In summary, when we consider just one gene, the means of the six basic generations are,

$$\begin{aligned}
\bar{P}_1 &= m + a_A \\
\bar{P}_2 &= m - a_A \\
\bar{F}_1 &= m + d_A \\
\bar{F}_2 &= m + \tfrac{1}{2}d_A \\
\bar{B}c_{1.1} &= m + \tfrac{1}{2}a_A + \tfrac{1}{2}d_A \\
\bar{B}c_{1.2} &= m - \tfrac{1}{2}a_A + \tfrac{1}{2}d_A.
\end{aligned} \qquad \text{[Eqns 2.1]}$$

2.2 *Two gene model with additive and dominance effects*

Extension of the single gene model to generations derived from parents which differ at two genes, say A and B, is relatively simple because the expectations derived above for gene A, apply to gene B also. For example, the previous expectation of the F_2 mean,

$$m + \tfrac{1}{2}d_A$$

is now modified to

$$m + \tfrac{1}{2}d_A + \tfrac{1}{2}d_B$$

to accommodate the effects of segregation at the B gene. There is, however, one complication which affects both the parental and the backcross scores and this concerns the distribution of alleles in the parents. The + and − alleles could either be associated,

$$\text{i.e.} \quad P_1 = A^+A^+B^+B^+, \quad P_2 = A^-A^-B^-B^-$$

or dispersed,

$$\text{i.e.} \quad P_1 = A^+A^+B^-B^-, \quad P_2 = A^-A^-B^+B^+$$

and this situation affects the composition and size of the additive component which would be equal to $a_A + a_B$ for association or $a_A - a_B$ for dispersion. By definition, both a_A and a_B are positive, therefore the value of the additive component will obviously be larger if the genes are in association than if they are in dispersion.

The mean and dominance components, on the other hand, will remain independent of gene association and dispersion. The former will still be equal to m because all the parameters (a_A, a_B, d_A and d_B) are defined as deviations from the mid-parent, while the latter will

become $d_A + d_B$. The dominance component $(d_A + d_B)$ will represent a net balance between the values of d_A and d_B because d may in some cases be positive and in others negative. Such a situation of positive and negative dominance at different genes is referred to as **ambidirectional dominance**, while dominance in one direction is called **unidirectional dominance**.

2.3 *Multiple gene model with additive and dominance effects*

One feature that is apparent from the one and two gene models is that whatever the number of genes, their additive and dominance effects will be measured by just two parameters. The effects of the individual genes will therefore be hidden in these parameters and it is difficult to visualize how the individual gene effects will be combined when there are several genes. This is particularly true of the additive genetic effects which will be affected by the distribution of the alleles and, by way of illustration, we will initially consider a special case of four genes controlling a trait. This will allow us to determine some general principles which can then be extended to accommodate any number of genes in our model. For simplicity, let us assume that the four genes are A, B, C and D, and that each gene has an additive genetic effect of $a = 2$ units.

Table 2.2 shows all possible combinations in which the alternative alleles can be present in P_1 and P_2 and gives the detailed composition of the additive component in terms of the gene effects. However, we can present the expectations of the additive genetic effect in a more general and meaningful form which is based on just two parameters: (i) half the difference between the parental means $(a_A + a_B + a_C + a_D)$ when there is complete association; and (ii) half the additive effects of those genes that are dispersed in the parents. For example, we can write the expectation of the additive genetic effect for situation (b) in Table 2.2 as $(a_A + a_B + a_C + a_D) - 2(a_D)$, or situation (g) as $(a_A + a_B + a_C + a_D) - 2(a_B + a_D)$. In other words, we can write the expectations of $\bar{P}_1$ and $\bar{P}_2$ in terms of (i) the total effects of the k genes and (ii) those (k') genes that are dispersed between the parents, as:

$$\bar{P}_1 = m + \sum^{k-k'} a_i - \sum^{k'} a_i \quad \text{or} \quad m + \sum^{k} a_i - 2\sum^{k'} a_i;$$

$$\bar{P}_2 = m - \sum^{k-k'} a_i + \sum^{k'} a_i \quad \text{or} \quad m - \sum^{k} a_i + 2\sum^{k'} a_i.$$

Table 2.2 The effect of gene dispersion on the composition and size of the additive genetic component when there are four genes

	A	B	C	D	Score	Expected value of additive component
(i) Complete association						
(a) P_1	++	++	++	++	$2+2+2+2=m+8$	$a_A+a_B+a_C+a_D=8$
P_2	−−	−−	−−	−−	$-2-2-2-2=m-8$	
(ii) Partial (one gene) dispersion						
(b) P_1	++	++	++	−−	$2+2+2-2=m+4$	$a_A+a_B+a_C-a_D=4$
P_2	−−	−−	−−	++	$-2-2-2+2=m-4$	
or						
(c) P_1	++	++	−−	++	$2+2-2+2=m+4$	$a_A+a_B-a_C+a_D=4$
P_2	−−	−−	++	−−	$-2-2+2-2=m-4$	
or						
(d) P_1	++	−−	++	++	$2-2+2+2=m+4$	$a_A-a_B+a_C+a_D=4$
P_2	−−	++	−−	−−	$-2+2-2-2=m-4$	
or						
(e) P_1	−−	++	++	++	$-2+2+2+2=m+4$	$-a_A+a_B+a_C+a_D=4$
P_2	++	−−	−−	−−	$2-2-2-2=m-4$	
(iii) Complete (two gene) dispersion						
(f) P_1	++	++	−−	−−	$2+2-2-2=m+0$	$+a_A+a_B-a_C-a_D=0$
P_2	−−	−−	++	++	$-2-2+2+2=m-0$	
or						
(g) P_1	++	−−	++	−−	$2-2+2-2=m+0$	$+a_A-a_B+a_C-a_D=0$
P_2	−−	++	−−	++	$-2+2-2+2=m-0$	
or						
(h) P_1	−−	++	++	−−	$-2+2+2-2=m+0$	$-a_A+a_B+a_C-a_D=0$
P_2	++	−−	−−	++	$+2-2-2+2=m-0$	

Earlier we specified that P_1 should always take the larger score, which means that,

$$\sum^{k} a_i \text{ must always be } \geq 2\sum^{k'} a_i.$$

We now further simplify the equations above by defining a **coefficient of gene association/dispersion** [2], r_a, where:

$$r_a = \left(\sum^{k} a_i - 2\sum^{k'} a_i\right) \Big/ \sum^{k} a_i.$$

Thus, $r_a = 1$ when there is complete association (as in Table 2.2(a)), i.e.

$$k' = 0$$

Genetic of Eye color in Human

There are several classes of eye color in man which produce continous variation. The difference in eye color is due to amount of melanin pigment in iris. The light blue eyes have the smallest amount of melanin, but the dark brown (black) eyes have the largest amount of pigment. However, there is no melanin in albino eyes

No. of active alleles in genotype	***Color of human eye***
0	Light blue
1	Medium blue
2	Dark blue
3	Gray
4	Green
5	Hazel
6	Light brown
7	Medium brown
8	Dark brown (black)

According to multiple-gene hypothesis, inheritance of this quantitative trait in human could be due to several pairs of genes. According to that

hypothesis, number of phenotypes in natural population or in F2 from crosses of animals or plants as following:- No. of phenotypes (9) 9=2n+1

n = 4 pairs of genes

So we can see that active alleles in each of these classes, (above) the first color has no alleles (active alleles), and as order of color increase the number of active alleles increases until black contains the largest number of active alleles, because of that, the amount of pigment in this class is more than other classes .

Transgressive Variation

With respect to this trail (cye color), it is observed in some families,that children are born with lighter or darker eye color than their parents . Another example is stature in some families, the children are born then become adult could be taller than their parents and their grandparents. In addition, there are many examples in animals and plants, in which some individuals express quantitative character more or less than their parents . One experiment in chickens, two inbred lines were crossed these two lines are different in body weight). The first line is called Sebright Beniam (700g), the second line is Golden hambrg (1100g). Punnett observed that F1 mean weight was (900g). In F2, there were many classes are few birds in F2 show extremes in body weight, the first, smaller than the small parent and the last extreme was larger than the big (heavy) parent. So, what is the genetic explanation of the results which differ from previous observations about expression of quantitative character ?

Can we explain these results on bases of over dominance This kind of dominance we can see in F1, That means that we obtain one extreme in F1, so we cannot explain these results on bases of over dominance

Punnet explained these results on base of transgressive variation
Suppose that the genotypes of parent were :-

P aa bb cc DD x AA BB CC dd

700g 1100g

F1 Aa Bb Cc Dd

900g

The segregation happened in F1 as well as other generations. From this segregation, there is probability of getting genotypes aa bb cc dd and AA BB CC DD, in addition there are many other genotypes in F2 :-

F2 aa bb cc dd , ……., AA BB CC DD

500g 1300g

The first genotype, (according to active alleles) will produce body weight smaller than the first parent. The last genotype (other extreme) will produce body weight more than the second parental line. We can calculate the contribution of one active allele in this experiment. As you see, the parents are different with two pairs of genes, difference between them in body weight (two parents) related to four active alleles

$$a = D/2n = 400/4 = 100g$$

So we expect that the first extreme is 500g, the second will be 1300g In F1 we observe that a segregation happened for (4) pairs of genes, while in parents, the reason of difference with weight is due to four active alleles .

Then this phenomenon (transgressive variation), we can make benefits it in animals and plants. This phenomenon also explain what observed about some quantitative traits in different families. For example, in some families, some children, their intelligence more than intelligence of their parents and their grand parents, while this phenomenon is not observed in qualitative traits. For example, there are no individuals, their blood groups more or less antigens than either parents .

$r_a = 0$ when there is complete dispersion (as in Table 2.2 (f to h)), i.e.

$$\sum^{k} a_i = 2\sum^{k'} a_i$$

and $0 < r_a < 1$ when there is partial association (as in Table 2.2 (b to e)), i.e.

$$\sum^{k} a_i > 2\sum^{k'} a_i.$$

We are now in a position to formulate a general definition of the additive component as,

$$[a] = r_a . \sum^{k} a_i.$$

The square brackets are used to denote that it is the net balance of additive genetic effects over all the genes which is being observed, after internal cancellations due to dispersion. The genetic values of the two parental lines thus become:

$$\bar{P}_1 = m + r_a . \sum^{k} a_i = m + [a]$$

$$\bar{P}_2 = m - r_a . \sum^{k} a_i = m - [a].$$

The genotype of the F_1 hybrid, on the other hand, will always be $A^+A^-B^+B^- \ldots K^+K^-$ irrespective of the degree of gene dispersion/association in the parents. Thus, the F_1 score will be equal to $m + d_A + d_B + d_C + \ldots + d_K$, which can be summarized as:

$$\bar{F}_1 = m + \sum^{k} d_i = m + [d].$$

Parameter $[d]$ represents the net balance of the dominance effects and indicates the direction of dominance at the majority of the k genes, weighted by the magnitudes of their effects.

We can illustrate the application of this simple model by means of the data for P_1, P_2 and F_1 from Table 2.1. Thus,

$$m = \tfrac{1}{2}(\bar{P}_1) + \tfrac{1}{2}(\bar{P}_2) = \tfrac{1}{2}(69.44) + \tfrac{1}{2}(59.04) = 64.24$$

$$[a] = \tfrac{1}{2}(\bar{P}_1) - \tfrac{1}{2}(\bar{P}_2) = \tfrac{1}{2}(69.44) - \tfrac{1}{2}(59.04) = 5.20$$

$$[d] = \bar{F}_1 - \tfrac{1}{2}(\bar{P}_1) - \tfrac{1}{2}(\bar{P}_2) = 83.44 - \tfrac{1}{2}(69.44) - \tfrac{1}{2}(59.04) = 19.20.$$

The fact that $[a]$ is relatively small could imply either that the two parents hardly differ genetically for the trait under study or

that there is gene dispersion, but the large positive value of $[d]$ clearly indicates that the parents are genetically different since $[d]$ is due solely to the heterozygosity of those genes for which the parents differ. A more detailed discussion of the interpretation and uses of these estimates will be deferred to a later section of this chapter.

2.4 *Extension to other generations*

The model we have just developed can easily be extended to other generations and the expectations of the six basic generations for a multiple gene situation are:

$$\begin{aligned}
\bar{P}_1 &= m + [a] \\
\bar{P}_2 &= m - [a] \\
\bar{F}_1 &= m + [d] \\
\bar{F}_2 &= m + \tfrac{1}{2}[d] \\
\bar{B}c_{1.1} &= m + \tfrac{1}{2}[a] + \tfrac{1}{2}[d] \\
\bar{B}c_{1.2} &= m - \tfrac{1}{2}[a] + \tfrac{1}{2}[d].
\end{aligned} \qquad \text{[Eqns 2.2]}$$

These are clearly analogous to those given earlier for a single gene (Equations 2.1). There are, however, other generations frequently produced by breeders and geneticists as a result of selfing or crossing F_2 or backcross individuals. For example, if we consider a single gene case when we self a random sample of the F_2 individuals, we obtain the following types of F_3 families:

Family	Frequency	Progeny genotypes	Mean
A^+A^+ self	$\frac{1}{4}$	all A^+A^+	$m + a_A$
A^+A^- self	$\frac{1}{2}$	$\frac{1}{4}A^+A^+$: $\frac{1}{2}A^+A^-$: $\frac{1}{4}A^-A^-$	$m + \frac{1}{2}d_A$
A^-A^- self	$\frac{1}{4}$	all A^-A^-	$m - a_A$

Thus, for a single gene difference, the F_3 mean has the expectation,

$$\tfrac{1}{4}(m + a_A) + \tfrac{1}{2}(m + \tfrac{1}{2}d_A) + \tfrac{1}{4}(m - a_A) = m + \tfrac{1}{4}d_A.$$

The mathematical relationship between F_1, F_2 and F_3 can be used to derive the expectation of any generation, F_n, which is obtained by applying $n - 1$ rounds of selfing to the above F_1 and the mean will

be equal to:

$$\bar{F}_n = m + \left(\frac{1}{2}\right)^{n-1} d_A.$$

When a random sample of $Bc_{1.1}$ individuals are again crossed to P_1 giving the second generation ($Bc_{2.1}$) of recurrent backcrossing, this generation, like the F_3, has a family structure as follows:

Family	Frequency	Progeny genotype	Mean
$A^+A^+ \times A^+A^+$	$\frac{1}{2}$	all A^+A^+	$m + a_A$
$A^+A^- \times A^+A^+$	$\frac{1}{2}$	$\frac{1}{2}A^+A^+$: $\frac{1}{2}A^+A^-$	$m + \frac{1}{2}a_A + \frac{1}{2}d_A$

Now, $\bar{B}c_{2.1} = \frac{1}{2}(m + a_A) + \frac{1}{2}(m + \frac{1}{2}a_A + \frac{1}{2}d_A) = m + \frac{3}{4}a_A + \frac{1}{4}d_A$. In the case of repeated backcrossing to P_1 for n generations the generation mean has the expectation:

$$\bar{B}c_{n.1} = m + \left(1 - \left(\frac{1}{2}\right)^n\right)a_A + \left(\frac{1}{2}\right)^n d_A.$$

By following the same approach we can derive the formulae for the first and the subsequent recurrent (n) backcrosses of P_2 (A^-A^-) to F_1 (represented by the symbol $Bc_{1.2}$ and $Bc_{n.2}$ respectively) and they are as follows:

$$\bar{B}c_{1.2} = m - \tfrac{1}{2}a_A + \tfrac{1}{2}d_A$$

$$\bar{B}c_{n.2} = m - \left(1 - \left(\frac{1}{2}\right)^n\right)a_A + \left(\frac{1}{2}\right)^n d_A$$

(where n = 1 for $Bc_{1.2}$).

Although we have derived these means for a single gene case, they translate simply into the multi-gene case and the general formulae for the expectations of other generations of the selfing and recurrent backcrossing series are:

$$\bar{F}_n = m + \left(\frac{1}{2}\right)^{n-1}[d];$$

$$\bar{B}c_{n.1} = m + \left(1 - \left(\frac{1}{2}\right)^n\right)[a] + \left(\frac{1}{2}\right)^n [d] \quad \text{and} \qquad \text{[Eqns 2.3]}$$

$$\bar{B}c_{n.2} = m - \left(1 - \left(\frac{1}{2}\right)^n\right)[a] + \left(\frac{1}{2}\right)^n [d].$$

Elaborations of these formulae for some of the frequently used generations in quantitative genetic studies are given in Table 2.3.

Table 2.3 Expectations of various generations on an additive-dominance model

Generation	m	$[a]$	$[d]$
P_1	1	1	0
P_2	1	−1	0
F_1	1	0	1
F_2	1	0	0.5
F_3	1	0	0.25
F_4	1	0	0.125
F_∞	1	0	0
$Bc_{1.1}$	1	0.5	0.5
$Bc_{2.1}$	1	0.75	0.25
$Bc_{1.2}$	1	−0.5	0.5
$Bc_{2.2}$	1	−0.75	0.25

It is clear from the above formulae that, with directional dominance, i.e. $[d] > 0$, the F_n mean will change with each generation of selfing, eventually reaching the mid-parental value, m. In the case of a cross exhibiting **hybrid vigour** or **heterosis**, that is where the F_1 mean exceeds the mean of the better parent, this decline in mean due to selfing is commonly referred to as **inbreeding depression**. This is well illustrated by the character final plant height in *Nicotiana rustica* (Figure 2.1).

2.5 *Relationships between generation means*

It is apparent from the theory so far that there are very simple relationships between the expected means of different generations. For example, the mean of $Bc_{1.1}$ has the expectation $m + \frac{1}{2}[a] + \frac{1}{2}[d]$ which is equal to the average of the means of the P_1 and F_1 generations. Similarly, the expectation of the $Bc_{1.2}$ mean is equal to the average of the means of P_2 and F_1. In fact, the expected mean of any generation that is derived from a cross between two pure breeding parents bears a simple relationship with the parental and F_1 means.

However, these expected relationships hold only if the generation means depend solely on the additive and dominance effects of genes. They would not be expected to hold if other factors such as differential viability, maternal effects or interactions between genes (often referred to as **epistasis** or **non-allelic interaction**) existed. Therefore, comparisons among these generation means can be used to provide tests

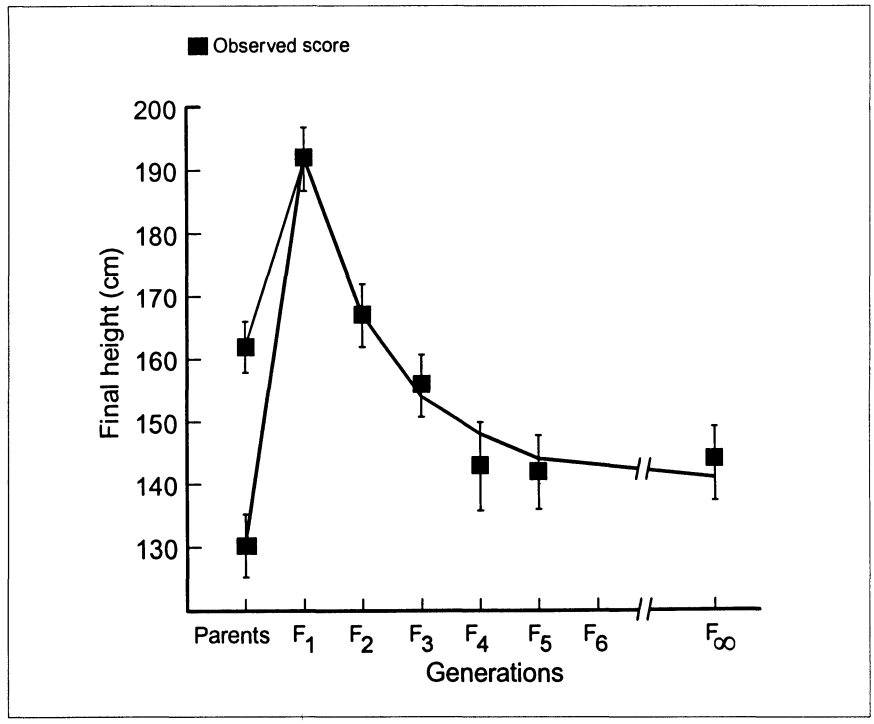

Figure 2.1 Heterosis and inbreeding depression for height in a cross between two inbred lines of *Nicotiana*. The solid line shows the expected decline in the mean with inbreeding.

for the presence of such complex factors. For example, when the additive–dominance model is adequate we expect:

$$\bar{B}c_{1.1} = \tfrac{1}{2}\bar{P}_1 + \tfrac{1}{2}\bar{F}_1$$

or

$$2\bar{B}c_{1.1} = \bar{P}_1 + \bar{F}_1$$

or

$$2\bar{B}c_{1.1} - \bar{P}_1 - \bar{F}_1 = 0.$$

On the other hand, if the additive–dominance model is inadequate, then it is likely that,

$$2\bar{B}c_{1.1} - \bar{P}_1 - \bar{F}_1 = \mathbf{A}.$$

Using the data in Table 2.1, for example, **A** is estimated as,

$$2 \times 76.03 - 69.44 - 83.44 = -0.82.$$

Clearly **A** is not zero, but in order to test if the additive–dominance model has failed in this case we need to know whether the estimate of **A** is significantly different from zero. In order to do this, we need the standard error of **A**.

Now, **A** is a linear sum of three means. It follows from basic statistical theory, that if some quantity, **Q**, is the sum of k_1y_1, k_2y_2, k_3y_3, etc. where the k_is are constants (i.e. $\mathbf{Q} = \sum k_i y_i$), then the

variance (s^2) of **Q** is,

$$s_{\mathbf{Q}}^2 = \sum k_i^2 s_{yi}^2,$$

providing the y_is are independent.

For **A**, the k_is are 2, −1 and −1 respectively, while the y_is are the generation means and hence the s_{yi}^2s are the variances of the relevant generation means. In Table 2.1 we have given the variance among the individuals of the various generations, s^2, and to obtain the variance of the generation means we need to divide these by the appropriate family size, *n*; e.g. for $Bc_{1.1}$, we obtain 0.8105 (= 81.05/100). The variance of **A** is therefore,

$$s_{\mathbf{A}}^2 = 4s_{\bar{Bc}_{1.1}}^2 + s_{\bar{P}_1}^2 + s_{\bar{F}_1}^2$$

$$s_{\mathbf{A}}^2 = 4 \times 0.8105 + 0.5973 + 0.5181 = 4.3574.$$

We can now apply Student's t test where,

$$t_{(df)} = \mathbf{A}/\sqrt{s_{\mathbf{A}}^2}.$$

The **degrees of freedom** (df) of t are the sum of the df of the three variances, 297.

$$t_{(297)} = -0.82/\sqrt{4.3574}$$

$$= -0.39 \text{ n.s.}$$

This shows that **A** is not significantly different from zero and hence there is no reason to reject the null hypothesis (H_0). We therefore conclude that the additive–dominance model adequately explains the variation among these means.

Mather [3] derived the above relationship and called it the '**A** Scaling Test'. He described two more scaling tests that are appropriately called **B** and **C** and they determine the conformity of the following relationships with the additive–dominance model.

$$\mathbf{B} = 2\bar{Bc}_{1.2} - \bar{P}_2 - \bar{F}_1 \qquad s_{\mathbf{B}}^2 = 4s_{\bar{Bc}_{1.2}}^2 + s_{\bar{P}_2}^2 + s_{\bar{F}_1}^2$$

$$\mathbf{C} = 4\bar{F}_2 - 2\bar{F}_1 - \bar{P}_1 - \bar{P}_2 \qquad s_{\mathbf{C}}^2 = 16s_{\bar{F}_2}^2 + 4s_{\bar{F}_1}^2 + s_{\bar{P}_1}^2 + s_{\bar{P}_2}^2.$$

Using the same procedure as for **A**, we obtain:

$$\mathbf{B} = 0.08, s_{\mathbf{B}}^2 = 4.8084, t_{(297)} = 0.04, \text{n.s.}; \quad \text{and}$$

$$\mathbf{C} = 2.08, s_{\mathbf{C}}^2 = 19.4468, t_{(396)} = 0.47, \text{n.s.}$$

Since neither **A**, nor **B** nor **C** is significant in the present example, we can accept that complicating effects, such as maternal effects or gene

interaction, are not involved in the genetical control of the character in this cross. We can not say that such complicating factors do not exist, but simply that we have no evidence for their existence; the data are consistent with their absence and the additive–dominance model is adequate. It is possible that more complex factors are involved, but the experiment is not sufficiently large and powerful to detect them. This creates a dilemma. A very small experiment will invariably not detect any departures from a simple model while a very large experiment may be so sensitive that it may detect all sorts of trivial departures. We will return to this later. If any of the three scaling tests indicate that the simple additive–dominance model fails then it is necessary to consider the more complex situations described in Chapters 11, 12 and 13.

2.6 *Estimating genetical parameters*

If we have determined that the differences between generation means for a particular character are explained adequately by the additive and dominance effects of genes, then it is sensible to obtain reliable estimates of m, $[a]$ and $[d]$ and to test their significance before attempting their interpretation. Earlier we explained how these three parameters could be estimated when we had just P_1, P_2 and F_1 means, but we now have the problem of estimating three parameters from six generation means. Clearly, we would like to use all the genetical information that is available to us in order to obtain the most accurate estimates which have the smallest possible standard errors. In other words, we want maximum likelihood estimates of m, $[a]$ and $[d]$.

Such estimates can be obtained by regression procedures and so we need to translate our data into a form amenable to a regression analysis. Consider the data in Table 2.1 and the expectations given earlier in this section for the six basic generations as shown in Table 2.4. If, for simplicity, we consider parameters m and $[a]$ only, we can write the expectations of the various generation means (y_i) in the form:

$$y_i = c + bx_i$$

where the coefficients of $[a]$ are the x_is and the intercept, c, equals m and the regression coefficient, b, equals $[a]$ (i.e. $y_i = m + [a]x_i$). The x_is are, in fact, measures of the relative contribution of A^+ and A^- alleles to each generation. Thus, P_1 is all A^+ ($x_i = 1$), $Bc_{1.1}$ consists of $\frac{1}{2}A^+A^+$ and $\frac{1}{2}A^+A^-$ individuals and hence there are $\frac{3}{4}A^+$ to $\frac{1}{4}A^-$

Table 2.4 Means of six basic generations, together with m, $[a]$ and $[d]$ model to illustrate the use of regression for parameter estimation

Generation	Mean (y variable)	m	$[a]$ (x_1 variable)	$[d]$ (x_2 variable)
P_1	69.44	1	1	0
$Bc_{1.1}$	76.03	1	$\frac{1}{2}$	$\frac{1}{2}$
F_1	83.44	1	0	1
F_2	74.36	1	0	$\frac{1}{2}$
$Bc_{1.2}$	71.28	1	$-\frac{1}{2}$	$\frac{1}{2}$
P_2	59.04	1	-1	0

($x_i = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$), F_1 is A^+A^- ($x_i = \frac{1}{2} - \frac{1}{2} = 0$), etc. If there is no dominance, therefore, the generation means, y_i, will vary with the relative balance of A^+ to A^- (x_i), which is sometimes called the **gene dosage**.

Using linear regression analysis with the data in Table 2.4, using x_i, we obtain the estimates,

$$c = m = 72.27$$

$$b = [a] = 5.11.$$

Statistically, these estimates are those which minimize deviations between the observed and the predicted values and also they are independent of each other. We can test the significance of each estimate by a Student's t-test. We can also determine whether the parameters have accounted for all of the significant differences between the generation means by testing the significance of the remainder mean square (MS) in the analysis of variance (ANOVA). When the remainder MS is significant, it indicates that there are other factors, apart from m and $[a]$, which affect the means, and these need to be identified and included in the model. In these circumstances, we need to extend the analysis to include several regression variables and linear regression then converts into multiple regression. This takes the form:

$$y_i = c + b_1.x_{1i} + b_2.x_{2i} + \ldots,$$

where $b_1 = [a]$, $b_2 = [d]$ etc.

So far we have assumed that each of the y_i values (generation means) is known with equal precision. This implies that the variances of the generation means (s_i^2/n_i) are all the same, but this is not generally likely in practice because some generations, e.g. F_2, will have much larger variances among individuals (s^2) than will others due to genetic segregation (see Chapter 3), as the data from Table 2.1

show, while there may also be large differences in family size (n_i). This heterogeneity among variances of the generation means makes the accuracy of the means unequal, which must be adjusted in the regression analysis by **weighting** the means differently. These weights are the reciprocals of the variances of the generation means [4], i.e. for the ith generation mean,

$$\text{weight}(wt_i) = \text{family size}(n_i)/\text{variance}(s_i^2).$$

The regression equation now becomes:

$$wt_i y_i = wt_i(c + b_1.x_{1i} + b_2.x_{2i} \ldots \text{etc.}).$$

Most statistical software packages available for modern computers should be capable of handling such weighted regression problems, but for those who wish to understand the methodology, a short theoretical explanation is given in Appendix F. Briefly, because there are N generation means, it is possible to estimate a total of N parameters, of which one is the mean, *m*. This leaves a maximum of $N-1$ genetical or other parameters that can be estimated, which is equivalent to the $N-1$ df between the N means. The object is to explain the variation between the observed generation means with as simple a model as possible. If only *m* is fitted, then there will be $N-1$ df remaining to test the adequacy of the model using a χ^2 or F test. Every time a further parameter is included in the model, there is one less df for testing the adequacy of the model.

Faced with a new set of data, how should the model fitting proceed? It always makes sense to try a model with just *m*, first. If this adequately explains the variation in the trait, then there is no need to fit any genetical parameters. If not, then further parameters should be introduced. For example, the next simplest model would involve the parameters *m* and [*a*]. If both parameters are significant and the model fits, then there is no need to proceed further. If both parameters are significant, but the χ^2 indicates that the model is still inadequate, then try a model comprising *m*, [*a*] and [*d*]. At each step, only significant parameters are retained in the model, while only sufficient parameters are added in total to provide an adequate fit to the data. It is quite possible for two different models to fit the data, in which case it is customary to accept the one with the fewest and biologically most plausible parameters. The process is illustrated in Table 2.5 with the data taken from Table 2.1.

With multiple regression situations, the parameters are seldom completely independent, i.e. they are correlated, and this means that the magnitude of a particular parameter will depend on those other

Table 2.5 The generation means, variances, weights and additive–dominance model for the six basic generations of a cross between two homozygous lines. (i) Basic data and model; (ii) parameters from sequential model fitting; (iii) best model

(i)

Generation	Mean	s_x^2	df	$s_{\bar{x}}^2$	Weight	m	$[a]$	$[d]$
P_1	69.44	59.73	99	0.5973	1.6742	1	1	0
P_2	59.04	65.71	99	0.6571	1.5218	1	−1	0
F_1	83.44	51.81	99	0.5181	1.9301	1	0	1
F_2	74.36	100.75	99	1.0075	0.9926	1	0	0.5
$Bc_{1.1}$	76.02	81.05	99	0.8105	1.2338	1	0.5	0.5
$Bc_{1.2}$	71.28	90.83	99	0.9083	1.1010	1	−0.5	0.5

(ii)

(a)	$m = 72.5420 \pm 0.3439^{***}$	$\chi^2_{(5)} = 542.73^{***}$
(b)	$m = 72.4172 \pm 0.3442^{***}$	
	$[a] = 4.8209 \pm 0.5148^{***}$	$\chi^2_{(4)} = 455.02^{***}$

(iii)

$m = 64.2474 \pm 0.5151^{***}$	
$[a] = 5.1252 \pm 0.5149^{***}$	
$[d] = 19.1989 \pm 0.9005^{***}$	$\chi^2_{(3)} = 0.47$ n.s.

parameters which are being estimated simultaneously. As additional parameters are added, it is to be expected, therefore, that the values of the parameters initially estimated will change. If the simple additive–dominance model fits, we can discount the presence of complicating factors such as maternal effects and interactions between genes. This approach thus provides a comprehensive test of the simple additive model which replaces the **A**, **B** and **C** scaling tests and has been termed the **Joint Scaling Test** [4].

2.7 *Interpretation: heterosis and potence ratio*

We have seen that it is possible to construct simple additive–dominance models for the means of generations derived from a pair of true breeding lines. We have also seen how it is possible to estimate the genetical parameters and test whether they alone are sufficient to explain the differences observed in the means of different generations. Where more complex models are required, such as with gene interaction, maternal effects, etc., the reader is referred to Chapters 11,

12 and 13. What insight, then, do these models and the parameters they generate give us into the genetical control of a trait?

Does $[a] = 0$ mean that there is no additive variation in the cross? No, not unless the parents differ for just a single gene (which is very unlikely), i.e. $[a] = a_A$. Similarly, we can conclude that there is no dominance only when $[d] = d_A = 0$. Further, for a single gene only, we can interpret $[d]/[a] = d_A/a_A$ as the **dominance ratio** and conclude the following:

if $d_A/a_A = +1$, allele A^+ is completely dominant to allele A^-;

if $d_A/a_A = -1$, allele A^- is completely dominant to allele A^+;

if $0 < d_A/a_A < +1$, allele A^+ is partially dominant to allele A^-;

if $-1 < d_A/a_A < 0$, allele A^- is partially dominant to allele A^+;

if $d_A/a_A = 0$, there is no dominance;

if $d_A/a_A > +1$, allele A^+ is over-dominant to allele A^-;

if $d_A/a_A < -1$, allele A^- is over-dominant to allele A^+.

With two or more genes, on the other hand, the situation becomes more complicated because

$$[d]/[a] = \left(\sum_{i=1}^{k} d_i\right) \Big/ \left(r_a \sum_{i=1}^{k} a_i\right).$$

The numerator could be zero as a result of ambi-directional dominance while the denominator could be zero as a result of gene dispersion. Thus, this ratio could take almost any value irrespective of the true degree of dominance. It is sometimes referred to as the **potence ratio** because it indicates which parent has the most dominant alleles and is therefore the more potent in the cross. Therefore, we can rarely equate the ratio $[d]/[a]$ with the true dominance ratio which is obtained only when dominance is uni-directional and there is complete association of alleles in the parents ($r_a = 1$). In other words, the true magnitude of the dominance ratio,

$$\left(\sum_{i=1}^{k} |d_i|\right) \Big/ \left(\sum_{i=1}^{k} a_i\right)$$

can only be estimated from the components of means when dominance is uni-directional at all the k genes for which the parents differ (i.e. $d_A, d_B \ldots d_K$, etc. have the same sign and $|[d]| = |d_A| + |d_B| \ldots + |d_K|$) and all the alleles with increasing effect are

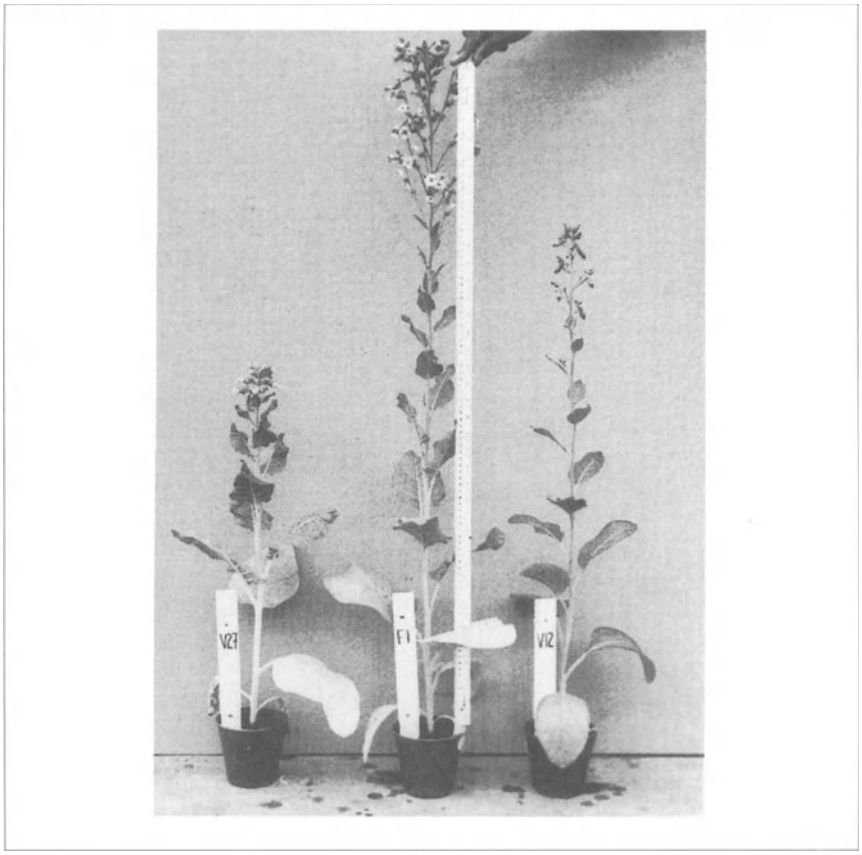

Figure 2.2 Heterosis for height in *Nicotiana rustica*.

present in P_1 (i.e. $\bar{P}_1 = m + a_A + a_B + \ldots + a_K$) and their counterparts are in P_2 ($\bar{P}_2 = m - a_A - a_B + \ldots - a_K$).

These conclusions concerning dominance are also relevant to interpreting the causes of **heterosis** or **hybrid vigour**. This phenomenon is of central importance in breeding and in evolutionary genetics, and concerns the superiority of the F_1 over the better parent, P_1 (see Figures 2.1 and 2.2). In the absence of gene interaction and other complicating factors, for which the tests have been described earlier, heterosis implies $\bar{F}_1 > \bar{P}_1$, i.e.

$$m + [d] > m + [a]$$

or

$$[d] > [a]$$

or

$$\sum d_i > r_a \sum a_i.$$

This relationship tells us that, as long as the average dominance ($\sum d_i / \sum a_i$) is greater than the degree of gene dispersion (r_a), then

there will be heterosis. It follows, therefore, that it requires very little dominance at individual genes to produce quite considerable heterosis if the genes are dispersed in the two parents. For example, if the better parent has 70% of the increasing alleles, then r_a will be $0.4(= 1 - 2 \times 0.3)$ and hence the average amount of dominance needs to be just greater than 0.4 to cause the F_1 to out-perform the better parent. If the better parent has 55% of the increasing alleles, the average dominance needs to be little more than 0.15. Clearly, characters such as fitness or yield, that are controlled by many genes, would require only very small amounts of dominance to produce very major hybrid superiority without the need to invoke overdominance. This also implies that it should be possible to produce lines that out-yield the F_1 hybrid and this possibility will be discussed in Chapter 15.

Summary

1. Six basic generations derived from a cross between two inbred lines, P_1, P_2, F_1, F_2, $Bc_{1.1}$ and $Bc_{1.2}$, appear in a whole range of experimental designs and their genetic models are central to quantitative genetics.
2. The means of these six basic generations can be described in terms of just three parameters which measure the mean, the additive and the dominance effects, m, $[a]$ and $[d]$, respectively.
3. It is possible to use these generations to provide powerful tests of the adequacy of a simple genetical model and, in particular, to test for complex effects such as epistasis, maternal effects, etc.
4. The parameterization of the additive and dominance effects shows that considerable heterosis can result from small amounts of dominance.

References

1. Fisher, R.A., Immer, F.R. and Tedin, O. (1932) The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics*, **17**, 107–24.

2. Mather, K. and Jinks, J.L. (1982) *Biometrical Genetics*, 3rd edn, Chapman & Hall, London.
3. Mather, K. (1949) *Biometrical Genetics*, 1st edn, Methuen, London.
4. Cavalli, L.L. (1952) An analysis of linkage in quantitative inheritance, in *Quantitative Inheritance* (eds E.C.R. Reeve and C.H. Waddington), HMSO, London, pp. 135–44.

3 Basic generations – variances

Variation is central to the study of quantitative traits because little can be deduced from populations or generations that are monomorphic. From studies of variation among genotypes, various questions concerning the genetic control of a trait can be asked, such as how much of the variation is heritable, which factors contribute to the genetic variability and what are their relative magnitudes and importance. Variation may be measured either among individuals when the population of individuals has no discernible structure, or among family means where there are naturally occurring or artificially produced families or clones.

In this chapter, we will restrict ourselves to defining, analysing and interpreting variation in the six basic generations while the variation among generations produced by selfing and sib-mating will be considered in Chapters 4 and 5.

3.1 Variation in the non-segregating generations

We can divide the six basic generations into two distinct groups with respect to their variances. The first group includes those generations that consist of genetically identical individuals, such as the P_1, P_2 and F_1 families. Because individuals within these generations do not show any genetic differences, they are referred to as the **non-segregating generations**. The second group includes the F_2, $Bc_{1.1}$ and $Bc_{1.2}$ generations which contain a mixture of genotypes resulting from segregation, random assortment and recombination of alleles at those loci for which P_1 and P_2 differ and the F_1 is heterozygous. Hence they are referred to as the **segregating generations**.

3.2 *Environmental variation*

Because the individuals within the P_1, P_2 and F_1 families are genetically identical, any variation between them cannot be genetical and it is conventional to refer to such variation as the **environmental variation within families**, V_E. This type of variation exists even between individuals which may be exposed to virtually identical environments and it is a major component of the total variation for quantitative traits, which frequently exceeds the genetical variation and the macro-environmental variation caused by such factors as blocks, locations or plots. Because V_E is such a large component it is worth exploring its underlying causes, of which there are several and not all are environmental in the strict sense.

3.2.1 Pre-replication factors

These include a variety of factors which influence the phenotype of the individual but which are determined directly by the mother or environmental effects on the mother. Consider, for example, plant height. This trait is affected as much by variation in, say, the quality of the seed as it is by soil fertility; poor quality seed produces weaker seedlings which, in turn, may lead to shorter plants. Seed quality is affected by various factors, such as the age of the mother, the age and position of the flower, age of the plant at the time of pollination, harvest time, conditions at the time of maturity and health of the maternal plant. For these reasons, even seeds harvested from the same plant often show extensive variation in quality including size, maturity and vitality. No two seeds are equally endowed, even those located next to each other on a maize cob, a wheat spike or in a tobacco capsule. A large proportion of such variation in seed quality, however, may be eliminated by taking precautions during seed production and grading the seed, but it is virtually impossible to obtain completely uniform seed. Consequently, some effects of seed quality will always be apparent on the plant height and they will appear confounded with V_E. Similar factors operate in animals. For example, monozygotic twins in humans are genetically identical; clones in fact. Yet, such twins are never completely identical at birth and all pairs show at least some differences in weight, height, health and personality which must have arisen in the uterus (Figure 1.3).

3.2.2 Developmental factors

Another important factor which affects V_E is developmental variation. This is an inherent property of the individuals which is difficult to measure, control or manipulate, but one example of where

developmental variation can be measured is the difference between duplicate measurements from the same individual. In animals, the two halves of the body are symmetrical and therefore provide duplicate measures on limbs, wings and bone structure, etc. In plants, differences between duplicate structures on opposite sides of a plant or repeat structures on different branches or tillers can be measured. Because these duplicate structures have received the same environment, the differences between them provide a measure of the developmental variation. In the case of traits such as height or weight, these developmental factors cannot be separated from the genuine within family variation.

3.2.3 Measurement errors

A third major factor which contributes significantly to V_E is measurement error which can be subdivided into that due to (a) measurement and (b) rounding off. The first group depends on the accuracy and consistency with which the data are scored and recorded, the magnitude of scoring error usually being largest when several people participate in scoring the experiment. Rounding off errors occur when a continuous variable is scored to a fixed interval of the measurement unit such as to the nearest centimetre or gram.

3.2.4 Statistical sampling error

Many quantitative traits are measured as a proportion, such as seed germination, disease attack or the answers to a personality test. These traits will additionally show statistical sampling variance because the proportion, p, will have the binomial variance, $p(1-p)/n$, where n is the number scored. Such variation is clearly related to p but is not otherwise determined by the environment.

3.2.5 True micro-environmental variation

The final factor which contributes to V_E is genuine micro-environmental variability. Because every individual occupies a unique part of the environment in the experimental area, it may be exposed to different levels of nutrition, temperature, sunlight, moisture, tillage, soil depth, etc.

It should be apparent that only the last of these five factors indeed represents differences due to the environment in which the genotypes are reared, and thus true micro-environmental variation often represents just a small part of the observed value of V_E (Table 3.1). However, there is generally no easy method of separating the contribution made to V_E by each component. The non-environmental components are difficult to manipulate while the true micro-environmental factors themselves may be very difficult to identify. It therefore follows

Table 3.1 Relative magnitudes of developmental variance as a proportion of V_E in *Drosophila* and *Nicotiana*

Organisms and traits	Developmental variance	V_E	% Developmental
Drosophila			
Sternopleural chaetae			
(a) Inbred lines	1.36	1.63	83
(b) F_1s	1.39	1.41	97
Nicotiana rustica			
(i) Height (12 weeks)	55.93	69.35	81
(ii) Leaf length	4.01	4.31	93

that V_E will always be present and will frequently be a major source of variation which will not be reduced easily by elaborate equipment for controlling the environment. In practice, therefore, one should obtain an accurate estimate of V_E, and reduce its effect on the variance of generation means by increasing the number of individuals. It is invariably easier and cheaper to increase the precision with which each mean is estimated by increased replication than by attempting to control the environment. The effect of greater replication on the variance of the means can be predicted and costed precisely while the results and cost of controlling the environment are unpredictable.

The components of V_E just discussed affect the phenotype of individuals within a family. Environmental differences can also exist between families, and such effects would be common to all progeny in that family. Thus, even though two mothers, plant or animal, may be genetically identical, differences in their environment such as nutrition or health, etc., may affect the phenotype of their progeny. Such environmental effects are represented by the symbol V_{EC}, which stands for the **common environment**. It is therefore very important in all studies of quantitative traits to take every precaution to ensure that the progeny are produced from parents that have been raised under similar conditions. For example, one should never raise F_1 plants from seed produced last year in the same experiment with P_1 plants from seed produced in the previous year.

3.3 Estimating environmental variance

Providing that the experiment is adequately randomized, the individuals of P_1, P_2 and F_1 are deemed to be exposed to the same

range of environmental conditions. Their variances should thus provide independent estimates of V_E which are not expected to differ from each other. For this reason V_E is also called the **additive environmental variance**. For example, from the variances given in Table 3.2, we can estimate V_E by averaging the values of s^2 of the three non-segregating generations. Because each generation is based on the same number of observations in the present case,

$$V_E = \tfrac{1}{3}(59.73 + 65.71 + 51.81) = 59.08.$$

Had the number of observations in each generation been different, then the variances would be pooled by summing their SS and dividing by their combined df.

However, on many occasions inbred lines (P_1 and P_2) and hybrids (F_1) respond differently to the same micro-environment, and this estimate of V_E will be meaningful only when the variances of the P_1, P_2 and F_1 generations are homogeneous, i.e. there is no interaction between the micro-environmental variation and the genotype. We can check whether $s^2P_1 = s^2P_2 = s^2F_1$ by means of a Bartlett's [1] or a Levene's [2] test. Alternatively, and more simply, an F-test can be applied using the ratio of the largest (65.71) to the smallest (51.81) of the three variances.

$$F = 65.71/51.81 = 1.27; \quad P \approx 0.10.$$

Because we have chosen the two extreme variances out of n (= 3) and divided the largest by the smallest, the value of P has to be corrected by multiplying by n(n − 1) (= 6). Thus, the correct probability that we obtain for the F test is $P \approx 0.6$, which shows that the parental and F_1 variances do not differ significantly from each other in this case and so we can accept the pooled estimate of $V_E = 59.08$.

3.4 Variation in the segregating generations

The individuals in the segregating generations will be subject not only to non-genetical, but also to genetical variation, as is suggested by the larger variances of the F_2 and Bc generations in Table 3.2. All the genes that were heterozygous in the F_1 should contribute to these variances and we can determine the nature and magnitude of this variation by deriving theoretical expectations of these variances from the genetic values and their frequencies in the segregating generations.

Let the expected genetic variance of an F_2 population at a single segregating locus be V_G. From statistical theory we find that

Table 3.2 Within family variances for the six basic generations, each based on 100 individuals (data taken from Table 2.1)

Generation	Within-family variance, s_x^2 (99df)
P_1	59.73
P_2	65.71
F_1	51.81
F_2	100.75
$Bc_{1.1}$	81.05
$Bc_{1.2}$	90.83

$V_G = \sum f_i(g_i - \text{mean})^2/(\sum f_i - 1)$ where f_i is the frequency of the ith genotype, and g_i is its genetic value. While this formula applies to small samples, it can be simplified to $V_G = \sum f_i g_i^2 - (\sum f_i g_i)^2$ when we are dealing with the F_2 population as a whole. The f_i now represents the proportion of individuals having the ith genotype and not the frequency, i.e. $\sum f_i = 1$ (not n). Applying this formula to gene A, for which the F_2 will consist of A^+A^+, A^+A^- and A^-A^- genotypes which will be present with the proportions of $\frac{1}{4} : \frac{1}{2} : \frac{1}{4}$ respectively, we get:

$$
\begin{aligned}
F_2 \text{ Mean} &= \sum f_i g_i \\
&= \tfrac{1}{4}(A^+A^+) + \tfrac{1}{2}(A^+A^-) + \tfrac{1}{4}(A^-A^-) \\
&= \tfrac{1}{4}(m + a_A) + \tfrac{1}{2}(m + d_A) + \tfrac{1}{4}(m - a_A) \\
&= m + \tfrac{1}{2}d_A \text{ (as in Chapter 2).}
\end{aligned}
$$

$$
\begin{aligned}
F_2 \text{ Variance} &= \sum f_i g_i^2 - \left(\sum f_i g_i\right)^2 \\
&= \tfrac{1}{4}(A^+A^+)^2 + \tfrac{1}{2}(A^+A^-)^2 + \tfrac{1}{4}(A^-A^-)^2 - (\text{mean})^2 \\
&= \{\tfrac{1}{4}(m + a_A)^2 + \tfrac{1}{2}(m + d_A)^2 + \tfrac{1}{4}(m - a_A)^2\} \\
&\quad - (m + \tfrac{1}{2}d_A)^2 \\
&= \tfrac{1}{2}a_A^2 + \tfrac{1}{4}d_A^2.
\end{aligned}
$$

The expected genetical variance of an F_2, V_G, for a single gene, thus consists of two parts, and we can write this expectation in terms of V_A^* and V_D^*, where $V_A^* = \frac{1}{2}a_A^2$, the **additive genetic** component of variance, and $V_D^* = \frac{1}{4}d_A^2$, the **dominance** or **non-additive** genetic component of variance.

For many genes (k), each of which shows independent gene action (i.e. no epistasis) and independent gene transmission at gamete formation (i.e. no linkage), V_A^* and V_D^* will represent the sums of

the additive and dominance variances of the individual genes, that is:

$$V_A^* = \tfrac{1}{2}(a_A^2 + a_B^2 + \ldots + a_K^2) = \tfrac{1}{2}\sum_{i=1}^{k} a_i^2$$

$$V_D^* = \tfrac{1}{4}(d_A^2 + d_B^2 + \ldots + d_K^2) = \tfrac{1}{4}\sum_{i=1}^{k} d_i^2.$$

(N.B. The * notation in V_A^*, V_D^* is to distinguish the special case of equal allele frequencies. See Chapter 9 for the general case.)

Thus, the expected variation among F_2 individuals is $V_A^* + V_D^* + V_E$, compared with just V_E for the non-segregating generations, P_1, P_2 and F_1. Consequently, the F_2 variance should be larger than V_E and significantly so where the F_2 is segregating at a large number of genes with small effect or a few genes with large effect.

In the case of $Bc_{1.1}$, a cross between F_1 and P_1, the genetic constitution of the family for gene A is $\frac{1}{2}A^+A^+ : \frac{1}{2}A^+A^-$, the scores of the genotypes are as given earlier, the mean genetic value is $m + \frac{1}{2}a_A + \frac{1}{2}d_A$ and the expected genetic variance is $\frac{1}{4}a_A^2 + \frac{1}{4}d_A^2 - \frac{1}{2}a_A d_A$. For k genes, this expectation becomes $\frac{1}{2}V_A^* + V_D^* - V_{AD}$ where V_A^* and V_D^* are as defined earlier and $V_{AD} = \frac{1}{2}(\delta_a a_A d_A + \delta_b a_B d_B + \ldots + \delta_k a_K d_K)$ [where $\delta_i = +1$ if the allele is in P_1 and $\delta_i = -1$ otherwise]; V_{AD} is thus a cross product of the additive and the dominance effects of the genes that are segregating in the cross. Thus, the phenotypic variance of $Bc_{1.1}$ will be $\frac{1}{2}V_A^* + V_D^* - V_{AD} + V_E$. Similarly, the phenotypic variance of $Bc_{1.2}$ $(= F_1 \times P_2)$ will be $\frac{1}{2}V_A^* + V_D^* + V_{AD} + V_E$ while the average of the $Bc_{1.1}$ and $Bc_{1.2}$ variances will be equal to $\frac{1}{2}V_A^* + V_D^* + V_E$. This average phenotypic variance should be smaller than that of the F_2 providing the additive genetic, dominance genetic and additive environmental model is adequate.

A universal characteristic of V_A^*, V_D^* and V_E is that, being components of variances, they cannot be negative. Sometimes the estimates may be negative due to sampling errors but they are unlikely to be significant. V_{AD}, on the other hand, is a covariance and its sign will depend on the direction of dominance. When alleles which decrease the phenotype are dominant, V_{AD} will be negative, and it will be positive for the opposite situation. Further, the magnitude of V_{AD} will be maximum when the alleles are completely associated between the parental lines and dominance is unidirectional, otherwise it will reflect the net balance of the cross products after internal cancellations. Therefore, the backcross to the parent with the largest number of dominant alleles will have the smaller phenotypic variance and, in the extreme case of a parent which contains all the dominant alleles and where dominance is complete ($d = a$), then the genetical variance will be zero because $\frac{1}{2}V_A^* + V_D^* = V_{AD}$. The data in Table 3.2 show

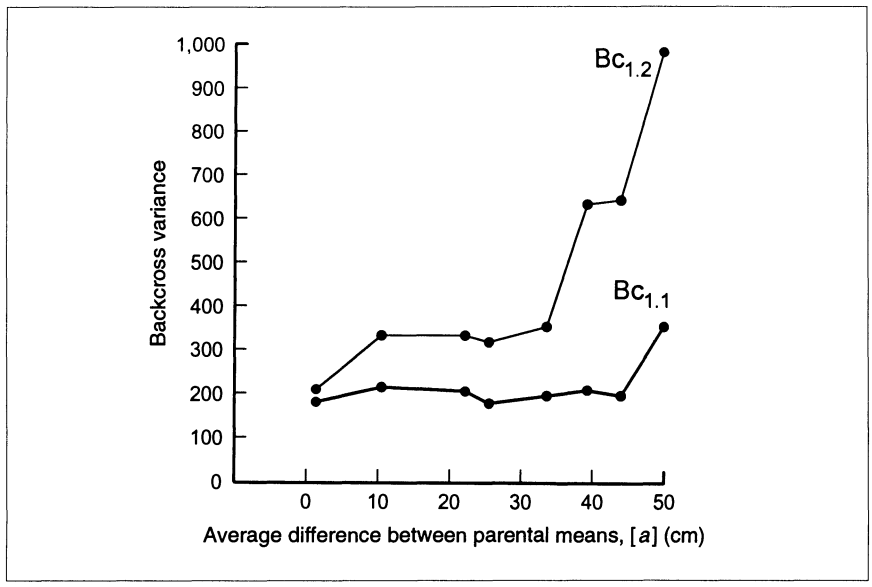

Figure 3.1 The observed relationship between backcross variance and the parental difference [*a*] for final height in *Nicotiana rustica* crosses.

that the backcross to P_1 ($Bc_{1.1}$) has the smaller variance, indicating that dominance is for the higher value of the trait and confirming what was shown by [*d*] in Chapter 2. With unidirectional dominance, therefore, V_{AD} will generally show a linear relationship with half the parental difference, [*a*], as is shown in Figure 3.1, for data collected on the basic generations derived from the recombinant inbred lines extracted from a cross between two pure breeding varieties of *Nicotiana rustica*.

3.5 *Estimation of genetical components*

Initially the breeder would like to know whether the cross shows significant genetic variation and, if so, how much of the variation is heritable and what types of gene effects are significant. The first and second questions are answered by comparing the variances of the segregating and the non-segregating generations. The simplest method is to use the pooled estimate of V_E derived earlier, 59.08, in an F-test to determine whether or not the F_2 and the backcross variances are significantly larger than V_E. Now,

for the F_2, $F_{(99,297)} = s^2_{F2}/V_E = 100.75/59.08 = 1.71^{**}$;

for the $Bc_{1.1}$, $F_{(99,297)} = s^2_{Bc1.1}/V_E = 81.05/59.08 = 1.37^{*}$;

for the $Bc_{1.2}$, $F_{(99,297)} = s^2_{Bc1.2}/V_E = 90.83/59.08 = 1.54^{**}$.

indicating that genetic variation exists in the generations derived from the cross. Having shown that the genetic variation exists, the breeder can then address the third question: what types of genetic variation are involved? This is achieved by estimating V_A^*, V_D^* and V_{AD} using the following equations.

$$V_A^* = (2s^2_{F_2} - s^2_{Bc_{1.1}} - s^2_{Bc_{1.2}})$$
$$V_D^* = (s^2_{Bc_{1.1}} + s^2_{Bc_{1.2}} - s^2_{F_2} - V_E) \qquad \text{[Eqn 3.1]}$$
$$V_{AD} = \tfrac{1}{2}(s^2_{Bc_{1.2}} - s^2_{Bc_{1.1}}).$$

Therefore:

$$V_A^* = (2 \times 100.75 - 81.05 - 90.83) = 29.62$$
$$V_D^* = (81.05 + 90.83 - 100.75 - 59.08) = 12.05$$
$$V_{AD} = \tfrac{1}{2}(90.83 - 81.05) = 4.89$$
$$\text{while } V_E = 59.08.$$

The relative magnitudes of these estimates are typical for the components of variances. Two of the components (V_D^* and V_{AD}) are relatively small and unlikely to be significant. Estimates of variances are more variable than those of means and the parameters are highly correlated, so it is very probable that we will find some of the genetic components non-significant, even when the overall genetic variation is highly significant, as in the present case. Clearly, as with generation means (Chapter 2), we need to select the best model which provides an adequate fit to the data, has all the estimated parameters significant and makes biological sense. This can be achieved by applying a modified form of the weighted least squares procedure described for fitting models to means. This modification is necessary in order to determine the appropriate weights for each variance which are not directly available as they are for means. With generation means, there are replicate observations of individual scores which provide an empirical estimate of the variance of the mean and hence an empirical weight. With variances, on the other hand, no such replicate variances are available, and hence, we do not have empirical weights. The theoretical variance of an observed variance (s^2) is equal to $2(\epsilon s^2)^2/df$, but the expected variance, ϵs^2, is not known. Hayman [3] solved this problem by using $df/2(s^2)^2$ as the initial weight in an iterative process. After the first iteration the parameter estimates are used to calculate expected variances which will approximate to the true values. These are then used to calculate new weights for a second iteration. This process is repeated through successive iterations until the test statistic, χ^2, reaches a minimum.

Table 3.3 Expectations of the within-family variances in terms of the additive dominance genetic and the additive environmental components of variation

Generation	Parameters			
	V_E	V_A^*	V_D^*	V_{AD}
P_1	1	0	0	0
P_2	1	0	0	0
F_1	1	0	0	0
F_2	1	1	1	0
$Bc_{1.1}$	1	$\frac{1}{2}$	1	-1
$Bc_{1.2}$	1	$\frac{1}{2}$	1	1

Model fitting in fact provides a comprehensive solution to all the problems that we have outlined earlier. Firstly, fitting a V_A^*, V_D^*, V_{AD} and V_E model (Table 3.3) to six variances yields a test of goodness of fit in the form of χ^2 for two degrees of freedom which also determines if a single V_E fits the variances of P_1, P_2 and F_1 generations. In other words, it establishes whether the variances of the non-segregating generations differ significantly or not. If they do not differ, fitting a single parameter V_E, tests if the six variances differ significantly, which becomes an unambiguous test of the genetic variation. We then proceed with the model fitting to obtain the best statistical *cum* biological model. This is illustrated in Table 3.4 for the data of Table 3.2 using model fitting with different combinations of parameters. The most appropriate model requires just two parameters, V_E and V_A^*, both of which are significant and the $\chi^2_{(4)}$ test of goodness of fit is non-significant. Alternatively, when the variances of P_1, P_2 and F_1 are shown to be heterogeneous, we need to replace V_E with three separate parameters, V_{E1}, V_{E2} and V_{E3}, to account for differences between them. Now, there are six parameters in the model and their contributions to the variances will be:

$$s^2_{P_1} = V_{E1}$$

$$s^2_{P_2} = V_{E2}$$

$$s^2_{F_1} = V_{E3}$$

$$s^2_{F_2} = V_A^* + V_D^* + \tfrac{1}{4}V_{E1} + \tfrac{1}{4}V_{E2} + \tfrac{1}{2}V_{E3}$$

$$s^2_{Bc_{1\,1}} = \tfrac{1}{2}V_A^* + V_D^* - V_{AD} + \tfrac{1}{2}V_{E1} + \tfrac{1}{2}V_{E3}$$

$$s^2_{Bc_{1.2}} = \tfrac{1}{2}V_A^* + V_D^* + V_{AD} + \tfrac{1}{2}V_{E2} + \tfrac{1}{2}V_{E3}.$$

On many occasions we may find that only one of the P_1, P_2 or F_1 variances is significantly different from the other two. In this case,

Quantitative Genetics

Quantitative genetics is a special branch of genetics, which is concerned with the inheritance of the differences between individuals that are measured in degree rather than in kind. These individual differences are referred to as quantitative differences or quantitative traits. Quantitative genetics is one of the disciplines of genetics dealing with the mechanism of quantitatively inherited traits. Classical quantitative genetics is also called statistical genetics or biometrical genetics .

Important Quantitative Concepts

The following sections are the core of this chapter. the six important concepts in quantitative behavioral genetics :

1_ heritability

2_ environmentability

3_ genetic correlation

4_environmental correlation

5_ geneenvironment interaction

6_ gene-environment correlation

Although these six are described at a conceptual level, it is important to recognize that behavioral geneticists try to quantify each of them—i.e., arrive at an actual number to estimate these six quantities and then judge how important this quantity is for a behavioral phenotype

Classical quantitative genetics

Discussion in this section includes genetic and environmental variances, relationships and genetic diversity, linkage and epistatic issues in populations .

Quantitative trait

Most traits encountered in plant breeding are quantitatively inherited. Many genes control such traits, each contributing a small effect to the .overall phenotypic expression of a trait

Variation in quantitative trait expression is without natural discontinuities (i.e.,the variation is continuous). The traits that exhibit .continuous variations are also called metric traits

Any attempt to classify such traits into distinct groups is only arbitrary. For example, height is a quantitative trait. If plants are grouped into tall versus short plants, relatively tall plants could be found in the short group and, similarly, short plants could be found in the tall group .

Qualitative genetics versus quantitative genetic

The major way in which qualitative genetics and quantitative genetics differ may be summarized as follows:

Qualitative genetics	*Quantitative genetics*
Characters of kind .	Characters of degree .
Discontinuous variation ,distinct phenotypic classes .	Continuous variation , phenotypic measurements from a spectrum .
Single gene effects .	Polygenic control ,effects of single genes too slight to be detected .
The phenotypic expression of a gene is not influenced by environment .	Environmental conditions effect the phenotypic expression of polygenes variously .
Analyzed by making counts and ratios.	In it analysis is made by statistical method .
Examples : Round or wrinkled seeds of pea ,Black or white coat of guinea pigs .	Examples : Skin color in Man , eye color in Man

Climate Change, Quantitative Genetics

Classical quantitative genetics is a powerful tool for investigating evolution in response to climate change .

Considering advances in our understanding of traits important for climate response at the molecular level, one may wonder why quantitative genetics is still a relevant approach? The answer is that despite the discovery of molecular mechanisms that underlie a few specific traits, our understanding is still woefully inadequate to describe more complex polygenic phenotypes as they interact with the environment in the wild. For example, timing of life history events, dispersal ability, thermal and drought tolerance, and competitive ability are traits with complex genetic underpinnings that will likely be the direct targets of selection with climate change

Although some of these traits are influenced by a few regions of chromosomes with major effects, called quantitative trait loci (QTLs)

Quantitative Genetics and Plant Breeding

Quantitative genetic principles apply to almost any area of plant breeding. Breeders recognize the need for more extensive testing for traits of low heritability than for traits of high heritability .

Who studies quantitative traits ?

Most biologists will find themselves having to work with quantitative characters at some time. There are, however, certain groups for whom a knowledge of such traits is essential. First among these are the breeders of plants and animals. Although there are still some advances to be made by manipulating single, major gene differences, particularly in the area of resistance to disease and herbicides, much of the present progress in breeding is achieved by improving quantitative characters such as quality, uniformity, growth rate, yield and response to fertilizers. Secondly, there are the population geneticists and students of evolution in natural populations. As with artificial, commercial selection, natural selection

also acts largely on the pool of quantitative variation. Indeed, fitness, in that it is the cumulative result of the whole genome in a given environment, is the

ultimate quantitative trait, while its component parts, such as competitive ability, adaptability and reproductive efficiency are all quantitative traits as well. Thirdly, we have the students of animal, particularly human, behaviour and personality - psychologists and sociologists. This group is interested in a variety of different, complex, quantitative characters such as intelligence, personality, social attitudes, etc. as well as abnormal behaviour, like schizophrenia. The fourth and final main group includes those interested in the genetics of human disease. Although many genetic diseases can be traced to identifiable major genes or chromosomal abnormalities, many others, such as propensity for heart disease, obesity and some cancers have a polygenic origin .

2 Basic generations – means

The genetical analysis of quantitative traits can not follow the standard procedures used to analyse major gene traits, such as looking for a one-gene (3:1) or two-gene (9:3:3:1) phenotypic ratio in an F_2, because it is not possible to follow the segregation of the separate, underlying polygenes. Instead, it is necessary to look at the degree of similarity or difference among related individuals and families using various statistics such as means, variances, covariances, regressions and correlations.

If we score a number of individuals from a particular family for a quantitative trait, the mean phenotype of those individuals and the variation among them will be due to the joint action and interaction of their genes and the environment. Different families will have different means and different variances because they contain different genotypes. The genetical contributions to these family means and variances can be investigated by searching for simple and plausible models which adequately explain the data.

In practice, there are a very large number of possible families that one might be working with but, for convenience and simplicity, we will start with those families that can be obtained from a pair of true breeding, homozygous lines. Table 2.1 gives an example of the means and within family variances of two such lines together with their F_1, F_2 and first backcross ($Bc_{1.1} = F_1 \times P_1$ and $Bc_{1.2} = F_1 \times P_2$) families. These six families are often referred to as the **six basic generations**. Not only do such families provide a simple conceptual framework from which to illustrate the construction and testing of the genetical models but they also appear in such a wide variety of contexts that they are the linchpin of quantitative analysis in most situations. In this chapter we deal with models for analysing the family means while the variances will be considered in Chapter 3. Unless otherwise stated, we will always assume that the organism is either diploid or, at least, shows disomic inheritance.

Table 2.1 The means and variances of the six basic generations for a quantitative trait. Each mean is based on 100 individuals, and hence the variance has 99 df

Generation	Mean ($\bar{x}$)	Variance (s_x^2)
P_1	69.44	59.73
P_2	59.04	65.71
F_1	83.44	51.81
F_2	74.36	100.75
$Bc_{1.1}$	76.03	81.05
$Bc_{1.2}$	71.28	90.83

2.1 *Single gene model with additive and dominance effects*

Let us suppose initially that we have two homozygous, inbred lines which are to be used as parents, P_1 and P_2, which are identical except for one gene. This gene, which we shall call A, has two alleles, A^+ and A^-, which are responsible for the higher and lower phenotypic scores respectively of the character being studied. If we adopt the convention that P_1 is the parent with the higher phenotype for the character under study, then the genotypes of P_1 and P_2 for this gene will be A^+A^+ and A^-A^- while their F_1 hybrid (from a cross of $P_1♀ \times P_2♂$ or $P_2♀ \times P_1♂$) will be A^+A^-. Let us further suppose that these two parents and the F_1 are raised in a replicated experiment and that we calculate their average phenotypic values for the character, $\bar{P}_1$, $\bar{P}_2$, $\bar{F}_1$.

If we use the average phenotype of the two parents, P_1 and P_2 as a base line, m, we can define the expected mean of all three genotypes as linear deviations from this as follows [1],

$$\begin{aligned} \bar{P}_1(A^+A^+) &\quad m + a_A \\ \bar{P}_2(A^-A^-) &\quad m - a_A \\ \bar{F}_1(A^+A^-) &\quad m + d_A. \end{aligned}$$

The parameters a_A and d_A are deviations from the mean, m, due to the effects of homozygous or heterozygous genotypes respectively; a_A is called the **additive** genetic and d_A the **dominance** genetic component of means. This model implies that a_A will always be positive and P_1 will always be greater than or equal to P_2, while the sign of d_A will be determined by the direction of dominance. Thus, d_A will be positive

when A^+ is dominant to A^-, negative when A^- is dominant to A^+ and zero when neither allele is dominant. It is further obvious that the absolute value of d_A ($|d_A|$) will be less than that of a_A when dominance is incomplete or partial, that $|d_A| = a_A$ for complete dominance and that $|d_A| > a_A$ for over-dominance or super-dominance.

These three parameters could be estimated using the following orthogonal comparisons among the three generation means,

$$m = \tfrac{1}{2}\bar{P}_1 + \tfrac{1}{2}\bar{P}_2$$
$$a_A = \tfrac{1}{2}\bar{P}_1 - \tfrac{1}{2}\bar{P}_2$$
$$d_A = \bar{F}_1 - \tfrac{1}{2}\bar{P}_1 - \tfrac{1}{2}\bar{P}_2.$$

The linear genetic model associated with each genotype ($m + a_A$; $m - a_A$; $m + d_A$) will be referred to as the **genetic value** of that genotype because it defines the underlying genetic contribution to the observed phenotype.

Selfing or intercrossing the $F_1(A^+A^-)$ produces an F_2 family which will have the following constitution:

Genotype	A^+A^+	A^+A^-	A^-A^-
Frequency	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Genetic value	$m + a_A$	$m + d_A$	$m - a_A$

The average phenotype of these genotypes obtained as,

$$\tfrac{1}{4}(m + a_A) + \tfrac{1}{2}(m + d_A) + \tfrac{1}{4}(m - a_A)$$

gives the expectation of the F_2 mean, which is equal to,

$$m + \tfrac{1}{2}d_A.$$

Backcrossing P_1 to F_1, on the other hand, produces the $Bc_{1.1}$ family which has two genotypes, A^+A^+ and A^+A^-, in equal proportions so that the average genetic value of these genotypes, the expected mean of $Bc_{1.1}$, is:

$$\tfrac{1}{2}(m + a_A) + \tfrac{1}{2}(m + d_A) = m + \tfrac{1}{2}a_A + \tfrac{1}{2}d_A.$$

Similarly, the two genotypes which will be present with equal frequency in the $Bc_{1.2}$ family, a cross between P_2 and F_1, are A^+A^- and A^-A^-, and their average genetic value, the $Bc_{1.2}$ family mean, is,

$$\tfrac{1}{2}(m + d_A) + \tfrac{1}{2}(m - a_A) = m - \tfrac{1}{2}a_A + \tfrac{1}{2}d_A.$$