mating. Hence, I_{rs} is expected to exceed I_s unless there is sufficient negative covariance between the mating success and fertility components ($R < -I_f$).

A second point is that if the variance in fitness is not independent of \bar{W} , comparisons of I values between populations are compromised. This occurs in cross-sectional studies that measure sexual selection by simply counting the number of mating pairs (in such studies, an unequal sex ratio further biases comparisons of I between the sexes). If the time scale is such that only single matings are observed, the fitness of any individual is either 1 (mating) or 0 (not mating). The resulting fitness of randomly-drawn individuals is binomially distributed with mean p (the mean copulatory success for the sex being considered) and variance $p(1-p)$, hence

$$I = \frac{p(1-p)}{p^2} \simeq \frac{1}{p} \quad \text{if } p \ll 1 \quad (19.9)$$

The mean and variance in individual fitness are not independent, and the opportunity for selection depends entirely on mean population fitness. As the time window for observing mating pairs decreases, fewer matings are seen and p decreases, increasing I . As the data plotted in Figure 19.2 illustrate, the opportunity for selection is often inflated if the observation period is short relative to the total mating period.

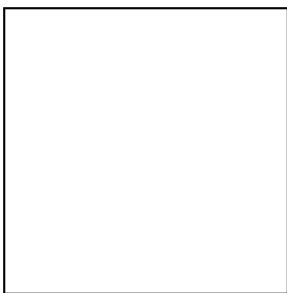


Figure 19.2. The ratio of the opportunity for selection on reproductive success to the lifetime opportunity for the coreid bug *Colpula lativentris* as a function of observation period. After Nishida (1989).

A second example of the lack of independence between \bar{W} and σ_W^2 was given by Downhower et al. (1987). Assuming that the number of mates for any given male follows a Poisson distribution, the variance in number of mates equals the mean number of mates, giving

$$I = \frac{\bar{W}}{\bar{W}^2} = \bar{W}^{-1}$$

where \bar{W} is the mean number of mates per male. Thus, differences in I between populations do not necessarily indicate *biological* differences in male mating ability. For example, in a population of 100 males, if only 5 females mate, average male mating success is $\bar{W} = 0.05$, while if 50 females mate, $\bar{W} = 0.5$. For this example, differences in I come solely from variation in the number of mating females, not biological differences between males in their ability to acquire mates. Downhower et al. conclude from this example that comparing I values with the Poisson prediction ($I = \bar{W}^{-1}$) or some other appropriate random distribution may help clarify the interpretation of I . For this case, values of I less than the Poisson prediction indicate a more uniform distribution of fitness than expected if mate choice is random, while values in excess of this expectation indicate disproportionately high fitness among a limited set of individuals.

This comparison of I to the value expected under a Poisson distribution of individual fitness is an attempt to account for differences in opportunities for selection due to differences in mean fitness. In effect, this is a problem of stabilizing the variances (Chapter 10). Since I is the squared coefficient of variation in fitness, it is plagued by the same statistical problems as the Roginskii-Yablokov effect — even if $\sigma^2(W)$ and \bar{W} are independent, recall from Chapter 10 that a negative correlation is often expected between x/y and y even when x and y are independent. Thus, in most cases we expect I to be somewhat dependent on \bar{W} .

The Poisson mating example further points out that random variation (differences in individual fitness not attributable to intrinsic differences between individuals) reduces the correlation between phenotypic value and relative fitness. For this reason, measures of selection based entirely upon variance in mating success have been criticized (Banks and Thompson 1985, Sutherland 1985a,b, Koenig and Albano 1986). Although carefully controlled studies can reduce the error variance induced by chance (e.g., Houck et al. 1985), accounting for inflation of the opportunity for selection by random effects remains a problem.

DESCRIBING PHENOTYPIC SELECTION: INTRODUCTORY REMARKS

Most of the previous discussion in this chapter dealt with the fitness of individuals, independent of any knowledge of their phenotypes. Selection can favor certain phenotypes, leading to our second major topic — how do particular character values influence the fitness of an individual? Our interest in a particular character may be in predicting how selection changes it over time, which requires knowledge of the genetics of that character. Alternately, we may simply wish to explore the ecological implications of a character by examining how expected fitness changes with character value. While only the relationship of total fitness to the character is needed to describe the evolutionary response to selection, partitioning fitness across episodes of selection can enhance our understanding of the ecology of that character.

One general way of detecting selection on a character is to compare the (fitness-weighted) phenotypic distribution before and after some episode of selection. One important caveat is that growth or other ontogenetic changes, immigration, and environmental changes can also change the phenotypic distribution. We must take great care to account for these factors. Typically, selection on a character is measured by considering changes in the mean and variance, rather than changes in the entire distribution. Indeed, as was discussed in Chapters 24 and 26, in many cases the entire selection response following a generation of selection can be reasonably predicted from the within-generation change in the mean and variance. As we will see below, there are a number of subtle issues in assigning fitnesses to phenotypes. A thoughtful review of these is given by Grafen (1988). Again, we remind the reader of the most important caveat, that of dealing with the effects of selection on phenotypically-correlated characters, which is considered in Chapter 32.

Fitness Surfaces

$W(z)$, the expected fitness of an individual with phenotype z , describes a **fitness surface** (or equivalently a **fitness function**), relating fitness and character value. The *relative* fitness surface $w(z) = W(z)/\bar{W}$ is often more convenient than $W(z)$, and we use these two somewhat interchangeably. The nature of selection on a

character in a particular population is determined by the local geometry of the individual fitness surface over that part of the surface spanned by the population (Figure 19.3). If fitness is increasing (decreasing) over some range of phenotypes, a population having its mean value in this interval experiences **directional selection**. If $W(z)$ contains a local maximum, a population with members within that interval experiences **stabilizing selection**. If the population is distributed around a local minimum, **disruptive selection** occurs. As is illustrated in Figure 19.3D, when the local geometry of the fitness surface is complicated (e.g., multimodal) the simplicity of description offered by these three types of selection breaks down, as the population can experience all three simultaneously.

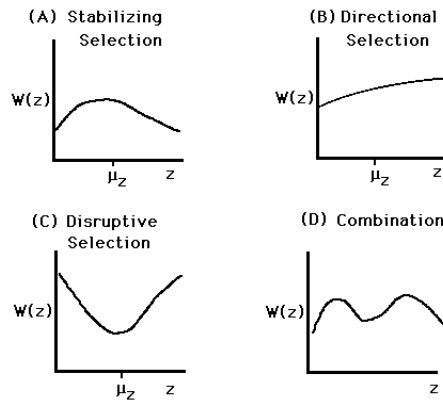


Figure 19.3. Selection is usually classified into three basic forms depending on the local geometry of $W(z)$: stabilizing (A), directional (B), and disruptive (C). As D illustrates, populations can simultaneously experience multiple forms of selection.

$W(z)$ may vary with genotypic and environmental backgrounds. In some situations (e.g., predators with search images, sexual selection, dominance hierarchies, truncation selection) the fitness of a phenotype depends on the frequency of other phenotypes in the population. In this case, fitnesses are said to be **frequency-dependent**.

Mean population fitness \bar{W} is also a fitness surface, describing the expected fitness of the population as a function of the distribution of phenotypes in that population,

$$\bar{W} = \int W(z) p(z) dz$$

Hence, mean fitness can be thought of as a function of the parameters of the phenotypic distribution. For example, if z is normally distributed, mean fitness

is a function of the mean μ_z and variance σ_z^2 for that population.

To stress the distinction between the $W(z)$ and \bar{W} fitness surfaces, the former is referred to as the **individual fitness surface**, latter as the **mean fitness surface**. Knowing the individual fitness surface allows us to compute the mean fitness surface for any specified $p(z)$ but the converse is not true. The importance of the mean fitness surface is that it provides one way of describing how the population changes under selection. When the breeders' equation holds, the first two partial derivatives of \bar{W} with respect to μ_z describes the change in mean and variance (Equations 25.11 and 31.16). More generally, partials of \bar{W} with respect to higher phenotypic moments describe the dynamics of selection in the Barton-Turelli response equations (e.g., Equations 27.29 and 27.33). Mean fitness surfaces are considerably smoother than the individual fitness surfaces that generate them (Figure 19.4). The individual fitness surface may have discontinuities and rough spots — regions where very small changes in phenotypic value result in large changes in individual fitness. Mean population fitness averages over $W(z)$, smoothing out these rough spots. This smoothing facilitates the existence the various partials of mean fitness used in the Barton-Turelli equations.

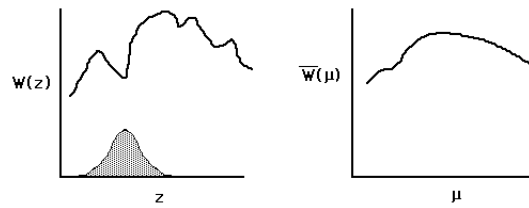


Figure 19.4. In this example, a small change in z can result in a large change in the individual fitness surface $W(z)$. However, since the mean population fitness $\bar{W}(\mu_z)$ averages individual fitnesses over the phenotypic distribution, shown as the stippled curve, small changes in μ_z result in only small changes in $\bar{W}(\mu_z)$.

DESCRIBING PHENOTYPIC SELECTION: CHANGES IN PHENOTYPIC MOMENTS

Selection for particular phenotypes changes in the phenotypic distribution (although it need not change all moments, for example, the mean may be unchanged). Thus, selection is detected by testing for differences between the distribution of phenotypes before and after some episode of selection. Nonparametric tests such as the Kolmogorov-Smirnov test (Sokal and Rohlf 1981) have the advantage of making no assumptions about the form of the distribution, but suffer from

low power. While complete distributions can be compared, the most common procedure for detecting selection is to test for changes in phenotypic moments. Standard statistical tests for differences in means (t -tests) and variances (F -tests) can be used, but these tests rely on normality assumptions that are often violated, and nonparametric tests are often more appropriate. Differences in means can be tested using the Wilcoxon-Mann-Whitney test, while Conover's squared rank test (Conover 1980) can be used to test for changes in variances. Other nonparametric tests for changes in variance exist, but care must be exercised, as some (e.g., the Siegel-Tukey test) are quite sensitive to differences in means; see Sprent (1989) and Conover (1980). While these issues are important, the main problem in detecting selection on a character is that changes in the moments may be due entirely to selection on phenotypically correlated characters (Chapter 31). Keeping this important caveat in mind, we now examine measures of selection for single characters.

Three measures of the within-generation change in phenotypic mean have been previously introduced: the directional selection differential s , the standardized directional selection differential (selection intensity) \bar{i} , and the directional selection gradient β . These measures are interchangeable for selection acting on a single character (e.g., Equations 25.5, 25.9, 25.11). When multiple characters are considered, the multivariate extension of β is the measure of choice, as it measures the amount of selection on a character over and above that attributable to selection on any other phenotypically correlated traits under consideration, while s (and hence \bar{i}) confounds these direct and indirect effects (Equation 32.4).

Similar measures can be defined to quantify the change in variance. At first glance this seems best described by $\sigma_{z^*}^2 - \sigma_z^2$, where $\sigma_{z^*}^2$ is the phenotypic variance following selection. The problem with this measure is that directional selection reduces the variance. Lande and Arnold (1983) showed that

$$\sigma_{z^*}^2 - \sigma_z^2 = \sigma [w, (z - \mu_z)^2] - s^2 \quad (19.10)$$

(proved for the multivariate version in Example 1 of Chapter 31). Hence, directional selection decreases the phenotypic variance by s^2 . With this in mind, Lande and Arnold suggest a corrected measure, the **stabilizing selection differential**

$$C = \sigma_{z^*}^2 - \sigma_z^2 + s^2 \quad (19.11)$$

that describes selection acting directly on the variance. As we will see below, the term stabilizing selection differential may be slightly misleading, so following Phillips and Arnold (1989) we refer to C as the **quadratic selection differential**. Correction for the effects of directional selection is important, as claims of stabilizing selection based on a reduction in variance following selection can be due entirely to reduction in variance caused by directional selection. Similarly, disruptive selection can be masked by directional selection (e.g., Example 6). Provided that selection does not act on characters phenotypically correlated with the one

under study, C provides information on the nature of selection on the variance. Positive C indicates selection to increase the variance (as would occur with disruptive selection), while negative C indicates selection to reduce the variance (as would occur with stabilizing selection). As we discuss shortly, $C < 0$ ($C > 0$) is *consistent* with stabilizing (disruptive) selection, but not *sufficient*. A further complication in interpreting C is that if the phenotypic distribution is skewed, selection on the variance changes the mean (e.g., Equations 24.21b, 27.32a, 27.33). This causes a non-zero value of s that in turn inflates C (Figure 19.5).

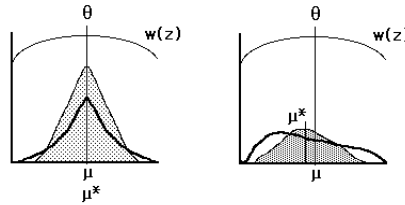


Figure 19.5. Even when a population is under strict stabilizing selection, the mean can change if the phenotypic distribution is skewed. A standard fitness function for stabilizing selection is $W(z) = 1 - b(\theta - z)^2$. O'Donald (1968) found that, even if the population mean is at the optimum value ($\mu_z = \theta$), s is nonnegative if the skew is nonzero ($\mu_{3,z} \neq 0$) as $s = -(b\mu_{3,z})/(1 - b\sigma_z^2)$. **Left:** If phenotypes are distributed symmetrically about the mean ($\mu_{3,z} = 0$), the distribution after selection (stippled) has the same mean when $\mu_z = \theta$. **Right:** If, however, the distribution is skewed, more of the distribution lies on one side of the mean than the other. Since the distribution is “unbalanced”, the longer tail experiences more selection than the shorter tail, changing the mean.

Analogous to s equaling the covariance between z and relative fitness, Equation 19.10 implies C is the covariance between relative fitness and the squared deviation of a character from its mean

$$C = \sigma [w, (z - \mu)^2] \quad (19.12)$$

As was the case with s , the opportunity for selection bounds the maximum possible within-generation change in variance (Arnold 1986). Expressing C as a covariance and using the definition of a correlation gives $C = \rho_{w, (z-\mu)^2} \sigma_w \sigma[(z - \mu)^2]$. Since $\rho_{w, (z-\mu)^2}^2 \leq 1$, we have

$$C^2 \leq \sigma_w^2 \sigma^2 [(z - \mu)^2] = I \cdot (\mu_{4,z} - \sigma_z^4)$$

The last equality follows from the definition of a variance, $\sigma^2 [(z - \mu)^2] = E[(z - \mu)^4] - E[(z - \mu)^2]^2$. Thus,

$$|C| \leq \sqrt{I (\mu_{4,z} - \sigma_z^4)} \quad (19.13a)$$

If z is normally distributed, $\mu_{4,z} = 3\sigma_z^4$ (Kendall and Stewart 1977), giving

$$|C| \leq \sigma_z^2 \sqrt{2I} \quad (19.13b)$$

The quadratic analogue of β , the **quadratic (stabilizing) selection gradient** γ , was suggested by Lande and Arnold (1983),

$$\gamma = \frac{\sigma [w, (z - \mu)^2]}{\sigma_z^4} = \frac{C}{\sigma_z^4} \quad (19.14)$$

A conceptual advantage of β and γ is that they describe the average local geometry of the fitness surface when phenotypes are normally distributed. When z is normal and individual fitness are not frequency-dependent, Equation 25.12 implies that β can be expressed in terms of the geometry of the *mean* fitness surface, $\beta = \partial \ln \bar{W} / \partial \mu_z = \bar{W}^{-1} \partial \bar{W} / \partial \mu_z$, the slope of the \bar{W} surface with respect to population mean (proved for the multivariate version Example 11 of Chapter 30). β can also be expressed as a function of the *individual* fitness surface. Lande and Arnold (1983) showed, provided z is normally distributed, that

$$\beta = \int \frac{\partial w(z)}{\partial z} p(z) dz \quad (19.15a)$$

implying that β is the average slope of the individual fitness surface, the average being taken over the population being studied (the multivariate version is proved in Chapter 31). Likewise, if z is normal,

$$\gamma = \int \frac{\partial^2 w(z)}{\partial z^2} p(z) dz \quad (19.15b)$$

which is the average curvature of the individual fitness surface (Lande and Arnold 1983). Thus, β and γ provide a measure of the geometry of the individual fitness surface averaged over the population being considered.

A final advantage of β and γ is that they appear as the only measure of phenotypic selection in equations describing selection response. We have already seen (Equation 25.11) that under the constraints of the breeders' equation, $\Delta\mu = \sigma_A^2 \beta$, which is independent of any other measure of the phenotype (note that σ_z^2 is missing). Similarly, for predicting changes in variance under the infinitesimal model, from Equation 27.7b the expected change in variance from a single generation of selection is

$$\begin{aligned} \Delta\sigma_z^2 &= \frac{h^4}{2} \delta\sigma_z^2 = \frac{\sigma_A^4}{2\sigma_z^4} (C - s^2) \\ &= \frac{\sigma_A^4}{2} (\gamma - \beta^2) \end{aligned} \quad (19.16)$$

which decomposes the change in variance into changes due to selection on the variance and changes due to directional selection. Note that, when expressed in terms of gradients, the phenotypic variance term vanishes.

While the distinction between differentials and gradients seems almost trivial in the univariate case (only a scale difference), the multivariate versions are considerably different. As we will see in Chapter 31, gradients have the extremely important feature of removing the effects of phenotypic correlations.

Partitioning Changes in Means and Variances into Episodes of Selection

Suppose the total amount of within-generation selection is partitioned into k episodes of selection. Let $\mu_{z(j)}$ and $\sigma_{z(j)}^2$ be the (fitness-weighted) mean and variance after the j th episode of selection ($\mu_z = \mu_{z(0)}$ and $\sigma_z^2 = \sigma_{z(0)}^2$ are the mean and variance before the first measured episode of selection).

The definitions of s , C , β , and γ suggest (Arnold and Wade 1984a) that appropriate measures for the j th episode of selection are given by

$$s_{(j)} = \mu_{z(j)} - \mu_{z(j-1)} \quad (19.17a)$$

$$C_{(j)} = \sigma_{z(j)}^2 - \sigma_{z(j-1)}^2 + s_{(j)}^2 \quad (19.17b)$$

$$\beta_{(j)} = \frac{s_{(j)}}{\sigma_{z(j-1)}^2} \quad (19.17c)$$

$$\gamma_{(j)} = \frac{C_{(j)}}{\sigma_{z(j-1)}^4} \quad (19.17d)$$

The properties for s and C hold for episodes of selection. Thus,

$$s_{(j)} = \sigma(w_j, z) \quad \text{and} \quad C_{(j)} = \sigma[w_j, (z - \mu_{z(j-1)})^2]$$

where w_j is the fitness for the j th episode of selection. Likewise, substituting $I(j)$ for I in Equations 19.5 and 19.13b bounds both $s_{(j)}$ and $C_{(j)}$.

Example 5. For Example 1, following male body size over two episodes of selection we find that $\hat{s}_{(1)} = \bar{z}(1) - \bar{z}(0) = -1$, $\hat{s}_{(2)} = -0.2$ and $\hat{s} = -1.2$. Likewise,

$$\hat{C}_{(1)} = \text{Var}[z(1)] - \text{Var}[z(0)] + \hat{s}_{(1)}^2 = 6.4 - 59.5 + (-1)^2 \simeq -52.1$$

Similar calculations yield $\hat{C}_{(2)} \simeq 61.3$ and $\hat{C} \simeq 9.64$. Based on this limited data set, there appears to be directional selection to reduce body size during both episodes. In addition, there is selection to reduce the variance in body size during the first episode (mate choice) countered by selection to increase this variance during the second episode (fertility per mating).

How do these individual episode measures relate to the total measure over all episodes? The partitions for s and C are additive, with

$$s = \sum_{j=1}^k s_{(j)} \quad (19.18a)$$

and

$$C = \sum_{j=1}^k C_{(j)} + \left(s^2 - \sum_{j=1}^k s_{(j)}^2 \right) \quad (19.18b)$$

Partitioning β and γ requires a little more care, as we have to account for changes in the phenotypic variance following each episode. Rewriting $s_{(j)} = \sigma_{z(j-1)}^2 \beta_{(j)}$ gives

$$\beta = \sum_{j=1}^k \beta_{(j)} \left[\frac{\sigma_{z(j-1)}^2}{\sigma_{z(0)}^2} \right] \quad (19.18c)$$

Thus, the total selection gradient is a *weighted* sum of the individual gradients associated with each episode. In their original paper, Arnold and Wade (1984a) stated that the total selection gradient is the sum of gradients associated with each selective episode, but the above expression shows this holds only when selection does not change the phenotypic variance (Kalisz 1986, Wade and Kalisz 1989). The partition of γ follows similarly,

$$\gamma = \sum_{j=1}^k \gamma_{(j)} \left[\frac{\sigma_{z(j-1)}^4}{\sigma_{z(0)}^4} \right] + \frac{1}{\sigma_{z(0)}^4} \left[s^2 - \sum_{j=1}^k s_{(j)}^2 \right] \quad (19.18d)$$

Standard Errors for Estimates of Differentials and Gradients

Since it is difficult to measure all individuals in a population, the effects of selection are usually estimated from a sample. Even with a longitudinal study, the cohort being followed is usually viewed as a representative sample of phenotypes from the population. This is not always the case — e.g., a cohort may be chosen intentionally to include the most extreme phenotypes at much higher frequencies than they are found in the population.

There are a number of statistical issues in extrapolating from these samples to the entire population, many of which still are unresolved. For example, individual fitness usually is measured with error. There is generally a bias to *underestimate* individual fitness — marked individuals may not be recaptured and hence recorded as having zero fitness, and the number of mates and/or offspring can be easily underestimated.

Assuming individual fitness is measured without error, the methods of Appendix 1 can be used to obtain large sample variances for estimators of differentials and gradients. The sampling variance for the directional selection differential is

$$\sigma^2(\hat{s}_{(j)}) = \frac{\sigma_{z^{(j)}}^2}{n_j} + \frac{\sigma_{z^{(j-1)}}^2}{n_{j-1}} \quad (19.19a)$$

where n_j is the sample size for the j th episode. From Example 1 in Appendix 1, the large-sample variance is approximately

$$\begin{aligned} \sigma^2(\hat{C}_{(j)}) &\simeq 4s_{(j)}^2 \sigma^2(\hat{s}_{(j)}) + 8s_{(j)} \left(\frac{\mu_{3,z^{(j)}}}{n_j} + \frac{\mu_{3,z^{(j-1)}}}{n_{j-1}} \right) \\ &\quad + \frac{\mu_{4,z^{(j)}} - \sigma_{z^{(j)}}^4}{n_j} + \frac{\mu_{4,z^{(j-1)}} - \sigma_{z^{(j-1)}}^4}{n_{j-1}} \end{aligned} \quad (19.19b)$$

If phenotypes are normally distributed, this reduces to

$$\sigma^2(\hat{C}_{(j)}) \simeq 4s_{(j)}^2 \sigma^2(\hat{s}_{(j)}) + 2 \left[\frac{\sigma_{z^{(j)}}^4}{n_j} + \frac{\sigma_{z^{(j-1)}}^4}{n_{j-1}} \right] \quad (19.19c)$$

If the scaled skewness k_3 (Equation 2.8) and kurtosis k_4 (Equation 2.13) are small, this normal approximation can be used. These results assume that the fitness-weighted distributions of phenotypes in episodes j and $j - 1$ are independent. If the same individuals are followed and the character measured only once, any measurement error in z for an individual carries over to all episodes, creating a correlation between episodes.

Example 6. Boag and Grant (1981) observed intense natural selection in *Geospiza fortis* (Darwin's medium ground finch) during a severe drought on Daphne Major Island in the Galápagos. The estimated mean weight and variance of 642 adults before the drought were respectively, 15.79 and 2.37, while the estimated mean and variance of 85 surviving adults after the drought was 16.85 and 2.43. Thus $\hat{s} = 16.85 - 15.79 = 1.06$ and Equation 19.19a gives the standard deviation of this estimate as

$$\text{SE}(\hat{s}) \simeq \sqrt{\frac{2.37}{642} + \frac{2.43}{85}} \simeq \sqrt{0.0323} \simeq 0.180$$

implying that the directional selection differential on body size is significantly positive. There appears to be very little selection on the variance when the uncorrected change in variance $\text{Var}(z^*) - \text{Var}(z) = 2.43 - 2.37 = 0.06$ is used. However,

$$\hat{C} = 0.06 + 1.06^2 = 1.14$$

consistent with disruptive selection in addition to directional selection. From Equation 19.19c, assuming body size is normally distributed before and after the drought,

$$\text{SE}(\hat{C}) \simeq \sqrt{4 \cdot (1.06)^2 \cdot 0.0323 + 2 \left[\frac{(2.37)^2}{642} + \frac{(2.43)^2}{85} \right]} \simeq 0.549$$

\hat{C} is 2.08 standard errors above zero, suggesting that it is significant.

As is discussed below, β and γ can be estimated from the coefficients of the linear and quadratic regressions (respectively) of relative fitness on phenotypic value. An advantage of this approach is that powerful resampling methods such as the jackknife method can be used to estimate approximate confidence intervals (e.g., Mitchell-Olds and Shaw 1987, Mitchell-Olds and Bergelson 1990b), and randomization tests can be used to test for significance (e.g., Moore 1990, Hews 1990). These procedures are very insensitive to the exact shape of the phenotypic distribution. The most significant advantage of using regressions is that this approach is easily extended to multiple characters, which removes the confounding effects of phenotypic correlations between measured characters (Chapter 31).

DESCRIBING PHENOTYPIC SELECTION: INDIVIDUAL FITNESS SURFACES

We can decompose the fitness W of an individual with character value z into the sum of its expected fitness $W(z)$ plus a residual deviation e ,

$$W = W(z) + e$$

The residual variance for a given z , $\sigma_e^2(z)$, measures the variance in fitness among individuals with phenotypic value z . Estimation of the individual fitness surface is thus a generalized regression problem, the goal being to choose a candidate function for $W(z)$ that minimizes the average residual variance $E_z[\sigma_e^2(z)]$. Since the total variance in fitness σ_W^2 equals the sum of the within-group (phenotype) and between-group variance in fitness,

$$\frac{\sigma_W^2 - E_z[\sigma_e^2(z)]}{\sigma_W^2}$$

is the fraction of individual fitness variation accounted for by a particular estimate of $W(z)$, and this provides a measure for comparing different estimates. In the

limiting case where fitness is independent of z (and any characters phenotypically correlated with z), $W(z) = \bar{W}$, so that the between-phenotypic variance is zero while $\sigma_e^2(z) = \sigma_W^2$.

There are at least two sources of error contributing to e . First, there can be errors in measuring the actual fitness of an individual (these are almost always ignored). Second, the *actual* fitness of a particular individual can deviate considerably from the *expected value* for its phenotype due to chance effects and selection on other characters besides those being considered. Generally, these residual deviations are heteroscedastic (Mitchell-Olds and Shaw 1987, Schluter 1988). To see how this arises naturally, suppose fitness is measured by survival to a particular age. While $W(z) = p_z$ is the probability of survival for an individual with character value z , the fitness for a particular individual is either 0 (does not survive) or 1 (survives). Thus the residual has only two possible values, $e = 1 - p_z$ with probability p_z and $e = -p_z$ with probability $1 - p_z$, giving $\sigma_e^2(z) = p_z(1 - p_z)$. Unless p_z is constant over z , the residuals are heteroscedastic. Note in this case that even after removing the effects attributable to differences in phenotypes, there still is substantial variance in individual fitness.

Inferences about the individual fitness surface are limited by the range of phenotypes in the population. Unless this range is very large, only a small region of the fitness surface can be estimated with any precision. Estimates of the fitness surface at the tails of the current phenotypic distribution are extremely imprecise, yet potentially very informative, suggesting what selection pressures populations at the margin of the observed range of phenotypes may be under. A further complication is that the fitness surface changes as the environment changes so that year to year estimates can differ (e.g., Kalisz 1986) and cannot be lumped together to increase sample size.

Linear and Quadratic Approximations of $W(z)$

The individual fitness surface $W(z)$ can be very complex and a wide variety of functions may be chosen to approximate it. The simplest and most straightforward approach is to use a low-order polynomial (typically linear or quadratic).

Consider first the simple linear regression of *relative* fitness w as a function of phenotypic value z . Since the directional selection gradient $\beta = s/\sigma_z^2 = \sigma(w, z)/\sigma_z^2$, it follows from regression theory (Equation 3.14b) that β is the slope of the least-squares linear regression of relative fitness on z ,

$$w = a + \beta z + e \quad (19.20a)$$

Hence the best linear predictor of relative fitness is $w(z) = a + \beta z$. Since the regression passes through the expected values of w and z , this can be written as

$$w = 1 + \beta(z - \mu_z) + e \quad (19.20b)$$

giving $w(z) = 1 + \beta(z - \mu_z)$. Assuming the fitness function is well described by a linear regression, β is the expected change in relative fitness given a unit change

in z . From Equation 3.17, the fraction of variance in individual fitness accounted for by this regression is

$$r_{z,w}^2 = \frac{\text{Cov}^2(z, w)}{\text{Var}(z) \cdot \text{Var}(w)} = \hat{\beta}^2 \frac{\text{Var}(z)}{\hat{I}} \tag{19.21}$$

If the fitness surface shows curvature, as might be expected if there is stabilizing selection and/or disruptive selection, a **quadratic regression** is more appropriate,

$$w = a + b_1z + b_2(z - \mu_z)^2 + e \tag{19.22a}$$

Since the regression passes through the mean of all variables, we can rewrite this as

$$w = 1 + b_1(z - \mu_z) + b_2 [(z - \mu_z)^2 - \sigma_z^2] + e \tag{19.22b}$$

The regression coefficients b_1 and b_2 nicely summarize the local geometry of the fitness surface. Evaluating the derivative of Equation 19.22 at $z = \mu_z$ gives

$$\left. \frac{\partial w(z)}{\partial z} \right|_{z=\mu_z} = b_1 \quad \text{and} \quad \left. \frac{\partial^2 w(z)}{\partial z^2} \right|_{z=\mu_z} = 2b_2 \tag{19.23}$$

Hence b_1 is the slope and $2b_2$ the second derivative (curvature) of the best quadratic fitness surface around the population mean. $b_2 > 0$ indicates that the best-fitting quadratic of the individual fitness surface has an upward curvature, while $b_2 < 0$ implies the curvature is downward. Lande and Arnold (1983) suggest that $b_2 > 0$ indicates disruptive selection, while $b_2 < 0$ indicates stabilizing selection. Their reasoning follows from elementary geometry in that a *necessary* condition for a local minimum is that a function curves upward in some interval, while a necessary condition for a local maximum is that the function curves downward. Mitchell-Olds and Shaw (1987) and Schluter (1988) argue that this condition is not *sufficient*. Stabilizing selection is generally defined as the presence of a local maximum in $w(z)$ and disruptive selection by the presence of a local minimum, while b_2 indicates curvature, rather than the presence of local extrema. As Figure 19.6 shows, the fitness function can curve downward without the population experiencing a local maximum or can curve upward without having a local minimum.

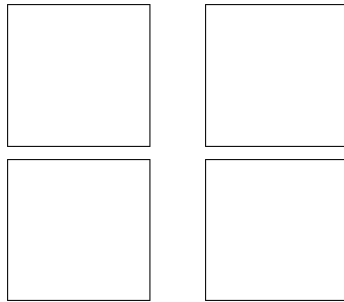


Figure 19.6. The relationship between γ and curvature of the fitness function (solid line). Dashed curve is the distribution of z . **A:** $W(z)$ is strictly linear, hence $\gamma = 0$. **B:** $W(z)$ curves downward, but has no maximum. Hence, $\gamma < 0$, implying stabilizing selection by the Lande-Arnold criterion, when in fact selection is entirely directional. **C:** Stabilizing selection. **D:** A combination of directional and stabilizing selection. From Mitchell-Olds and Shaw (1987).

We solve for the regression coefficients b_1 and b_2 by transforming Equation 19.22 into a standard multiple regression problem by setting $x_1 = z$ and $x_2 = (z - \mu_z)^2$ and applying the methods of Chapter 7. To proceed, we need expressions for $\sigma(x_1, x_2)$, $\sigma(x_1, w)$, and $\sigma(x_2, w)$. From Equation A1.14, $\sigma(x_1, x_2) = \sigma(z, (z - \mu_z)^2) = \mu_{3,z}$, the skew of the phenotypic distribution before selection. Likewise, from Equations 3.21 and 19.12, $\sigma(x_1, w) = \sigma(z, w) = s$ and $\sigma(x_2, w) = \sigma((z - \mu_z)^2, w) = C$. Substituting these into Equations 7.6 and 7.7, and noting that $\sigma^2((z - \mu_z)^2) = \mu_{4,z} - \sigma_z^4$ gives

$$b_1 = \frac{\sigma^2(x_2) \cdot \sigma(x_1, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{(\mu_{4,z} - \sigma_z^4) \cdot s - \mu_{3,z} \cdot C}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (19.24a)$$

$$b_2 = \frac{\sigma^2(x_1) \cdot \sigma(x_2, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{\sigma_z^2 \cdot C - \mu_{3,z} \cdot s}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (19.24b)$$

The estimators of b_1 and b_2 are obtained by replacing $\mu_{k,z}$ with their sample estimates and using \hat{C} and \hat{s} .

Provided z is normally distributed before selection, $\mu_{3,z} = 0$ and $\mu_{4,z} - \sigma_z^4 = 2\sigma_z^4$. In this case, Equations 25.11 and 19.14 imply, respectively, that $b_1 = \beta$ and $b_2 = \gamma/2$, giving the univariate version of the **Lande-Arnold regression**,

$$w = 1 + \beta(z - \mu_z) + \frac{\gamma}{2} \left((z - \mu_z)^2 - \sigma_z^2 \right) + e \quad (19.25)$$

developed by Lande and Arnold (1983), motivated by Pearson (1903). The Lande-Arnold regression thus provides a connection between selection differentials (directional and stabilizing) and quadratic approximations of the individual fitness surface.

An important point from Equation 19.24a is that if skew is present ($\mu_{3,z} \neq 0$), $b_1 \neq \beta$ and the slope term in the linear regression (the best *linear* fit) of $w(z)$ differs from the slope term in the quadratic regression (the best *quadratic* fit) of $w(z)$. This arises because the presence of skew generates a covariance between z and $(z - \mu_z)^2$. The biological significance of this can be seen by reconsidering Figure 19.5, wherein the presence of skew in the phenotypic distribution results in a change in the mean of a population under strict stabilizing selection (as measured by the population mean being at the optimum of the individual fitness surface). Skew generates a correlation between z and $(z - \mu_z)^2$ so that selection

acting only $(z - \mu_z)^2$ generates a correlated change in z . From the Robertson-Price identity (Equation 3.21), the within-generation change in mean equals the covariance between phenotypic value and relative fitness. Since covariances measure the amount of *linear* association between variables, in describing the change in mean, the correct measure is the slope of the best *linear* fit of the individual fitness surface. If skew is present, using b_1 from the quadratic regression to describe the change in mean is incorrect, as this quadratic regression removes the effects on relative fitness from a linear change in z due to the correlation between z and $(z - \mu_z)^2$.

Hypothesis Testing and Approximate Confidence Intervals

While there is a large body of theory for testing the significance of regression coefficients, much of it assumes homoscedastic and normally distributed residuals. As mentioned above, these two assumptions are almost always violated with fitness data, invalidating standard tests for significance found in most canned statistical programs (Mitchell-Olds and Shaw 1987). Fortunately, there are a variety of resampling methods available for hypothesis testing that are robust to heteroscedasticity and non-normal residuals, and we briefly mention three procedures (jackknife confidence intervals, randomization tests of significance, and cross-validation) here. Other approaches, such as probit analysis for 0/1 fitness data, are also available (Price 1984).

Jackknife estimates were introduced by Tukey (1958) as a generalized statistical tool. A nice introduction can be found in Sokal and Rohlf (1981, pp. 791-799), with a more detailed review in Miller (1974) and Wu (1986). The idea is simple: to base parameter estimates on the behavior of the estimate in subsamples of the original data. Consider the estimator of β for the linear regression given by Equation 19.20a. Denote by $\hat{\beta}$ the standard least-squares estimate of β using the full data set of n individuals, and let $\hat{\beta}_i$ denote the estimator using the complete data set minus data for the i th individual. The resulting jackknife estimator is

$$\hat{\beta}_{jack} = \frac{1}{n} \sum_{i=1}^n \phi_i = \bar{\phi} \quad \text{where} \quad \phi_i = n\hat{\beta} - (n-1)\hat{\beta}_i \quad (19.26a)$$

which has approximate large-sample variance

$$\text{Var}(\hat{\beta}_{jack}) \simeq \frac{1}{n(n-1)} \sum_{i=1}^n (\phi_i - \bar{\phi})^2 \quad (19.26b)$$

Approximate large-sample confidence intervals follow using Equation 19.26b and the fact that $\hat{\beta}_{jack}$ is approximately t -distributed with $n - 1$ degrees of freedom. The jackknife estimator and its sampling variance are well behaved even when the residuals are heteroscedastic, allowing for valid hypothesis testing (Wu 1986). Wu gives a slightly improved jackknifed estimator by weighting the ϕ_i values,

but the difference between the weighted and unweighted estimates is usually small for large sample sizes. Mitchell-Olds (1989) has developed a program for computing both the weighted and unweighted jackknife estimates for multiple character quadratic fitness regressions. Mitchell-Olds and Bergelson (1990b) use this method to examine the significance of several fitness regressions in the annual plant *Impatiens capensis*.

Randomization tests provide another approach for testing the significance of a regression. Again, the idea behind this class of tests (reviewed in Sokal and Rohlf 1981, pp. 787–795) is simple but computationally intensive. A particular value of $\hat{\beta}$ under the hypothesis of no association between fitness and z is generated by assigning the n individual fitnesses at random to the observed phenotypic values and estimating β for this scrambled (randomized) data set. By repeating this resampling procedure several hundred times we generate a distribution of regression coefficients under the hypothesis of no association between individual fitness and character value. Suppose we obtain a standard least-square estimate (assuming a linear regression) of $\hat{\beta} = 1.25$ and upon subsequent randomization of the same data set we find that only 7 out of 500 randomized data sets (1.4%) have $\hat{\beta}$ values in excess of 1.25. This suggests that this value is significantly different from zero at the 5%, but not 1% level. See Moore (1990) and Hews (1990) for applications of randomization tests to fitness data.

A final issue is assessing the validity of the particular model chosen to fit $W(z)$. This is a difficult task since by their nature fitness data are inherently noisy — the residual variance can be rather large, even if we have perfectly fit $W(z)$. One approach for checking model validity is **cross-validation** (Snee 1977, Picard and Cook 1984) wherein the original data are split into two samples at random. The fitness regression for the particular model being assumed is estimated using the first sample and the predictive ability of this model is then checked by seeing how well this regression predicts fitnesses in the second sample.

Schluter's Cubic-Spline Estimate

A serious problem with quadratic regressions as estimators of $W(z)$ is that the fitted curve allows for at most only a single local maximum or minimum. Fitness surfaces with multiple local maxima are thus very poorly described by a quadratic. Figure 19.7 gives a particularly illustrative example, showing that a quadratic fit to a truncation selection fitness function creates a spurious local minimum.

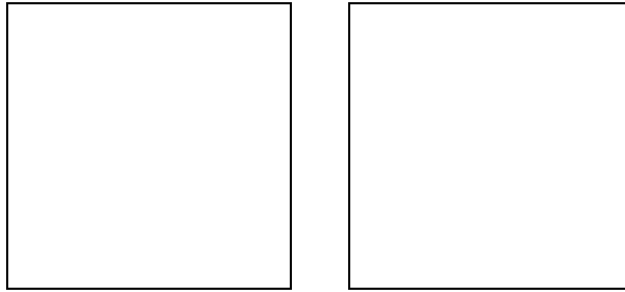


Figure 19.7. Examples of a misleading approximation of $W(z)$ resulting from using a quadratic regression. **Left.** A hypothetical example wherein phenotypes are normally distributed with only individuals exceeding one phenotypic standard deviation surviving. The best quadratic regression erroneously suggests the presence of disruptive selection (by introducing a false minimum), rather than the strict directional selection that is actually occurring. From Schluter (1988). **Right.** Data from Mitchell-Olds and Bergelson (1990b) on individual fitness as a function of the character $z =$ late growth rate for the annual plant *Impatiens capensis*. The data clearly depart from linearity, showing curvature. The best-fitting quadratic (plotted) indicates a minimum in fitness (disruptive selection) around $x \simeq -2.4$. However, assuming exponentially increasing fitness with $w(z) + 0.5 = \exp(0.52 + 0.46 \cdot z - 0.002 \cdot z^2)$ gives a better fit of the data, suggesting that strict directional selection is acting on z as this function monotonically increases over the character range measured. After Mitchell-Olds and Bergelson (1990b).

In order to more reliably estimate the fitness function, Schluter (1988) has developed a non-parametric method that makes no assumptions about the functional form of the fitness surface. Schluter's approach fits the data using a series of cubic splines (a series of cubic polynomials that join smoothly together) using a jackknife method as the "best fit" criterion. This requires assumptions about the distribution of the residuals e as a function of phenotypic value z . A program to estimate $W(z)$ assuming either normally, binomially, or Poisson distributed residuals is available from Schluter. Binomially distributed residuals arise naturally with survival data, while Poisson distributed residuals are a reasonable model for number of offspring or mates. Schluter's program also has a resampling procedure that generates rough confidence intervals on estimates of $W(z)$. Examples of fitness surfaces estimated using this approach are given in Figure 19.8. Interestingly, when one of one of the classic examples of stabilizing selection, the data of Karn and Penrose (1951) relating survival and human birth weight, is reanalyzed using Schluter's method, the local maximum is not significant. Parametric tests of the significance of estimated local maximum/minimum are discussed by Mitchell-Olds and Shaw (1987) for quadratic regressions, while nonparametric

tests are discussed by Schluter (1988).

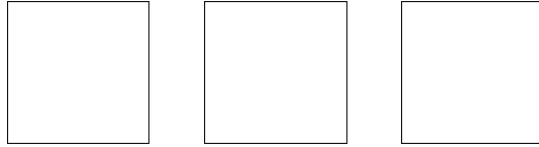


Figure 19.8. Examples of fitness surfaces generated using Schluter's method. The actual fitness values for individuals are indicated by +, the solid curve indicates the cubic-spline estimate of $W(z)$. **Left:** Probability of survival as a function of beak depth in Darwin's finch *Geospiza fortis*. The dashed curve indicates the estimate of the surface by a quadratic regression, which generates a spurious minimum. **Center:** Number of young produced as a function of tarsus length in song sparrows. **Right:** Survival of male human infants as a function of birth weight. From Schluter (1988).

The Importance of Experimental Manipulation

Several authors have stressed that regression approaches should be viewed as only the preliminary step in any analysis of the actual agents of selection, treating any regression estimates as an initial hypothesis to be further tested by experimental manipulation (Mitchell-Olds and Shaw 1987, 1990; Schluter 1988; Wade and Kalisz 1990). Spurious correlations between a character and fitness can be generated in a variety of ways: environmental correlations between character value and fitness, selection on unmeasured characters correlated with the observed character, and loci with direct fitness effects having pleiotropic effects on the character being measured. Recall Example 6 in Chapter 23, in which loci with overdominant effects on fitness also had additive effects on a character z not under selection. In this example, while there was a perfect linear regression of $W(z)$ on z with larger values having higher mean fitness, z declined as selection proceeds. In this case, selection was not acting on z , contrary to what we would surmise from a fitness regression. Likewise, when individuals in the population differ in amount of inbreeding (such as occurs in many plants), highly inbred individuals may suffer a reduction in fitness due to inbreeding depression. If the measured character being considered also suffers inbreeding depression, this generates a correlation between with fitness that is due entirely to levels of inbreeding rather than intrinsic differences between phenotypic values (Willis 1993).

Mitchell-Olds and Shaw (1987) and Wade and Kalisz (1990) suggest that interactions between environmental effects and fitness are extremely important. For example, Breden and Wade (1989) observed a positive relationship between group size and fitness in the imported willow leaf beetle. However, when preda-

tors were excluded, there was no relationship. Thus, in this case β is correlated with the environment (presence/absence of predators). A second (hypothetical) example is the case where individuals reared in higher-quality environments both obtain a larger size and also have more offspring than individuals from lower-quality environments. This generates a correlation between body size and fitness. However, it is the quality of the environment, not body size *per se*, that is the causal agent influencing the number of offspring in this case. Wade and Kalisz suggest computing fitness regressions in several different environments, and looking for correlations between β (and/or γ) and the environmental treatment. Such correlations strongly imply that the environmental character is a causal agent of selection.