28

Individual Fitness and the Measurement of Univariate Selection

Draft Version 14 November 2008

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 unrealistic situation offers the advantage of allowing basic methodological points to be stressed without the additional complications inherent in a multivariate analysis. In Chapter 29 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. Excellent discussions on the detection of selection in natural populations can be found in Endler (1986), Manly (1985), Primack and Kang (1989), Brodie et al. (1995), and Kingsolver and Pfenning (2007), while summaries of the estimates of selection values can be found in Endler (1986), Kingsolver et al. (2001), Hoekstra et al. (2001), Conner (2001), Hereford et al. (2004), and Kingsolver and Pfenning (2007). The use of individual fitness data is developed in Arnold and Wade (1984a,b) and Lande and Arnold (1983), whose approach we largely follow.

EPISODES OF SELECTION AND THE ASSIGNMENT OF FITNESS

Selection can often be subdivided into discrete 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, **more current references?**). 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

We start with two simplifying assumptions. The first is that generations are discreet and non-overlapping, so that the actual *timing* of reproduction is unimportant. Second, we assume that parental phenotypes (or more generally, phenotypes of other individuals that interact with our focal individual) have no influence on that individual's fitness. Some of consequences when these assumptions fail will be considered shortly.

Under these simplifying assumptions, the **lifetime** (or **total**) **fitness** of an individual is the number of descendants it leaves at the start of the next generation. Likewise, under these assumptions 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 can 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 accomplished (although see Chapter 16). 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. (**More recent references?**)

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? The nature of fitness components is often set by ecological or behavioral, rather than evolutionary, concerns. Ideally, components should be *sequential*, with one episode finishing before the next begins, or (as with the plant example) represent non-overlapping events. This, of course, need not be the case. For example, total number of mates (which is seemingly a measure of sexual selection) can be the result of (say) average number of mates per day times number of days lived. The former is a strict measure of sexual selection, the latter (viability) is a measure of natural selection, so that number of mates here is actually a compound measure of both natural and sexual selection.

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. As mentioned, age-structured populations pose further complications that will be considred shortly.

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 $1 \le r \le n$) is followed through several discrete (non-overlapping) episodes of selection. Let $W_j(r)$ be the fitness measure for the jth episode of selection for the rth individual. For example, if we are following viability then W_j is either zero (dead) or one (alive) at the census period. **Relative fitness** components $w_j(r) = W_j(r)/\overline{W}_j$ will turn out to be especially useful (as $\overline{w}_j = 1$). At the start of the study, the frequency of each individual is $1/n$, giving for the first (observed) episode of selection

$$\overline{W}_1 = \frac{1}{n}\sum_{r=1}^{n} W_1(r) \tag{28.1a}$$

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 rth individual is $w_1(r)/n$, implying that the mean fitness $\overline{W}_2$ for the second episode of selection is given by

$$\overline{W}_2 = \sum_{r=1}^{n} W_2(r)\cdot w_1(r)\cdot\left(\frac{1}{n}\right) \tag{28.1b}$$

In general, for the jth episode of selection,

$$\overline{W}_j = \sum_{r=1}^{n} W_j(r)\cdot w_{j-1}(r)\cdot w_{j-2}(r)\cdots w_1(r)\cdot\left(\frac{1}{n}\right) \tag{28.1c}$$

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

$$p_j(r) = w_j(r)\cdot p_{j-1}(r) = \frac{1}{n}\prod_{i=1}^{j} w_i(r) \tag{28.2a}$$

The first term $(1/n)$ is the initial frequency of individual r, while the product is the total relative fitness for individual r following the j episodes of selection. Thus, Equation 28.1c can also be expressed as $\overline{W}_j = \sum W_j(r)\cdot p_{j-1}(r)$. Using these weights allows fitness-weighted moments to be calculated, e.g., the mean of a particular character following the jth episode satisfies

$$\overline{z}_j = \sum z(r)\cdot p_j(r) \tag{28.2b}$$

where $z(r)$ is the value of the character of individual r. Likewise, the sample variance is

$$\mathrm{Var}(z_j) = \frac{n}{n-1}\left(\sum z^2(r)\cdot p_j(r) - [\,\overline{z}_j\,]^2\right) \tag{28.2c}$$

Example 28.1. 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 examine these fitness components for five males.

Male	z	W_1	W_2	w_1	p_1	w_2	p_2
1	145	1	25,820	0.714	0.143	1.628	0.233
2	128	1	22,670	0.714	0.143	1.429	0.204
3	148	0	0	0.000	0.000	0.000	0
4	138	2	7,230	1.429	0.286	0.456	0.130
5	141	3	15,986	2.143	0.429	1.008	0.432

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

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

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 it's parent is male 2 is 0.143. The mean fertility *per mating* is

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

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* $\overline{W}_2$ 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 28.2a, 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 z (in mm). Using the body sizes for the males given above, the pre-selection the mean and variance are $\overline{z}_0 = 140.0$ and $\mathrm{Var}[z_0] = 59.5$. From Equation 28.2b, the fitness-weighted mean following the first episode of selection is

$$\overline{z}_1 = 145\cdot 0.143 + 128\cdot 0.143 + 148\cdot 0 + 138\cdot 0.286 + 141\cdot 0.429 = 139.0$$

Similarly,

$$\overline{z_1^2} = 145^2\cdot 0.143 + 128^2\cdot 0.143 + 148^2\cdot 0 + 138^2\cdot 0.286 + 141^2\cdot 0.429 = 19,325$$

Hence

$$\mathrm{Var}(z_1) = \frac{5}{4}(19,325 - 138.996^2) = 6.39$$

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 139.0. Likewise, $\overline{z}_2 = 138.8$ and $\mathrm{Var}[z_2] = 67.7$.

Potential Issues With Assigning Discrete Fitness Values

In many studies, fitness falls cleanly into discrete categories. Often these are simply binary such as alive/dead or mated/unmated. Blanckenhorn et al. (1999) note that biased sampling is not uncommon in such situations. There is a tendency to oversample the rare fitness class in order to increase the power on measured trait. For example, if 10 percent of the population dies, there can be a tendency to have more than ten percent of the measures on individuals showing mortality in order to increase the statistical power to detect small differences between the fitness classes. Such a sampling scheme generates biased estimates of phenotypic selection. Oversampling of a rare fitness class results in an *underestimation* of selection differentials. Conversely, when selection is on the common class (suppose we have 85% mortality in our sample window), then oversampling the surviving individuals results in an *overestimation* of the selection differential. Zuk (1988) and Blanckenhorn et al. (1999) present corrections of these sources of bias, *provided* we know the true population frequency of each fitness class.

Example 28.2. Suppose 10 percent mortality occurs during an episode of selection and one is examining whether a focal trait influences fitness. The trait may be somewhat difficult to measure relative to simply scoring individuals as dead or alive, and thus it is tempting to oversample dead individuals in order to gain more power for testing whether a different in trait value occurs in the living versus dead groups. Suppose the mean of the trait is 100 in surviving individuals and 50 in those showing mortality. With 10% mortality, the mean (before selection) is just $0.1 \cdot 50 + 0.1 \cdot 100 = 95$ giving a selection differential of $100 - 95 = 5$. However, if one oversamples dead individuals to gain more precision in the mean of the trait in this group, say with 25 percent of the sample being dead individuals, then the sample (before selection) mean becomes $0.25 \cdot 50 + 0.75 \cdot 100 = 87.5$ and the resulting selection differential is overestimated as $S = 100 - 87.5 = 12.5$.

A second issue with assigning fitness values to discrete classes was noted by Brodie and Janzen (1996): the absolute value of fitness assigned to a class can, in some cases, influence measures of selection. In the binary case where fitness is scored as 0 and x (such as dead/living, not mated/mated) any (strictly positive) value can be chosen for x as the relative fitnesses are independent of the choice of x. However, with three (or more) discrete fitness classes, this is no longer true. They present an example where laboratory survivorship was followed over a four year period in turtles. One could code this data as simply 0 (do not survive) or 1 (survive). However, the data could also be scored as zero through four, depending on the latest year of survival. They also note that it might make sense to double the assigned fitness value in year four, as this is when reproduction typically starts in nature. These two schemes give different relative fitness values for the surviving age classes and hence different measures of selection. The problem arises here in that only one component (viability) of selection is measured, but (in nature) reproduction would also be occurring which is not measured here, and the different proposed fitness-weighting schemes attempt to partially accommodate this.

Assigning Components of Offspring Fitness to Their Mothers

Generally when assigning reproduction fitness, offspring should be counted at the zygote

stage. If offspring are counted later in life, they can have experienced selection based on their own, as opposed to their parents', phenotypes (Lande and Arnold 1983, Cherverud and Moore 1994), confounding the targets of selection. However, it is not uncommon for avian and mammalian evolutionary geneticists to count only "successful" offspring. For example, eggs in the nest is a close measure of number of zygotes produced, but offspring are often scored as hatchlings, fledglings, or even only when they themselves join the reproductive pool. These measures move increasingly away from number of zygotes and can reflect selection on features of the offspring. Many biologists would counter that maternal care is critical, and success of offspring is a maternal, rather than offspring, trait (Grafen 1988). Where indeed does one draw the line? The strict "line in the sand" at zygotes is formally correct *provided* parental phenotypes have no influence on offspring fitness. However, it is also clear than in species with significant maternal investment in offspring care (such as birds and mammals, as well are plants with significant maternal endosperm contribution in their seeds) that the genotype and phenotype of the mother can influence the fitness of her offspring *independent* of offspring phenotype.

Much of this apparent confusion arises from thinking about this as a univariate selection problem. In reality this is a multiple trait problem, with offspring survival potential involving both the **direct** effect of an offspring on its own fitness and an **indirect** effect from the genotype/phenotype of their mother (Kirkpatrick and Lande 1989). Recall that we have seem similar discussions before in Chapter 18, wherein the fitness of an individual is influenced by its own phenotype (its direct effect) and the **associate effects** of those individuals (potentially parents, other relatives, or even unrelated) around it. Chapter 36 examines the dynamics of such direct/indirect maternal effects models in details. From the standpoint of assigning fitnesses, when maternal phenotype has no effect on the fitness of her offspring, assigning the fitness of the mother by counting offspring *after* potential episodes of selection (such as viability selection) have occurred is misleading. Conversely, if the maternal phenotype *does* impact survival of her offspring independent of their own direct effects, then failure to include this also creates a misleading picture (Wolf and Wade 2001). In terms of correlated traits, one can view maternal preference as one trait and offspring performance as another. Both can influence the fitness of an offspring. If these are uncorrelated (no genetic or phenotypic correlation between the direct and maternal effects), maternal fitness should be assigned as the number of offspring *following* the episodes of selection in the offspring that are influenced by the maternal performance trait. For example, a trait may influence survival to fledging, but then have no future impact. Simply counting number of eggs laid misses the maternal contribution. Conversely, even if the maternal contribution is important, if there are correlations between direct and maternal traits, then a misleading picture of selection can arise unless we uncouple these. Given this somewhat schizophrenic view, how should one proceed? One suggestion is to do the analysis with several measures of reproductive success (based on counting offspring at different stages) and see how consistent the results are. There is also no substitute for a detailed biological understanding of the system being studied, which may suggest when one should be concerned about potential maternal effects.

Example 28.3. Wilson et al. (2005) examined the effects of birth weight, birth date, and litter size on reproductive success in a population of Soay sheep (*Ovis aries*) in Scotland. Two surrogate measures of lifetime fitness were used: lifetime breeding success (LBS, the number of progeny produced during lifetime) and lifetime reproductive success (LRS, which the authors defined for this study as the number of progeny produced that survived to age one). LRS is a measure that potentially combines offspring and maternal fitness, while LBS ignores viability effects after birth (although, of course, they could still be present before birth). Likewise,

fitness-trait associations were consider from two different perspectives, the direct value in the offspring and the indirect value from their mother. Consider the association of birth-weight with total fitness. From an offspring's perspective, this is just the association between their birth weight and either the LBS or LRS measure of total fitness. Conversely, from the maternal perspective, this is the association between the average weight *of her offspring* with her own LBS or LRS. Wilson et al. found that selection gradients on birth weights were much greater through offspring LBS than through maternal LBS. Analysis using LRS at the fitness measure found qualitatively and quantitatively similar results.

Concurrent Episodes, Reproductive Timing and Individual Fitness λ_{ind}

The partitioning of selection into discrete episodes is, of course, an abstraction of the real world. Often this is simply done for the convenience of a researcher who wishes to measure individual fitness by considering just one or two episodes, potentially looking for tradeoffs between them. Selection episodes (e.g., viability and reproduction) often occur *concurrently*, rather than sequentially. For example, in the turtle data of Brodie and Janzen, one has only information about viability (in the laboratory), and how to weight this to account for viability translating into reproductive success raises issues in assigning fitness values. Methods to partition concurrent episodes into their components have been proposed (e.g., Hamon 2005). The more general problem is assigning an appropriate measure of fitness that accounts for reproductive timing.

We have been stressing the use of lifetime reproductive success (LRS), but this is a *rate-insensitive* measure, counting only the total number of offspring, not the actual timing at which they are produced. If reproduction occurs in a discrete window and generations are non-overlapping, LRS does indeed provide a good measure of fitness. However, when generations overlap, the actual *timing* of reproductive events can be more important than the total number of offspring. Consider three individuals, all of which produce a total of forty offspring during their lifespan, so that each have the same fitness as measured by LRS. However, suppose individual one has 10 offspring each at ages 2, 3, 4, and 5, individual 2 has 20 offspring at both ages 4 and 5, and individual three has 20 offspring at ages 2 and 3. Clearly, individual three has a higher *rate* of offspring production, even though it has the same LRS as the other two individuals. How do we account for this?

One solution is based on age-projection (or **Lesile**) matrices (Leslie 1945, 1948; Caswell 1989, 2001). While first proposed by Lenski and Service (1982), who worried that such measured might be biased, it was McGraw and Caswell (1996) who forcefully argued for this approach as a measure of individual fitness in age-structured populations. Recall from Chapter 23 that the growth rate of an age-structure population can be determined from ℓ_x, the probability of surviving from age (or class) x to $x+1$, and f_x, the fecundity in class x. This information is expressed in a $k \times k$ matrix, where k is the upper age limit on reproduction. The first row of the Lesile matrix contains the fecundities, while the below-diagonal line consists of the ℓ_x values,

$$\mathbf{L} = \begin{pmatrix} f_1 & f_2 & \cdots & f_{k-1} & f_k \\ \ell_1 & 0 & \cdots & 0 & 0 \\ 0 & \ell_2 & \cdots & 0 & 0 \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \ell_{k-1} & 0 \end{pmatrix} \tag{28.3a}$$

If $\mathbf{n}(t)$ is a vector of the number of individuals in each age/stage class at some time t, then $\mathbf{n}(t+1) = \mathbf{L}\mathbf{n}(t)$. The (asymptotic) growth rate λ for this population is given by the largest eigenvalue of $\mathbf{L}$ (Chapter 23, Appendix 4). Now consider a particular individual who last

reproduces at age m. The resulting Lesile matrix for *this individual* is

$$\mathbf{L}_{ind} = \begin{pmatrix} f_1 & f_2 & \cdots & f_{m-1} & f_m \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{28.3b}$$

where f_x is half the number of offspring the individual produced at age x (the half to account for both males and females contributing to the population and hence we do not wish to double-count offspring). Note that $LRS = 2\sum_j^k f_j$. While $\mathbf{L}_{ind}$ is very similar to a population growth matrix, the fecundities are those observed for the focal individual (as opposed to population averages for each age class), and we know this individual survives to (at least) age m so that we replace the ℓ_x (average survival per age class) with ones, indicating actual survival. The "growth" rate λ_{ind} for this individual is simply given by the largest eigenvalue of $\mathbf{L}_{ind}$. Alternatively, this satisfies the equation

$$\sum_{i=1}^{m} f_i \, \lambda_{ind}^{-i} = 1 \tag{28.3c}$$

Lenski and Service (1982) expressed concern that λ_{ind} is a biased measure, as the average of the λ_{ind} does not equal the population average growth rate λ. McGraw and Caswell (1996) note two sources of bias. First, for any given phenotypic class, the particular realization $\mathbf{L}_{ind}$ for an individual in that class is a biased estimator of the matrix for that class (i.e., the trait-value z dependent survival $f_x(z)$, and fecundity $\ell_x(z)$). The reason is that random death prevents a proper estimation of the f_x values, especially for later ages. The second source of bias is that the average of the leading eigenvalue for each realization of $\mathbf{L}_{ind}$ does not necessarily equal the leading eigenvalue for the average matrix. Despite these concerns, McGraw and Caswell (1996) still favor this measure, while Lenski and Service (1982) developed a resampled measure with less bias. Alternatively, if the mean population growth rate λ is known, then an unbiased estimator for the relative fitness for individual j is

$$w(j) = \lambda \sum_{i=1}^{k} f_i(j) \; \lambda^{-i} \tag{28.3d}$$

where $f_i(j)$ denotes j's fecundity at age i.

Example 28.4. When generations overlap, the *timing* of reproduction can be at least as important as the lifetime reproductive success (LRS, the total number of offspring produced). Consider the three hypothetical individuals mentioned above, all of whom have a LRS of 40, but have differences in the timing of reproduction. The Leslie matrices for these three individuals become

$$\mathbf{L}_1 = \begin{pmatrix} 0 & 5 & 5 & 5 & 5 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{L}_2 = \begin{pmatrix} 0 & 0 & 0 & 10 & 10 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{L}_3 = \begin{pmatrix} 0 & 10 & 10 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

Here, $\lambda_1 = 2.77$, $\lambda_2 = 1.97$, and $\lambda_3 = 3.58$, Hence, individuals one and two are 77% and 55% (respectively) as fit as individual 3.

Example 29.1 shows that individuals with the same LRS can have very different λ_{ind}, indeed the rankings of individuals based on LRS versus λ_{ind} can be rather different (Figure 29.1). It should not be surprising that Brommer et al. (2002) note several cases where inferences on selection are different when λ_{ind} is used in place of LRS. An especially telling example in shown in Figure 29.1, which plots these two measures of fitness for female Ural owls (*Strix uralensis*). Note that there is a diminishing return in λ_{ind} with increasing LRS. This occurs because reproductive contributions later in life are increasingly down-weighted. While there are strengths to using λ_{ind}, it is not without problems. Brommer et al. (2002) make the important point that the stage *when* offspring are scored (e.g., eggs vs. hatchlings) is critical. Although LRS can also change given differences in the stage of scoring, the effects can be much more dramatic for λ_{ind}, which differentially-weights offspring produced at different ages.

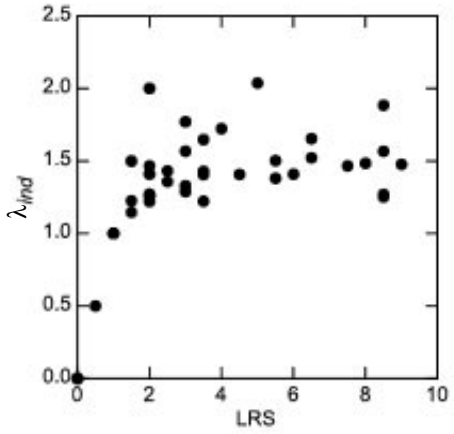

Figure 29.1. Lifetime reproductive success (LRS) versus λ_{ind} in 56 Ural owls where offspring as measured as fledglings. Note here that increasingly LRS results in diminishing returns in λ_{ind}, which occurs because additional reproduction later in life is increasingly down-weighted by λ_{ind}. After Brommer et al. (2002).

We remark in closing that there is a robust discussion in the literature on the appropriate measures of selection in age-structured populations (a partial list includes Charlesworth 1980, Lande 1982, Charlesworth 1983, Lenski and Service 1982, Travis and Henrich 1986, Partridge and Harvey 1988, Henle 1991, de Gong 1994, Metz et al. 1992, Kozlowski 1993, Charlesworth et al. 1994, McGraw and Caswell 1996, Brommer 2000, Brommer et al. 2003, Metcalf and Pavard 2007). Indeed, Stearns (1976) said it best: "Fitness: something everyone understand but no one can define precisely." In part this arises because different questions (e.g., can a new phenotype invade a population versus change in a trait value) can result in different appropriate fitness definitions. The issue of selection response in such age-structured populations was considered in Chapter 23 and will be again in Chapter 36.

VARIANCE IN INDIVIDUAL FITNESS

How do we compare the amount of selection acting on different populations? At first glance, one might consider using the standardized selection differential (the selection intensity) $\overline{i} = S/\sigma$ for comparing the relative strength of individual selection between populations. The drawback with $\overline{i}$ as a measure of *overall* selection on populations is that it is *character specific*. Hence, $\overline{i}$ 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 $\overline{i}$ 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:

$$I = \sigma_w^2 = \frac{\sigma_W^2}{\overline{W}^2} \tag{28.4}$$

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

$$\widehat{I} = \text{Var}(w) = \frac{n}{n-1}\left(\overline{w^2} - 1\right) \tag{28.5}$$

Crow noted that if fitness is perfectly heritable (e.g., $h^2(\text{fitness}) = 1$), then $I = \Delta\overline{w}$, the scaled change in fitness. Following Arnold and Wade (1984a,b) we call I the opportunity for selection, as any variation in individual fitness represents an *opportunity* for a within-generation change in a trait. The opportunity for selection bounds the maximum value of $\overline{i}$. This follows by using (respectively), the definition of a correlation ρ, the Price-Robertson identity $S = \sigma(z, w)$, and the fact that $|\rho| \leq 1$, to give

$$|\,\rho_{z,w}\,| = \frac{|\sigma(z,w)|}{\sigma_z\sigma_w} = \frac{|S|}{\sigma_z\sqrt{I}} \leq 1,$$

implying

$$|\,\overline{i}\,| \leq \sqrt{I} \tag{28.6}$$

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

Example 28.5. To estimate I for the bullfrog data used in Example 28.1, 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 for the five chosen males. Thus

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

giving

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

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 28.1 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 $\bar{i}$ depends on the correlation between relative fitness and the character being considered. Figure 28.2 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.

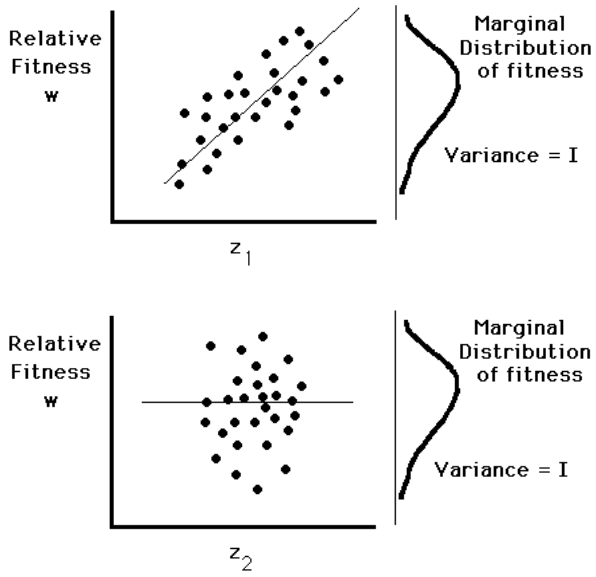

Figure 28.2. The fraction of the opportunity for selection I that is translated into a change in the mean depends on the correlation between relative fitness and the character. Characters z_1 and z_2 have the same marginal distribution of fitness, but only the regression of w on z_1 is significant. Thus (within a generation) selection changes the mean of z_1 but not z_2.

In many cases individual fitnesses are not recorded, instead average fitness for each phenotypic class is estimated. In such cases, we can still obtain a lower bound for I, as the following example illustrates.

Example 28.6. O'Donald (1970, 1971) 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 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 1971 for details).

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

Here number represents the number of females in a particular eyespot class in the laboratory sample. Thus,

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

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

Thus

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

and

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

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). 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 any particular episode. Denote the opportunity of selection associated with the jth 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 jth fitness component:

$$I_j = \sigma^2(w_j) = E(w_j^2) - 1 \tag{28.7a}$$

which is estimated by

$$\begin{aligned}\widehat{I}_j = \text{Var}(w_j) &= \frac{n}{n-1}\left(\overline{w_j^2} - 1\right) \\ &= \frac{n}{n-1}\left(\sum_r w_j^2(r)\, p_{j-1}(r) - 1\right)\end{aligned} \tag{28.7b}$$

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

$$I = \sum_{j=1}^{k} I_j + R \tag{28.8}$$

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

Example 28.7. Compute the estimates of I_1 and I_2 using the data from Example 28.1. Using the relative fitnesses given in the table, $p_0(r) = 1/5$ giving

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

Likewise

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

Hence

$$\widehat{I}_1 = \frac{5}{4}\,(1.531 - 1) \simeq 0.664$$

and

$$\widehat{I}_2 = \frac{5}{4}\,(1.165 - 1) \simeq 0.206$$

Since $\widehat{I} = 0.62$ (Example 28.5), $\widehat{I}_1 + \widehat{I}_2 = 0.87 \neq \widehat{I}$. From Equation 28.8,

$$\widehat{R} = 0.62 - 0.87 = -0.25$$

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).

Correcting Lifetime Reproductive Success for Random Offspring Mortality

As noted by Clutton -Brock (1988), the variance in female lifetime reproductive success tends to increase with the age at which the offspring are counted. For example, in birds I typically increases as we measure reproductive success by counting eggs, hatchings, and fledged offspring. One might expect to see this pattern when maternal care is important, but it can also arise simply from random offspring mortality. Cababa and Kramer (1991) noted that a correction by Crow and Morton (1955), originally suggested for allele frequency change, can be used to adjust for random mortality. Suppose (before an episode of selection) that the mean family size is μ with variance σ_1^2. If offspring mortality is entirely random, with s the probability of survival, then the mean after selection is just $s\mu$ and the variance σ_2^2 satisfies

$$\frac{\sigma_2^2/(s\mu) - 1}{s\mu} = \frac{\sigma_1^2/\mu - 1}{\mu} \tag{28.9a}$$

Thus, Cababa and Kramer find that the opportunity for selection under random offspring mortality is given by

$$I_2 = I_1 + \frac{1/s - 1}{\mu} \tag{28.9b}$$

Thus, random offspring mortality always inflates the variance in reproductive success, with the effect inversely proportional to the survival probability s and mean family size μ. The effect of random mortality is most exaggerated for small mean family size and high random mortality. Cababa and Kramer apply this correction to 43 case studies of vertebrate reproductive success. In all these cases, the mean decreased and the variance increased with time of offspring enumeration. A study of 16 birds found the predicted value of I_2 from Equation 28.9b was roughly 45% of the observed value, while a study of 12 birds and mammals found it to be around 67% of the observed value. Thus, not all of the reduction in I_2 over time is due to random mortality (at least as given by the Crow-Morton model). However, a large fraction of the reduction can indeed to accounted for by this simple model.

Variance in Mating Success: Bateman's Principles

As we (briefly) review in Chapter 42, selection on mate choice – sexual selection — is a major field of study in evolutionary biology. The total reproductive output of an individual is the product of their number of mates and the average fecundity (number of offspring) per mating. Number of mates represent sexual selection, while fecundity per mating is a measure of natural selection. In a classic paper that was overlooked for many years, Bateman (1948) used data from mating success and fecundity in laboratory populations of *Drosophila* to propose several principles regarding sexual selection. These can be expressed in terms of opportunities for selection (Arnold 1994, Arnold and Duvall 1994):

1. *Males show a greater variance in number of offspring than do females*. The opportunity for selection based on total fecundity, a composite measure of both sexual and natural selection, is greater for males than females.
2. *Males show a greater variance in number of mates than do females.* The opportunity for selection when measured by number of mates in larger in males than in females. This is a direct measure of sexual selection.
3. *The total number of offspring in males is an increasing function of number of mates, but is largely independent of number of mates in females.* Provided that females have mated, their fecundity is largely independent of the number of mates. Thus, in females, the average fecundity/mating should be a *decreasing* function of number of mates, as her total fecundity, (number of mates) * (fecundity per mating), is largely independent of the number of mates (provided they have at least one).

The essence of Bateman's observations is that the fitness of females is not increased by additional matings, while male fitness is, so that the third principle drives the first two. While these three principles serve as a useful baseline for thinking about sexual selection, Arnold (1994) and Arnold and Duvall (1994) show that each of these can be violated in different mating systems. If female fecundity also increases with number of matings (such as can occur when males provide nuptial gifts), then we expect sex-specific differences in the opportunity for selection based on both total reproductive success and number of mates to decrease. A hypothetical (extreme) example is in some mantids, wherein the female eats her mate, accruing additional nutrition. In such cases, "successful" males mate just once, while female fecundity could easily (in theory) be an increasing function of number of mates. Arnold and Duvall (1994) suggest that the regression of total fecundity on number of mates (a measure of principle three) provides a metric of the impact of sexual selection on total fitness, and denote this regression coefficient as the **sexual selection gradient**.

Some Caveats in Using the Opportunity for Selection

There are several 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 (Trial 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 28.8, $I_{rs} = I_s + I_f + R$, 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 $\overline{W}$, 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

$$I = \frac{p(1-p)}{p^2} \simeq \frac{1}{p} \quad \text{if } p << 1 \tag{28.10}$$

In this case, 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 28.3 illustrate, the opportunity for selection is often inflated if the observation period is short relative to the total mating period.

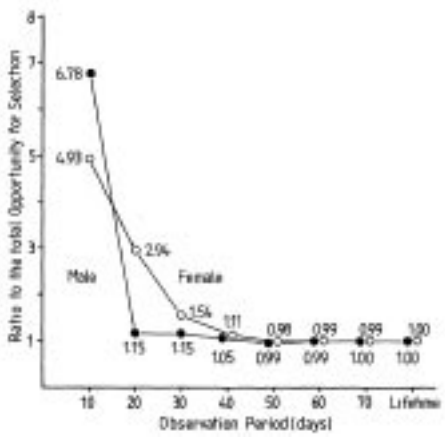

Figure 28.3. The ratio of the opportunity for selection on reproductive success to the lifetime opportunity for the coreid bug *Colpula lativentris* as a function of observation period. Males are given by filled circles, females by open circles. Note the inflation in I when very short intervals are considered. After Nishida (1989).

A second example of the lack of independence between $\overline{W}$ and σ^2_W was given by Downhower et al. (1987). The Poisson distribution is a reasonable model for the number of mates under random mating. Indeed, Joshi et al. (1999) found that the Poisson distribution provides an excellent fit for male mating success in laboratory populations of *Drosophila melanogaster*. 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

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

where $\overline{W}$ 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 $\overline{W} = 0.05$, while if 50 females mate, $\overline{W} = 0.5$. 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 ($I = \overline{W}^{-1}$), 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. In a study of four species with **lek** mating systems (matings typically occur only in a specified area), Mackenzie et al. (1995) found that the combination of random mating and variation in male attendance at the lek account for roughly 40% of the variation in male mating success. Thus, the variance in success was inflated by random factors, but still contains a substantial fraction of unexplained variance.

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 (LW Chapter 11). 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 $\overline{W}$ are independent, recall from LW Chapter 11 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 $\overline{W}$.

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, **more recent refs?**). 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

The discussion thus far in this chapter largely 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 trait.

One simple approach for detecting selection on a character is to compare the (fitness-weighted) phenotypic distribution before and after some episode of selection. We assume (for

this chapter) that any such within-generation change following an episode of selection is *not* due to selection on a phenotypically-correlated trait, relaxing this very unrealistic assumption in Chapter 29. Further, recall from Chapter 16 that a spurious fitness-trait correlation arises if both are correlated with an environmental variable. Even if there is no selection acting on traits phenotypically-correlated to our focal character nor trait-fitness correlations through an environmental variable, an 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 10-14, in many cases the entire selection response following a generation of selection can be reasonably predicted from the within-generation change in these two moments. As we have already seen, there are a number of subtle issues in assigning fitnesses to phenotypes. A thoughtful review of some of these issues is given by Grafen (1988), Cherverud and Moore (1994) and Wolf and Wade (2001).

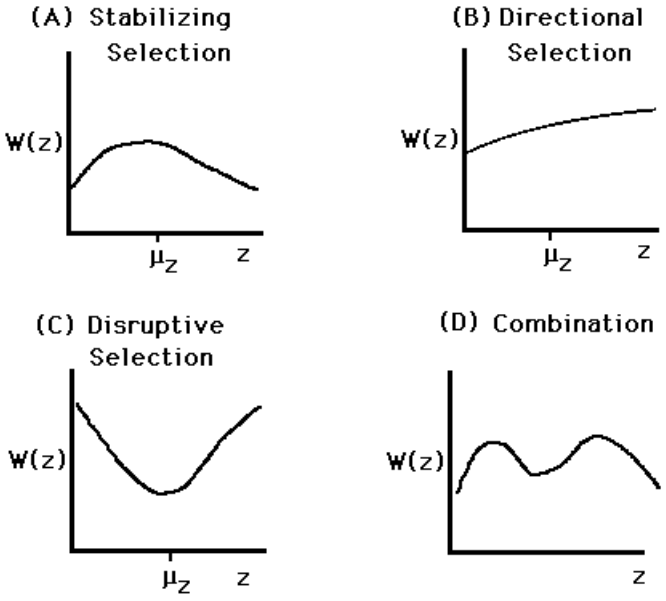

Figure 28.4. Phenotypic selection has historically been roughly classified into three basic forms depending on the local geometry of $W(z)$: stabilizing (**A**), directional (**B**), and disruptive (**C**). As **D** illustrates, populations can simultaneously experience multiple forms of selection.

Fitness Surfaces

$W(z)$, the expected fitness of an individual with phenotype z, describes a **fitness surface** (or equivalently a **fitness function** or **fitness profile**), relating fitness and character value. The *relative* fitness surface $w(z) = W(z)/\overline{W}$ is often more convenient than $W(z)$, and we use the 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 28.4). If fitness is increasing (decreasing) over some range of phenotypes, a population having its mean value in this interval

experiences **directional selection**. Curvature of the fitness surface is important as well, with stabilizing or disruptive selection being possible when curvature is present. 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. More generally, we can have positive (upward or **concave**) or negative (downward or **convex**) curvature without having a maximum/minimum, and hence not formally stabilizing or disruptive selection (respectively). As is illustrated in Figure 28.4D, 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.

$W(z)$ 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 $\overline{W}$ is also a fitness surface, describing the expected fitness of the *population* as a function of the distribution of phenotypes ($p(z, \boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ is the vector of distribution parameters such as the mean and variance) in that population,

$$\overline{W}(\boldsymbol{\Theta}) = \int W(z)\, p(z, \boldsymbol{\Theta})\, dz$$

Hence, mean fitness can be thought of as a function of the parameters of the phenotypic distribution, and we are interested in how change in these parameters changes $\overline{W}$. 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.

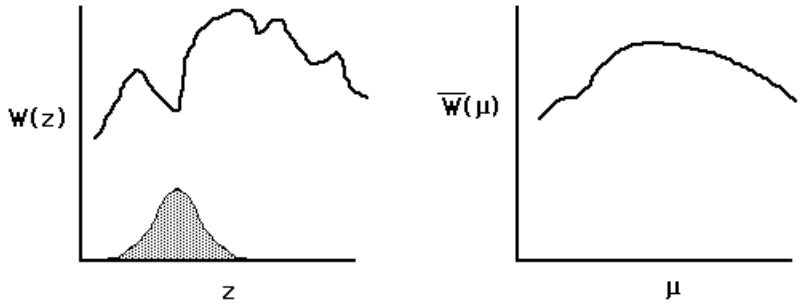

Figure 28.5. In this example, a small change in z can result in a large change in the *individual* fitness surface $W(z)$. However, since the mean population fitness $\overline{W}(\mu_z)$ averages individual fitnesses over the phenotypic distribution, shown as the stippled curve, small changes in μ_z result in only small changes in $\overline{W}(\mu_z)$.

To stress the distinction between the $W(z)$ and $\overline{W}$ fitness surfaces, the former is referred to as the **individual fitness surface**, the latter the **mean fitness surface**. Knowing the individual fitness surface allows us to compute the mean fitness surface for any specified phenotypic distribution $p(z)$, 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 $\overline{W}$ with respect to μ_z describes the change in mean and variance (Equations 10.23c and 28.17). More generally, partials of $\overline{W}$ with respect to higher phenotypic moments describe the dynamics

of selection in the Barton-Turelli response equations (e.g., Equations 24.26 and 24.29). Mean fitness surfaces are considerably smoother than the individual fitness surfaces that generate them (Figure 28.5). 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 (Chapter 24).

DESCRIBING PHENOTYPIC SELECTION: CHANGES IN PHENOTYPIC MOMENTS

Selection for particular phenotypes changes in the trait 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 1994) 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 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 1999) 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 Conover (1999) and Sprent and Smeeton (2007). 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 29). Keeping this important caveat in mind, we now examine measures of selection for single characters.

Directional Selection

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**) $\bar{\imath}$, and the **directional selection gradient** β. These measures are interchangeable (using an appropriate scaling factor) for selection acting on a single character (e.g., Equations 10.5, 10.9 and 10.23a). 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 $\bar{\imath}$) confounds these direct and indirect effects (Equation 29.4).

Quadratic Selection

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

$$\sigma^2_{z^*} - \sigma^2_z = \sigma\left[w, (z-\mu_z)^2\right] - S^2 \tag{28.11}$$

(proved, for the multivariate version, in Example 29.2). 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**

$$C = \sigma^2_{z^*} - \sigma^2_z + S^2 \tag{28.12}$$

which describes selection acting directly on the variance. As we will see below, the term stabilizing selection differential may be 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 28.10). 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 5.44b, 24.27, 24.28). This causes a non-zero value of S that in turn inflates C (Figure 28.6).

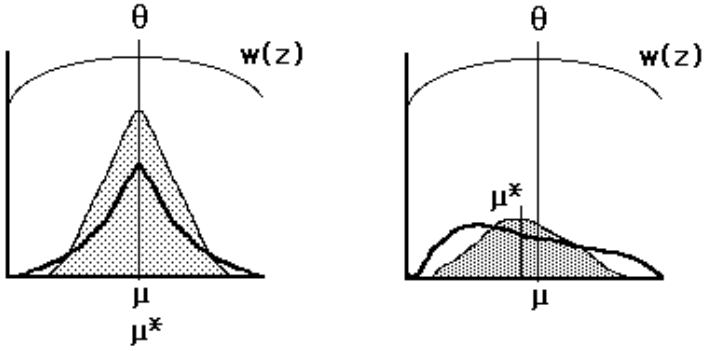

Figure 28.6. Even when a population is under strict stabilizing selection, the mean can change if the phenotypic distribution is skewed. A standard fitness function for stabilizing selection is $W(z) = 1 - b(\theta - z)^2$. O'Donald (1968) found that, even if the population mean is at the optimum value ($\mu_z = \theta$), S is nonnegative if the skew is nonzero ($\mu_{3,z} \neq 0$) as $S = -(b\mu_{3,z})/(1 - b\sigma_z^2)$. The solid line and stippled curve represent the pre- and post-selection phenotypic distributions. **Left**: If phenotypes are distributed symmetrically about the mean ($\mu_{3,z} = 0$), the distribution after selection 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 28.11 implies C is the covariance between relative fitness and the squared deviation of a character from its mean

$$C = \sigma\left[w, (z-\mu)^2\right] \tag{28.13}$$

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 $\rho^2 \leq 1$, we have

$$C^2 \leq \sigma_w^2\,\sigma^2[(z-\mu)^2] = I \cdot \left(\mu_{4,z} - \sigma_z^4\right) \tag{28.14a}$$

where $\mu_{4,z} = E[(z-\mu)^4]$. Equation 28.14a follows from $\sigma^2[(z-\mu)^2] = E[(z-\mu)^4] - E[(z-\mu)^2]^2$.

Thus,

$$|C| \leq \sqrt{I\left(\mu_{4,z} - \sigma_z^4\right)} \tag{28.14b}$$

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

$$|C| \leq \sigma_z^2\sqrt{2I} \tag{28.14b}$$

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

$$\gamma = \frac{\sigma\left[w, (z-\mu)^2\right]}{\sigma_z^4} = \frac{C}{\sigma_z^4} \tag{28.15}$$

As was the case of β, in its univariate form γ appears as a simple rescaling of C, while its multivariate form removes the influence of phenotypic correlations among the measured traits (Chapter 29).

Gradrients Describe the Local Geometry of the Fitness Surface

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, we show in Example A5.4 that β can be expressed in terms of the geometry of the *mean* fitness surface, $\beta = \partial \ln \overline{W}/\partial\mu_z = \overline{W}^{-1}\partial\overline{W}/\partial\mu_z$, the slope of the $\overline{W}$ surface with respect to population mean. β can also be expressed as a function of the *individual* fitness surface. Lande and Arnold (1983) showed, provided z is normally distributed, that

$$\beta = \int \frac{\partial w(z)}{\partial z}\, p(z)\, dz \tag{28.16a}$$

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 29). Likewise, if z is normal,

$$\gamma = \int \frac{\partial^2 w(z)}{\partial z^2}\, p(z)\, dz \tag{28.16b}$$

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.

Gradrients Appear in Selection Response Equations

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 10.23b) 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 13.7b the expected change in variance from a single generation of selection is

$$\begin{aligned}\Delta\sigma_z^2 = \frac{h^4}{2}\,\delta_{\sigma_z^2} &= \frac{\sigma_A^4}{2\,\sigma_z^4}\left(C - S^2\right) \\ &= \frac{\sigma_A^4}{2}\left(\gamma - \beta^2\right)\end{aligned} \tag{28.17}$$

which decomposes the change in variance into changes due to selection on the variance and changes due to directional selection. Again, 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 29, 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 μ_j and σ_j^2 be the (fitness-weighted) mean and variance after the jth episode of selection ($\mu = \mu_0$ and $\sigma^2 = \sigma_0^2$ 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 jth episode of selection are given by

$$S_j = \mu_j - \mu_{j-1} \tag{28.18a}$$

$$C_j = \sigma_j^2 - \sigma_{j-1}^2 + S_j^2 \tag{28.18b}$$

$$\beta_j = \frac{S_j}{\sigma_{j-1}^2} \tag{28.18c}$$

$$\gamma_j = \frac{C_j}{\sigma_{j-1}^4} \tag{28.18d}$$

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

$$S_j = \sigma(w_j, z) \qquad \text{and} \qquad C_j = \sigma[w_j, (z - \mu_{j-1})^2]$$

where w_j is the fitness for the jth episode of selection. Likewise, substituting I_j for I in Equations 28.6 and 28.14c bounds both S_j and C_j.

Example 28.8. For Example 28.1, following male body size over two episodes of selection we find that $\widehat{S}_1 = \overline{z}_1 - \overline{z}_0 = -1$, $\widehat{S}_2 = -0.25$ and $\widehat{S} = -1.25$. Likewise,

$$\widehat{C}_1 = \text{Var}[z_1] - \text{Var}[z_0] + \widehat{S}_1^2 = 6.4 - 59.5 + (-1)^2 \simeq -52.1$$

Similar calculations yield $\widehat{C}_2 \simeq 61.4$ and $\widehat{C} \simeq 9.80$. 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

$$S = \sum_{j=1}^{k} S_j \tag{28.19a}$$

and

$$C = \sum_{j=1}^{k} C_j + \left(S^2 - \sum_{j=1}^{k} S_j^2 \right) \tag{28.19b}$$

Partitioning β and γ requires a little more care, as we have to account for changes in the phenotypic variance following each episode. Rewriting $S_j = \sigma^2_{j-1}\beta_j$ gives

$$\beta = \sum_{j=1}^{k} \beta_j \left[\frac{\sigma^2_{j-1}}{\sigma^2_0}\right] \tag{28.19c}$$

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,

$$\gamma = \sum_{j=1}^{k} \gamma_j \left[\frac{\sigma^4_{j-1}}{\sigma^4_0}\right] + \frac{1}{\sigma^4_0}\left[S^2 - \sum_{j=1}^{k} S_j^2\right] \tag{28.19d}$$

Choice of the Reference Population: "Independent Partitioning"

Using the above **additive partitioning** scheme, selection differentials and gradients for a particular episode are based on their fitness-weighted values from the previous episode and are (with appropriate weighting) additive across episodes. Several authors have suggested that using an **independent partitioning** instead can provide additional insight into the nature of selection (Conner 1988, Koenig and Albano 1987, Koenig et al. 1991, Preziosi and Fairbairn 2000). Under an independent partitioning, one uses the observed distribution of phenotypes as the reference population (as opposed to their fitness-weighted distribution) when computing statistics for each episode. Such a partition does not weight for previous selection (and hence provides a misleading picture for evolutionary response), but it may provide insight into the action of selection in particular episode, as Example 28.9 highlights.

Example 28.9. To illustrate the difference between additive and independent partition, Koenig et al. (1991) present the following hypothetical data set relating body size with natural (survival) and sexual (mating rate) selection.

Size z	Suriviorship (Days alive), W_1	Mating Success (Mates/days alive), W_2	Total Mates $W_1\,W_2$
11	1	3	3
12	1	2	2
13	1	1	1
11	10	1	10
12	10	2	20
13	10	3	30

Here, the mean of z before selection is 12. Focusing on total mates, the mean number of mates is 11, and the fitness-weighted mean is

$$\frac{1}{6}\left[11\left(\frac{3+10}{11}\right) + 12\left(\frac{2+30}{11}\right) + 13\left(\frac{1+20}{11}\right)\right] = 12.3$$

Thus, the total selection differential on size due to size-related differences in number of mates of 12.3-12 = 0.3. However, the selection differential in survivorship is zero, and so the positive

differential on total mates arises due to the fitness-weighted differences in W_2. However, suppose one simply followed mating success per day and had no knowledge of survivorship. Taking all six individuals being equally-weighted (1/6) as the reference population, $\overline{W}_2 = 2$, giving the mean after selection as

$$\frac{1}{6}\left[11\left(\frac{3+1}{2}\right) + 12\left(\frac{2+2}{2}\right) + 13\left(\frac{1+3}{2}\right)\right] = 12$$

Thus, the selection differential on mating rate using this reference population is zero. If we ignore differences in survival, there are no differences in sexual selection (mating rate). The net differential on total mates arises because individuals with lower fitness in the first episode are given less weight when computing the additive partition, while there are equally weighed using the independent partition.

The example highlights the critical important of accounting for all selection. If the study simply followed survivorship or simply followed mates/day, no trait-fitness associations would have been seen. Grafen (1988) has coined the very appropriate term of the **invisible fraction** for that population undergoing selection that is not seen by the investigator. This missing data will significantly bias estimates of selection (e.g., Bennington and McGraw 1995, Hadfield 2008).

Koenig et al. (1991) stress that the additive and independent partitions are complementary measures. The additive partition correctly accounts for how fitness builds up over episodes of selection, while the independent partition examines whether or not a trait has fitness consequences *independent* of selection at other stages. Examining both measures can provide the investigator with insight into the biology of the system being examined. Further, if episodes are *not* sequential, then the additive partition is not appropriate, while the independent partition can still be used to evaluate potential targets of selection.

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 to intentionally 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 LW Appendix 1 can be used to obtain approximate large sample variances for estimators of differentials and gradients. The (exact) sampling variance for the directional selection differential is

$$\sigma^2\left(\widehat{S}_j\right) = \frac{\sigma_j^2}{n_j} + \frac{\sigma_{j-1}^2}{n_{j-1}} \tag{28.20a}$$

where n_j is the sample size for the jth episode. Using the approximation methods from LW

Appendix 1, the large-sample variance for C is approximately

$$\sigma^2\left(\widehat{C}_j\right) \simeq 4S_j^2\,\sigma^2\left(\widehat{S}_j\right) + 8S_j\left(\frac{\mu_{3,j}}{n_j} + \frac{\mu_{3,j-1}}{n_{j-1}}\right) + \frac{\mu_{4,j} - \sigma_j^4}{n_j} + \frac{\mu_{4,j-1} - \sigma_{j-1}^4}{n_{j-1}} \tag{28.20b}$$

If phenotypes are normally distributed, this reduces to

$$\sigma^2\left(\widehat{C}_j\right) \simeq 4S_j^2\,\sigma^2\left(\widehat{S}_j\right) + 2\left[\frac{\sigma_j^4}{n_j} + \frac{\sigma_{j-1}^4}{n_{j-1}}\right] \tag{28.20c}$$

If the scaled skewness k_3 (LW Equation 2.8) and kurtosis k_4 (LW Equation 2.12a) are small, this normal approximation can 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 28.10. 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 and variance for body weight in 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 $\widehat{S} = 16.85 - 15.79 = 1.06$ and Equation 28.20a gives the standard deviation of this estimate as

$$\text{SE}\left(\widehat{S}\right) \simeq \sqrt{\frac{2.37}{642} + \frac{2.43}{85}} \simeq \sqrt{0.0323} \simeq 0.180$$

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, using the quadratic selection differential to correct for the reduction in the variance from directional selection gives

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

consistent with selection to increase the variance in addition to directional selection. From Equation 28.20c, assuming body size is normally distributed before and after the drought,

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

$\widehat{C}$ 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 can be used

to estimate approximate confidence intervals (e.g., Mitchell-Olds and Shaw 1987, Mitchell-Olds and Bergelson 1990), and randomization tests can be used to test for significance (e.g., Moore 1990, Hews 1990). These procedures are rather 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 29).

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 miminizes 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 (individuals with the same trait value) 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), $W(z) = \overline{W}$, 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, although this can induce serious, even fatal, biases, see Hadfield 2008)). 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, consider fitness 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. Finally, as emphasized in Chapter 16, organisms often modify their environments as they evolve, so that the biotic environment can change through selection, modifying the nature of future selection.

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 (LW Equation 3.14b) that β is the slope of the least-squares linear regression of relative fitness on z,

$$w = a + \beta z + e \tag{28.21a}$$

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 (1 and μ, respectively), this can be written as

$$w = 1 + \beta(z - \mu_z) + e \tag{28.21b}$$

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

$$r_{z,w}^2 = \frac{\mathrm{Cov}^2(z,w)}{\mathrm{Var}(z)\cdot\mathrm{Var}(w)} = \widehat{\beta}^2\frac{\mathrm{Var}(z)}{\widehat{I}} \tag{28.22}$$

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

$$w = a + b_1 z + b_2(z - \mu_z)^2 + e \tag{28.23a}$$

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

$$w = 1 + b_1(z - \mu_z) + b_2\left[(z - \mu_z)^2 - \sigma_z^2\right] + e \tag{28.23b}$$

The regression coefficients b_1 and b_2 nicely summarize the local geometry of the fitness surface. Evaluating the derivative of Equation 28.23 at $z = \mu_z$ gives

$$\left.\frac{\partial w(z)}{\partial z}\right|_{z=\mu_z} = b_1 \qquad \text{and} \qquad \left.\frac{\partial^2 w(z)}{\partial z^2}\right|_{z=\mu_z} = 2b_2 \tag{28.24}$$

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 (**concave**) curvature, while $b_2 < 0$ implies the curvature is downward (**convex**). 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 is concave in some interval, while a necessary condition for a local maximum is that the function is convex. 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 28.7 shows, the fitness function can curve downward without the population experiencing a local maximum or can curve upward without having a local minimum.

We solve for the regression coefficients b_1 and b_2 by transforming Equation 28.23 into a standard multiple regression problem by setting $x_1 = z$ and $x_2 = (z - \mu_z)^2$ and applying the methods of LW Chapter 8. To proceed, we need expressions for $\sigma(x_1, x_2)$, $\sigma(x_1, w)$, and $\sigma(x_2, w)$. From LW Equation A1.14, $\sigma(x_1, x_2) = \sigma\left(z, (z - \mu_z)^2\right) = \mu_{3,z}$, the skew of the

phenotypic distribution before selection. Likewise, from Equations 10.7 and 28.13, $\sigma(x_1, w) = \sigma(z, w) = S$ and $\sigma(x_2, w) = \sigma(\,(z - \mu_z)^2, w) = C$. Substituting these into the results from LW Example 8.3 (which gives exact expressions for the partial regression coefficients in a bivariate regression), and noting that $\sigma^2(\,(z - \mu_z)^2\,) = \mu_{4,z} - \sigma_z^4$, gives

$$b_1 = \frac{\sigma^2(x_2) \cdot \sigma(x_1, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{(\mu_{4,z} - \sigma_z^4) \cdot S - \mu_{3,z} \cdot C}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (28.25a)$$

$$b_2 = \frac{\sigma^2(x_1) \cdot \sigma(x_2, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{\sigma_z^2 \cdot C - \mu_{3,z} \cdot S}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (28.25b)$$

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

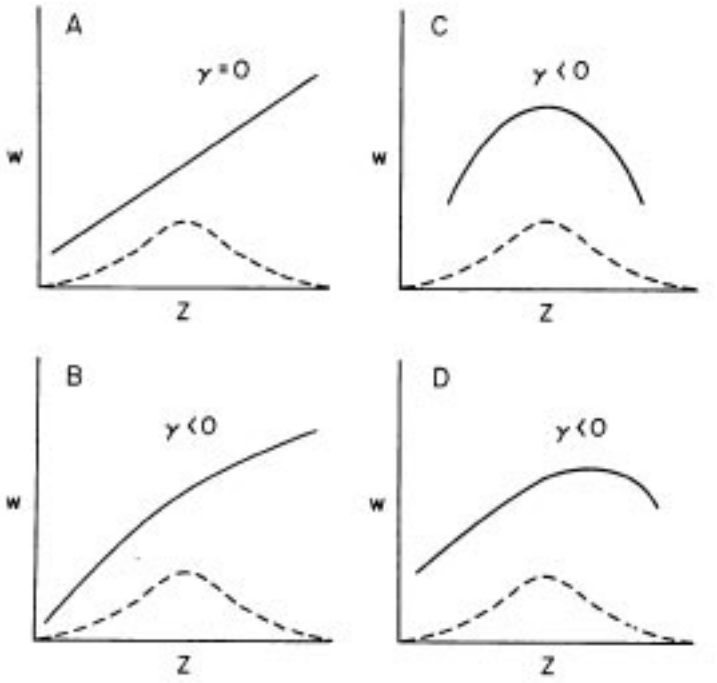

Figure 28.7. The relationship between γ and curvature of the fitness function (solid line). Dashed curve is the distribution of z. **A**: $W(z)$ is strictly linear, hence $\gamma = 0$. **B**: $W(z)$ curves downward (is convex), but has no maximum. Hence, $\gamma < 0$, implying stabilizing selection by the Lande-Arnold criterion, when in fact selection is entirely directional. **C**: Stabilizing selection only, as there is no change in the mean. **D**: A combination of directional and stabilizing selection (as the population mean is not under the optimal fitness value). From Mitchell-Olds and Shaw (1987).

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, the definitions of β and γ imply, respectively, that $b_1 = \beta$ and $b_2 = \gamma/2$, giving the univariate version of the **Lande-Arnold regression**,

$$w = 1 + \beta(z - \mu_z) + \frac{\gamma}{2}\left((z - \mu_z)^2 - \sigma_z^2\right) + e \quad (28.26)$$

developed by Lande and Arnold (1983), motivated by Pearson (1903), and hence we also occasionally refer to this as the **Pearson-Lande-Arnold regression**. The Lande-Arnold regression thus provides a connection between selection gradients (directional and stabilizing) and quadratic approximations of the individual fitness surface.

An important point from Equation 28.25a 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 on $(z - \mu_z)^2$. The biological significance of this can be seen by reconsidering Figure 28.6, where the presence of skew in the phenotypic distribution results in a change in the mean of a population under strict stabilizing selection (defined as 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 10.7), 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 standard 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.

Jackknife estimates were introduced by Tukey (1958) as a generalized statistical tool. A nice introduction can be found in Sokal and Rohlf (1994), with more detailed treatments in Miller (1974), Wu (1986), Shao and Tu (1996), and Manly (1997). 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 28.21a. Denote by $\widehat{\beta}$ the standard least-squares estimate of β using the full data set of n individuals, and let $\widehat{\beta}_i$ denote the estimator using the complete data set minus data for the ith individual. The resulting jackknife estimator is

$$\widehat{\beta}_{jack} = \frac{1}{n}\sum_{i=1}^{n} \phi_i = \overline{\phi} \quad \text{where} \quad \phi_i = n\widehat{\beta} - (n-1)\widehat{\beta}_i \tag{28.27a}$$

which has approximate large-sample variance

$$\text{Var}\left(\widehat{\beta}_{jack}\right) \simeq \frac{1}{n(n-1)} \sum_{i=1}^{n} (\phi_i - \overline{\phi})^2 \tag{28.27b}$$

Approximate large-sample confidence intervals follow using Equation 28.27b and the fact that $\widehat{\beta}_{jack}$ 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. A program for computing both the weighted and unweighted jackknife estimates for multiple character quadratic fitness regressions has been developed by Mitchell-Olds (1989).

Randomization tests provide another approach for testing the significance of a regression. Again, the idea behind this class of tests is simple but computationally intensive. A particular value of $\widehat{\beta}$ 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 null hypothesis of no association between individual fitness and character value. Suppose we obtain a standard least-square estimate (assuming a linear regression) of $\widehat{\beta}$=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 $\widehat{\beta}$ values in excess of 1.25. This suggests that this value is significantly different from zero at the five percent, but not the one percent, level. See Moore (1990) and Hews (1990) for applications of randomization tests to fitness data. (**updated refs?**)

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, **updated references?**) 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.

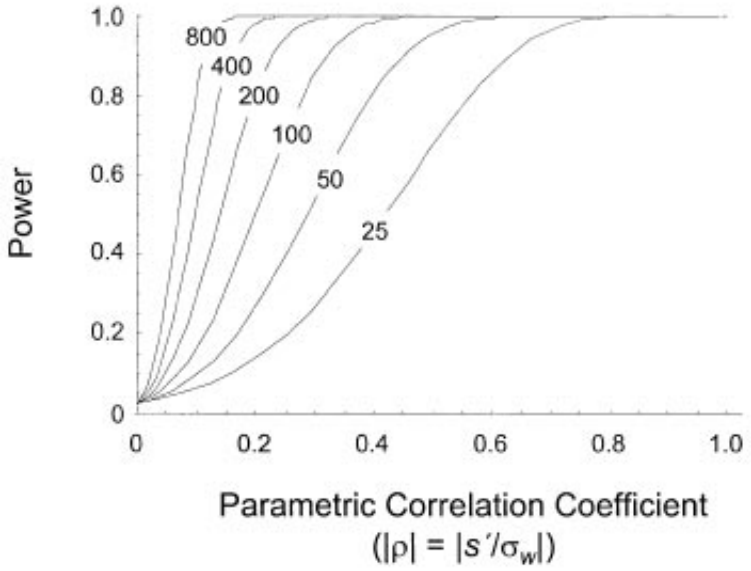

Figure 28.8. The power of a univariate regression to detect a directional selection gradient is a function of the correlation ρ between trait value and relative fitness, where $\rho = \overline{i}/\sigma_w = \beta\,\sigma_z/\sigma_w$. Power is plotted as a function of ρ with curves for sample size starting at n = 25 and successively doubling until 500. Here power is the probability that the sample correlation is declared significantly different from zero using a test of significance of $\alpha = 0.05$. After Hersch and Phillips (2004).

Power

Another critical issue is the power of a regression to detect selection. As discussed in LW Appendix 5, power is simply the probability of detecting (i.e., declaring significant) an effect given a preset significance level. Our focus here is on the power to detect a directional selection gradient β in a univariate regression. A convenient way to compute power for a regression is to consider the correlation ρ between the trait value and relative fitness,

$$\rho = \frac{\sigma(z,w)}{\sigma_z\,\sigma_w} = \frac{S}{\sigma_z\,\sigma_w} = \frac{\overline{\imath}}{\sqrt{I}} \tag{28.28a}$$

which is the ratio of the selection intensity and the square root of the opportunity for selection (Hersch and Phillips 2004). Note that we can also express ρ as

$$\rho = \frac{S}{\sigma_z\,\sigma_w} = \beta\,\frac{\sigma_z}{\sigma_w} \tag{28.28b}$$

The power to detect a directional selection gradient is thus a function of both the strength of selection (measured by the selection intensity $\overline{\imath} = \beta\sigma_z$) as well as the variance in total fitness (I). A strong amount of selection (a large $\overline{\imath}$) per se does not imply high power, rather it is the strength of selection *relative* to the total variance in fitness that is critical. Assuming normally-distributed residuals, Hersch and Phillips (2004) show that the (adjusted) sample correlation

$$r' = \sqrt{\frac{AR(w)}{\sigma_w^2}} \cdot \sqrt{\frac{\sigma_z^2}{Var(z)}} \cdot r \tag{28.28c}$$

can be scaled to follow a students-t distribution,

$$\sqrt{n-1}\left(\frac{r'-\rho}{\sqrt{1-\rho^2}}\right) \sim t_{n-1} \tag{28.28d}$$

When the sample correlation r is sufficiently large, the regression is declared to be significant. Using these results, power calculations follow from standard approaches (see similar examples in LW Appendix 5). As shown in Figure 28.8, sample sizes typically need to be in the hundreds to have significant power of detecting even modest selection. The above expressions for power assume *normally-distributed residuals*. As we have mentioned, this assumption is often violated, especially with viability data. With such 0/1 coded fitness data, the residuals follow binomial, rather than normal, distributions and display heteroscedasity. Simulations by Hersch and Phillips (2004) show that using the above expressions *overestimates* power when residuals are binomially–distributed.

Quadratic Surfaces Can be Very Misleading

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, or even sharp transitions, are thus very poorly described by a quadratic. Figure 28.9 gives a particularly illustrative example, showing that a quadratic fit to a truncation selection fitness function creates a spurious local minimum.

Given this potential for a very misleading view of the fitness surface, why all the focus on quadratic regressions? There are two primary motivations. First, the quadratic is the simplest function that allows for curvature, and hence the simplest estimate of any non-linearity in the fitness surface. Second, and much more importantly, when the conditions for the breeders' equation hold (Chapters 10, 13, 24), the sole measures of phenotypic change entering into

the selection response equations for mean and variance are the coefficients from the best linear (β) and quadratic (γ) fit. Hence, even if the fitness surface is very poorly described by a quadratic, even to the point of being very misleading (Figure 28.9), one would still extract the coefficients for the response in the mean and variance by using the quadratic estimate of this surface. There is thus the potential for conflict in the use of quadratic regressions between ecologists (who wish to ascertain how traits influence fitness) and evolutionary biologists (who wish to see how these traits will evolve). In reality, both viewpoints are correct. The more accurate the description of the fitness surface, the more ecological insight into the trait, but this fitness surface also needs to be translated into the evolutionary dynamics of the trait.

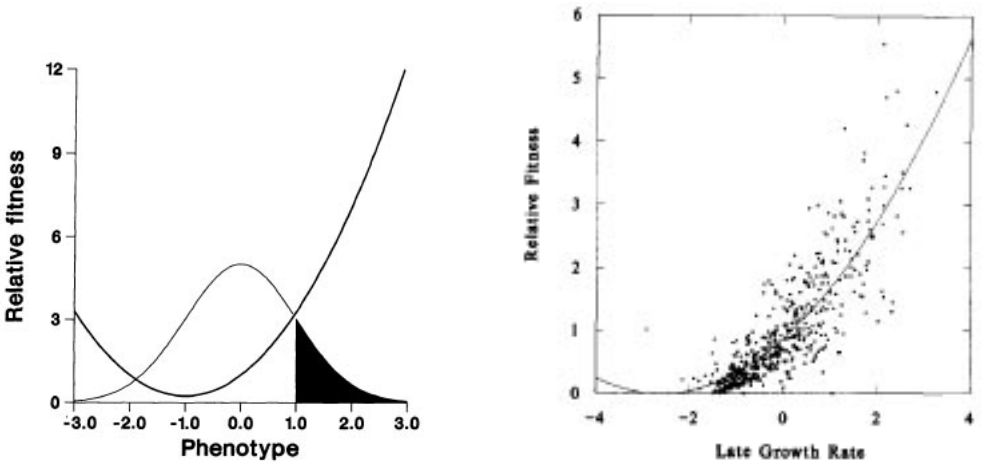

Figure 28.9. Examples of a misleading approximation of $W(z)$ resulting from using a quadratic regression. **Left**. A hypothetical example wherein phenotypes are normally distributed with only individuals exceeding one phenotypic standard deviation surviving (so that $W(z)$ is the square-wave function). 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 (1990) 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 mimimun 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 (1990).

Fitting Other Parametic Fitness Functions

If our sole focus is in describing the fitness surface (as opposed to extracting components of selection response), then other parameteric forms besides the simple quadratic are possible. One obvious candidate is a Gaussian fitness function,

$$W(z) = a \exp\left(-\frac{(z-\theta)^2}{\omega}\right) \tag{28.29}$$

Weldon (1901), in one of the first studies of selection on a quantitative trait in nature, remarked that the Gaussian seems to be a good description of the fitness function, and it was

used by Cavalli-Sforza and Bodmer (1972) to model the human birth weight-survival relationship seen by Karn and Penrose (1951). Likewise, the exponential function can also be used (Mitchell-Olds and Bergelson 1990, see Figure 29.8).

For viability data, a variety of approaches from survival analysis can be used. Since individual viability fitness data is coded as zero/one, **logistic regression** is a natural choice (Janzen and Stern 2006), with

$$W(z) = \Pr(\text{survival} \,|\, z) = \frac{\exp(-a + bz)}{1 + \exp(-a + bz)} \tag{28.30}$$

Note that $W(z)$ increases from zero to one as z increases. Logistic regressions naturally handle the heteroscedatic nature of residuals for viability, and maximum likelihood (LW Appendix 3) can be used to estimate model parameters. Modifications of logistic regressions can also be used to estimate fitness using capture-recapture data (Kingsolver and Smith 1995).

More general approaches to survival functions come from the analysis of clinical trails (Klein and Moeschberger 1997, Lawless 2003) and time-to-failure failure analysis from industrial statistics (Kenett and Zacks 1998), both example of **surivial analysis**. Manly (1976) suggested use of the **double exponential fitness function** for viability data, where

$$W(z) = \Pr(\text{survival} \,|\, z) = \exp(-\exp(a + bz) \tag{28.31}$$

This is a special case of the **proportional hazards model** (Cox 1972), where the probability of survival to time t is given by a general risk for every individual in the population (specified by some monotonically non-decreasing function $g(t)$, so that probability of survival declines with increasing t), and a specific (proportional) risk for the particular phenotype z, giving

$$W(z,t) = \Pr(\text{survival} \,|\, z,t) = \exp(-\exp(ag(t) + bz) \tag{28.32a}$$

Likewise, if survival at stage j is measured, the proportional hazards model can be written as

$$W(z,j) = \Pr(\text{survival} \,|\, z,t) = \exp(-\exp(h_j + bz) \tag{28.32b}$$

One advantage of using such models is that they can allow for certain types of missing, or **censored**, data (Little and Rubin 2002).

Nonparametric Approaches: Schluter's Cubic-Spline Estimate

In order to more reliably estimate the fitness function, Schluter (1988) developed a nonparametric 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. Schluter developed a program to estimate $W(z)$ assuming either normally, binominally, or Poisson distributed residuals. 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 28.10. 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).

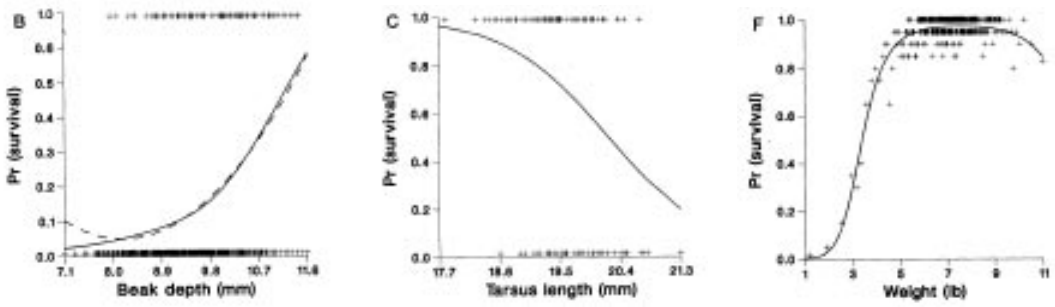

Figure 28.10. Examples of fitness surfaces generated using Schluter's method. The actual fitness values for individuals are indicated by +, the solid curve indicates the cubic-spline estimate of $W(z)$. **Left**: Probability of survival as a function of beak depth in Darwin's finch *Geospiza fortis*. 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, Kingsolver and Pfenning 2007). Spurious correlations between a character and fitness can be generated in a variety of ways: environmental correlations between character value and fitness (Chapter 16), selection on unmeasured characters correlated with the observed character (Chapter 29), and loci with direct fitness effects having pleiotropic effects on the character being measured (Chapter 5). Recall Example 5.8, 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 trait-fitness correlation that is entirely due to levels of inbreeding rather than intrinsic differences between phenotypic values (Willis 1996). Directional and concave (stabilizing) selection will appear stronger when inbreeding is present, overestimating their effects.

Mitchell-Olds and Shaw (1987) and Wade and Kalisz (1990) suggest that interactions between environmental effects and fitness are extremely important, a point also discussed at some lenght in Chapter 16. For example, Breden and Wade (1989) observed a positive relationship between group size and fitness in the imported willow leaf beetle. However, when predators 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.

Given all of these potentials for false associations, the most direct test of a trait-fitness association suggested by a regression is **phenotypic manipulation**, the artificial modification of trait values followed by a fitness assessment in nature (Sinervo and Basolo 1996, Travis and Reznick 1998).

Example 28.11. Grether (1996) examined male mating success and survival in a California population of the rubyspot damselfly (*Hetaerina americana*). Mature males of this species are marked by a red spot at the base of their wings, and Grether was interested in whether variation in the size of this spot influences sexual and/or natural selection. Three fitness components were measured: reproductive life span (a measure of survival selection), mating rate (a measure of sexual selection), and lifetime mating success (a combination of both survival and sexual selection). All three of these showed significant selection gradients with red spot size, but not with body size. Thus, the Lande-Arnold regression suggests that larger spot size is favored by both sexual (mating rate) and natural (reproductive life span) selection. As a test of this, Grether examined these components in three additional groups: an "enlarged " group where red ink was used to increase spot size, a "sham" group where the same area was filled with clear ink, and an unmanipulated group. Mate success (mates per day) was significantly greater in the enlarged group, while unchanged in the sham versus control groups. Thus, experimental manipulated confirms the role of red spot size in sexual selection. Surprisingly, however, males with enhanced spots had mortality rates 23% higher than the controls, while the sham and control groups showed no differences. Thus, contrary to what the fitness regressions suggest, direct manipulation show that increasing spot size decreases survival. Hence, unmeasured traits (or other factors) generated an apparently strong selection gradient that could not be verified by phenotypic manipulation.

Literature Cited

Arnold, S. J. 1983. Sexual selection: the interface of theory and empircism. *In* P. Bateson (ed.), *Mate choice*, pp. 67–107. Cambridge University Press, Cambridge. [28]

Arnold, S. J. 1986. Limits on stabilizing, distruptive, and correlational selection set by the opportunity for selection. *Amer. Nat.* 128: 143–146. [28]

Arnold, S. J. 1994. Bateman's principles and the measurement of sexual selection in plants and animals. *Amer. Natl.* 144: S126–S149. [26]

Arnold, S. J., and D. Duvall. 1994. Animal mating systems: a synthesis based on selection theory. *Amer. Natl.* 143: S317–348. [26]

Arnold, S. J., and M. J. Wade. 1984a. On the measurment of natural and sexual selection: theory. *Evolution* 38: 709–719. [28]

Arnold, S. J., and M. J. Wade. 1984b. On the measurment of natural and sexual selection: applications. *Evolution* 38: 720–734. [28]

Banks, M. J., and D. J. Thompson. 1985. Lifetime mating success in the damselfly *Coenagrion puella*. *Anim. Behav.* 33: 1175–1183. [28]

Bateman, A. J. 1948. Intra-sexual selection in *Drosophila*. *Heredity* 2: 349–368. [28]

Bateson, P. (Ed.) 1983. *Mate choice*. Cambridge University Press, Cambridge. [28]

Bennington, C. C., and J. B. McGraw. 1995. Phenotypic selection in an artificial population of *Imatiens pallida:* the importance of the invisible fraction. *Evolution* 49: 317–324. [28]

Blanckenhorn, W. U., M. Reuter, P. L. Ward, and A. D. Barbour. 1999. Correcting for sampling bias in quantitative measures of selection when fitness is discrete. *Evolution* 53: 286–291. [28]

Boag, P. T., and P. R. Grant. 1981. Intense natural selection in a population of Dawrin's finches (Geospizinae) in the Galápagos. *Science* 214: 82–84. [28]

Bradbury, J. W., and M. B. Andersson. 1987. *Sexual selection: testing the alternatives*. John Wiley and Sons, New York, New York. [28]

Brakefield, P. M. 1984. The ecologicial genetics of quantitative characters of *Maniola jurtina* and other butterflies. *In* R. J. Vane-Wright and P. R. Ackery (eds.), *The biology of butterflies*, pp. 167–190. Princeton Univeristy Press, Princeton. [28]

Breden, F. J. and M. J. Wade. 1989. Selection within and between kin groups of the imported willow leaf beetle. *Amer. Nat.* 134: 35–50. [28]

Brodie, E. D. III, and F. J. Janzen. 1996. On the assignment of fitness values in statistical analysis of selection. *Evolution* 50: 437–442. [28]

Brodie, E. D. III, A. J. Moore, and F. J. Janzen. 1995. Visualizing and quantifying natural selection. *Trends Ecol. Evol. Biol.* 10: 313–318. [28]

Brommer, J. E. 2000. The evolution of fitness in life-history theory. *Biol. REv. Cambridge Phil. Sco.* 75: 377–404. [28]

Brommer, J. E., J. Merilä, and H. Kokko. 2002. Reproductive timing and individual fitness. *Ecol. Letters* 5: 802–810. [28]

Bumpus, H. C. 1899. The elimination of the unfit as illustrated by the introduced House Sparrow,*Passer domesticus*. *Biol. Lectures, Marine Biol. Lab., Woods Hole*: 209-226. [28]

Byerly, H. G., and R. E. Michod. 1991. Fitness and evolutionary explanation. *Biol and Philos.* 6: 1–22. [28]

Cabana, G., and D. L. Kramer. 1991. Random offspring mortality and variaation in parental fitness. *Evolution* 45: 228–234. [28]

Caswell, H. 1989. *Matirx population models.* Sinauer, Sunderland, Ma. [28]

Caswell, H. 2001. *Matirx population models, 2nd Edition* Sinauer, Sunderland, Ma. [28]

Charlesworth, B. 1980. *Evolution in age-structured populations*. Cambridge University Press [28]

Charlesworth, B. 1983. Natural selection on multivariate traits in age-structured populations. *Proc. Roy. Soc. B* 251: 47–52. [82]

Charlesworth, B., C. Cannings, and F. C. Hoppensteadt. 1994. *Evolution in age-structured populations, 2nd edition*. Cambridge University Press [28]

Cheverud, J. M., and A. J. Moore. 1994. Quantitative genetics and the role of the environment provided by relatives in the evolution of behavior. *In* C. R. B. Boake (Ed), *Quantitative genetic studies of behavioral evolution.*, pp. 67–100. University of Chicago Press.

Clutton-Brock, T. H. (ed.) 1988. *Reproductive success: studies of individual variation in contrasting breeding systems*. University of Chicago Press, Chicago. [28]

Conner, J. 1988. Field measurements of natural and sexual selection: applications. *Evolution* 42: 736–749. [28]

Conover, W. J. 1999. *Practical nonparametric statistics, 3rd edition*. Wiley and Sons, New York. [28]

Connner, J. K. 2001. How strong is natural selection? *Trends in Ecology and Evolution* 16: 215–217. [28]

Cox, D. R. 1972. Regression models and life tables. *J. Roy. Statist. Soc.* 34: 187-220. [28]

Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30: 1–13. [28]

Crow, J. F., and N. E. Morton. 1955. Measurement of gene frequency drift in small populations. *Evolution* 9: 202–214. [28]

Darwin, C. 1859. *The origin of species by means of natural selection.* Murray, London. [28]

Darwin, C. 1871. *The descent of man and selection in relation to sex*. John Murray, London. [28]

de Jong, G. 1994. The fitness of fitness concepts and the description of natural selection. *Q. Rev. Biol.* 69: 3–29. [28]

Dowdeswell, W. H. 1961. Experimental studies on natural selection in the butterfly, *Maniola jurrina*. *Heredity* 16: 39–52. [28]

Downhower, J. F., L. S. Blumer, and L. Brown. 1987. Opportunity for selection: an appropriate measure for evaluating variation in the potential for selection? *Evolution* 41: 1395–1400.

Endler, J. A., 1986. *Natural selection in the wild*. Princeton University Press, Princeton, New Jersey. [28]

Fisher, R. A. 1930. *The genetical theory of natural selection.* Clarendon. [28]

Grafen, A. 1988. On the uses of data on lifetime reproductive success. *In* T. H. Clutton-Brock (ed.), *Reproductive success: studies of individual variation in contrasting breeding systems* pp. 454-471. University of Chicago Press, Chicago. [28]

Grether, G. F. 1996. Sexual selection and survival selection on wing coloration and body size in the rubyspot damselfly *Hetaerina americana. Evolution* 50: 1939–1948. [28]

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. Roy. Soc. B* 275: 723–734. [28]

Hamon, T. R. 2005. Measurement of concurrent selection episodes. *Evolution* 59: 1096–1103. [28]

Henle, K. 1991. Some reflections on evolutionary theories with a classification of fitness. *Acta Biotheor.* 39: 91–106. [28]

Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. *Eovlution* 58: 479–485. [28]

Hews, D. K. 1990. Examining hypotheses generated by field measures of sexual selection on male lizards, *Uta palmeri*. *Evolution* 44: 1956–1966. [28]

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *PNAS* 98: 9157–9160 [28]

Houck, L. D. , S. J. Arnold, and R. A. Thisted. 1985. A statistical study of mate choice: Sexual selection in a plethodontid salamander (*Desmognathus ochrophaesus*). *Evolution* 39: 370–386. [28]

Howard, R. D. 1979. Estimating reproductive success in natural populations. *Amer. Nat.* 114: 221–231. [28]

Joshi, A., M. H. Do, and L. D. Mueller. 1999. Poisson distribution of male mating success in laboratory populations of *Drosophila melanogaster*. *Genet. Res.* 73: 239–249. [28]

Karn, M. N., and L. S. Penrose. 1951. Birth weight and gestation time in relation to maternal age, parity and infant survival. *Ann. Eugen.* 16: 147–164. [28]

Kalisz, S. 1986. Variable selection on the timing of germination in *Collinsia verna (Scrophulariaceae)*. *Evolution* 40: 479–491. [28]

Kenett, R., and S. Zacks. 1998. *Modern industrial statistics: the design and control of quality and reliability.* Duxbury. [28]

Kendall, M., and A. Stuart. 1983. *The advanced theory of statistics. Volume 3: Design and analysis, and time-series*. 4th Ed. Griffin, London. [28]

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P.Gilbert, and P. Beerli. 2001. The strenght of phenotypic selection in natural populations. *Amer. Natl.* 157: 245–261. [28]

Kingsolver, J. G., and D. W. Pfenning. 2007. Patterns and power of phenotypic selection in nature. *BioScience* 57: 561–572. [28]

Klein, J. P., and , M. L. Moeschberger. 1997 *Survival analysis: techniques for censored and truncated data.* Springer. [28]

Kirkpatrick, M., and R. Lande. 1989. The evolution of maternal characters. *Evolution* 43: 485–503. [28]

Koenig, W. D., and S. S. Albano. 1986. On the measurement of sexual selection. *Amer. Nat.* 127: 403–409. [28]

Koenig, W. D., and S. A. Albano. 1987. Lifetime reproductive success, selection, and the opportunity for selection in the white-tailed skimmer *Plathemis lydia* (Odonta: Libellulidae). *Evolution* 41: 22–36. [28]

Koenig, W. D., S. A. Albano, and J. L .Dickinson. 1991. A comparison of methods to partition selection via components of fitness: do larger male bullfrogs have greater hatching success? *J. Evol. Biol.* 4: 309–320. [28]

Kozlowski, J. 1993. Measuring fitness in life-history studies. *Trends Ecol. Evol.* 8: 84–85. [28]

Lande, R. 1982. A quantitative genetic theory of life history evolution. *Ecology* 63: 607–615.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37: 1210–1226. [28]

Lawless, J. F. 2003. *Statistical models and methods for lifetime data*, 2nd Ed. Wiley, New York. [28]

Lenski, E. E. and P. M. Service. 1982. The statistical analysis of population growth rates calculated from schedules for survivorship and fecundity. *Ecology* 63: 655–662. [28]

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. *Biometrika* 33: 183–212. [28]

Leslie, P. H. 1948. Some further notes on the use of matrices in certain population mathematics. *Biometrika* 35:213–245. [28]

Little, R. J. A., abd D. B. Rubin. 2002. *Statistical analysis with missing data*, 2nd Ed. Wiley, New YOrk. [28]

Mackenzia, A., J. D. Reynolds, V. J. Brown, and W. J. Sutherland. 1995. Variation in male mating success on leks. *Amer. Natl.* 145: 633–652. [28]

Manly, B. F. J. 1985. *The statistics of natural selection*. Chapman and Hall, London. [28]

Manly, B. F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology, 2nd edition*. Chapman and Hall, London. [28]

McGraw, J. B., and H. Caswell. 1996. Estimation of individual fitness from life-history data. *Amer. Natl.* 147: 47–64. [28]

Metcalf, C. J., and S. Pavard. 2007. Why evolutionary biologists should be demographers. *Tends Ecol. Evol.* 22: 205–212. [28]

Metz, J. A. J., R. M. Nisbet, and S. A. H. Green. 1992. How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evol.* 7: 198–202. [28]

Miller, R. G. 1974. The jackknife — a review. *Biometrika* 61: 1–15. [28]

Mitchell-Olds, T., and J. Bergelson. 1990. Statistical genetics of *Impatiens capensis.* II. Natural selection. *Genetics* 124: 417–421. [28]

Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41: 1149–1161. [28]

Mitchell-Olds, T. and R. G. Shaw. 1990. Comments on the causes of natural selection. *Evolution* 44: 2158. [28]

Moore, A. J. 1990. The evolution of sexual dimorphism by sexual selection: the separate effects of intrasexual selection and intersexual selection. *Evolution* 44: 315–331. [28]

Nishida, T. 1989. Is lifetime data always necessary for evaluating the "intensity" of selection? *Evolution* 43: 1826–1827.

O'Donald, P. 1970. Change of fitness by selection for a quantitative character. *Theor. Pop. Biol.* 1: 219–232. [28]

O'Donald, P. 1971. Natural selection for quantitative characters. *Heridity* 27: 137–153. [28]

Partridge, L., and P. H. Harvey. 1988. The ecological context of life history evolution. *Science* 241: 1449–1455. [28]

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. Roy. Soc. Lond.* A 200: 1–66. [28]

Picard, R. R. and R. D. Cook. 1984. Cross-validation of regression models. *Technometrics* 79: 575–583. [28]

Postma, E. 2006. Implications of the differences between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol*. 19: 309–320. [28]

Preziosi, R. F., and D. J. Fairbairn. 2000. Lifetime selection on adult body size and components of body size in a waterstrider: opposing selection and the maintenance of sexual size dimorphism. *Evolution* 54: 558–566. [28]

Price, T. D., and P. R. Grant. 1984. Life history traits and natural selection for small body size in a population of Darwin's finches. *Evolution* 38: 483–494. [28]

Price, T. D., P. R. Grant, H. L. Gibbs, and P. T. Boag. 1984. Recurrent patterns of natural selection in a population of Darwin's finches. *Nature* 309: 787–789. [28]

Primack, R. B., and H. Kang. 1989. Measuring fintess and natural selection in wild plant populations. *Annu. Rev. Ecol. Sys.* 20: 367–396. [28]

Prout, T. 1965. The estimation of fitness from genotypic frequencies. *Evolution* 19: 546–551. [28]

Prout, T. 1969. The estimation of fitness from population data. *Genetics* 63: 949–967. [28]

Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46: 616–626. [28]

Scheiner, S. M., K. Donohue, L. A. Dorn, S. J. Mazer, and L. M. Wolfe. 2002. Reducing environmental bias when measuring natural selection. *Evolution* 56: 2156–2167. [28]

Schluter, D. 1988. Estimating the form of natural selection on a quantitiative triat. *Evolution* 42: 849–861. [28]

Shao, J. and D. Tu. 1996. *The Jackknife and Bootstrap.* Springer. [28]

Sino, B., and A. L. Basolo. 1996. Testing adaptation using phenotypic manipulation. *in* M. R. Rose an G. V. Lauder (eds) si Adaptation, pp. 149–185.. Academic Press, New York. [28]

Sokal, R. R. and J. F. Rohlf 1994. *Biometry: The Principles and Practices of Statistics in Biological Research*. W. H. Freeman [28]

Snee, R. D. 1977. Validation of regression models: methods and examples. *Technometrics* 19: 415–428. [28]

Sprent. P. , and N. C. Smeeton. 2007. *Applied nonparametric statistical methods, 4th edition*. Chapman and Hall, New York. [28]

Stearns, S. C. 1976. Life-history tactics: a review of the idea. *Q. Rev. Biol.* 51: 3–47. [28]

Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Ruahser, and R. Mauricio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Amer. Natl.* 160: 511–523. [28]

Sutherland, W. J. 1985a. Chance can produce a sex difference in variance in mating success and explain Bateman's data. *Anim. Behav.* 33: 1349–1352. [28]

Sutherland, W. J. 1985b. Measures of sexual selection. *Oxford Surveys Evol. Bio.* 2: 90–101. [28]

Sved, J. A. 1989. The measurement of fitness in *Drosophila*. *In* W. G. Hill, and T. F. C. Mackay, (eds.), *Evolution and animal breeding*, pp. 113–120. C. A. B. International, Wallingford. [28]

Trail, P. W. 1985. The intensity of selection: intersexual and interspecific comparisons require consistent measures. *Amer. Nat.* 126: 434–439. [28]

Travis, J. and S. Henrich. 1986. Some problems in estimating the intensity of selection through fertility differences in natural and experimental populations. *Evolution* 40: 786–790. [28]

Travis, J., and D. N. Reznick. 1998. Experimental approaches to the study of evolution. *In* J. Resitartis and J. Bernardo (eds) *Issues and perspectives in experimental biology*, pp. 437–459. Oxford, New York. [28]

Tukey, J. W. 1958. Bias and confidence in not quite large samples (Abstract). *Ann. Math. Stat.* 29: 614. [28]

Wade, M. J. and S. Kalisz. 1989. The additive partitioning of selection gradients. *Evolution* 43: 1567–1569. [28]

Weldon, H. F. R. 1901. A first study of natural selection in *clausilia laminata* (Montagu). *Biometrika* 1: 109–124. [28]

Willis, J. H. 1996. Measures of phenotypic selection are biased by partial inbreeding. *Evolution* 50: 1501–1511 [28]

Wilson, A. J., J. G. Pilkington, J. M. Pemberton, D. W. Coltman, A. D. J. Overall, K. A. Byrne, and L. E. B. Kruuk. 2005. Selection on mothers and offspring: whose phenotype is it and does it matter? *Evolution* 59: 451–463. [28]

Wolf. J. B., and M. J. Wade. 2001. On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *J. Evol. Biol.* 14: 347–356. [28]

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Stat.* 14: 1261–1295. [28]

Zuk, M. 1988. Parasite load, body size, and age of wild-caught male field crickets (Orthoptera: Gryllidae): effects on sexual selection. *Evolution* 42: 696–976. [28]