

# Introduction to Basic Concepts in Quantitative Genetics

Izel Fourie Sørensen & Peter Sørensen

2022-03-18

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Quantitative Genetics</b>	<b>1</b>
2.1	Infinitesimal model . . . . .	2
2.2	Single locus model for a quantitative trait . . . . .	5
2.3	Multiple locus model for a quantitative trait . . . . .	12

Replace plant/animal examples with human examples

## 1 Introduction

This section introduces the basic concepts in Quantitative Genetics such as:

- Genetic value and variance for a quantitative trait
- Genetic parameters (heritability, genetic variance and correlation)
- Single locus model, multiple locus model and infinitesimal model

These concepts are relevant for a range of genetic and statistical analyses of complex traits and diseases in human populations, including:

- Estimating the effect of single locus (or marker) for gene discovery
- Estimating the effect of multiple loci (or markers) for genomic prediction
- Estimating the heritability of a trait (the part of its variability due to genetics)
- Estimating genetic predisposition by pedigree or genomic information

## 2 Quantitative Genetics

Quantitative genetics, also referred to as the genetics of complex traits, is the study of quantitative traits. Quantitative genetics is based on models in which many genes influence the trait, and in which non-genetic factors may also be important. Quantitative traits such as height, obesity or longevity vary greatly among individuals. Their phenotypes are continuously distributed phenotypes and do not show simple Mendelian inheritance (i.e., phenotypes that are distributed in discrete categories determined by one or a few genes). The quantitative genetics framework can also be used to analyze categorical traits like number of children given birth to (which consist of discrete counts like 0, 1, 2, 3, ...) or binary traits like survival to adulthood

(which consist of 0 or 1, ‘dead’ or ‘alive’, etc.), provided that they have a polygenic basis (i.e., they are determined by many genes). The quantitative genetics approach has diverse applications: it is fundamental to an understanding of variation and covariation among relatives in natural and managed populations; it is also used as basis for predicting genetic predisposition in humans as well as selective breeding methods in animal and plant populations.

## 2.1 Infinitesimal model

The infinitesimal model, also known as the polygenic model, is a widely used genetic model in quantitative genetics. Originally developed in 1918 by Ronald Fisher, it is based on the idea that variation in a quantitative trait is influenced by an infinitely large number of genes, each of which makes an infinitely small (infinitesimal) contribution to the phenotype, as well as by environmental (non-genetic) factors. In the most basic model the phenotype (P) is the sum of genetic effects (G), and environmental effects (E):

$$P = G + E \quad (1)$$

The genotypic effect (G) in the model can be divided into additive effects (A), dominance effects (D), and epistatic effects (I) such that the expanded infinitesimal model becomes:

$$P = A + D + I + E \quad (2)$$

The genotypic effect may also depend on the environment in which they are expressed (e.g., in plants a drought-tolerance gene may have a favorable effect on grain yield under water-limited conditions, but may be useless under irrigation). Therefore we may consider an extended version of the infinitesimal model where the phenotype (P) is the sum of genotypic effects (G), environmental effect (E), and genotype-environment interaction effects (G×E):

$$P = G + E + G \times E \quad (3)$$

In practice, the genotype-environment interaction effect can be important for the phenotype of individuals, but for the sake of simplicity we will ignore them in the remainder of this section. Therefore, in the following, we will assume that genotypic effects are not impacted by environmental factors.

### 2.1.1 Genotypic effects

[EDIT THIS SECTION UNTIL EPISTATIC GENETIC EFFECTS] The genotypic effect (G) in the model can include additive effects (A), dominance effects (D), and epistatic effects (I). Additive effects are the summed effects of individual alleles. Dominance effects are interactions between alleles within loci. Epistatic effects are interactions between alleles in different loci and can therefore only occur if two or more loci affect the trait.

Consider an individual that is diploid, like humans (i.e., they carry two copies of every gene, except in their sexual chromosomes). Assume that one locus in its genome exists under two possible alleles:  $A_1$  and  $A_2$ , with respective allele effects +1 and -1.

Generally, the value of each genotype will depend on the combination of alleles within one locus ( $G = A + D$ ) or across multiple loci ( $G = A + D + I$ ). For example, in presence of dominance, the value of each possible genotype may be -2 if the individual is  $A_2A_2$ , +1 if it is  $A_2A_1$  (or  $A_1A_2$ ), and +2 if it is  $A_1A_1$ .

**2.1.1.1 Additive Effects** If genetic effects are entirely additive, then the value of each possible genotype is the sum of their respective allele effects, i.e., -2 if the individual is  $A_2A_2$ , 0 if it is  $A_2A_1$  (or  $A_1A_2$ ), and +2 if it is  $A_1A_1$ . In general additive effects are the summed effects of average allele effects.

**2.1.1.2 Dominance Effects** Dominance genetic effects are the interactions among alleles at a given locus. This is an effect that is extra to the sum of the additive allele effects. Each genotype has its own dominance effect, denoted by  $\delta_{ij}$ , for the specific combination of alleles  $i$  and  $j$ , (e.g.,  $\delta_{A_1A_2}$ ), and each of them are non-zero quantities.

**2.1.1.3 Epistatic Genetic Effects** Epistatic genetic effects encompass all possible interactions among the loci impacting the trait, whenever there is more than one such loci. This includes all two-way interactions (e.g., interactions between loci A and B, A and C), three-way interactions (e.g., joint interaction among A, B and C), etc. Epistasis can be decomposed, so it includes interactions between additive effects at different loci, interactions between additive effects at one locus with dominance effects at a second locus, and interactions between dominance effects at different loci.

## 2.1.2 Infinite number of loci each with small effect on the phenotype

Quantitative traits do not behave according to simple Mendelian inheritance laws. More specifically, their inheritance cannot be explained by the genetic segregation of one or a few genes. Even though Mendelian inheritance laws accurately depict the segregation of genotypes in a population, they are not tractable with the large number of genes which typically affect quantitative traits. To better understand the infinitesimal model assume Mendelian inheritance to occur at every locus in the genome. Let's say there are 30,000 gene loci in the genome. The number of alleles at each locus varies from 2 to 30 or more. If we assume that there are only two alleles (3 possible genotypes) per locus, and gene loci segregate independently, then the number of possible genotypes (considering all loci simultaneously) would be  $3^{30000}$  which is large enough to give the illusion of an infinite number of loci. Furthermore each of these loci could contribute additive and dominance effects in addition to interaction effects.

**2.1.2.1 Distribution of genotypic and phenotypic values in single locus model** [EDIT: I recommend following] A single-locus model is applied to (when) model (ling) the genetic basis of a quantitative trait when a single locus affects the trait of interest. [OR] When a single locus affects a quantitative trait, a single-locus model is used to model the genetic basis of the trait.

[THEN DELETE TWO SENTENCES:] First we will consider how to model the genetic basis of a quantitative trait when a single locus affects the trait of interest. We call this a single-locus model. The distribution of the genotypic values for a set of individuals will be discrete. The frequency of the genotypic values depend on genotype frequencies, which in turn depend on allele frequencies of  $A_1$  and  $A_2$ . The phenotype is however also influenced by the environment. If we assume that the environmental effects are normally distributed (e.g.  $\mathcal{N}(0, \sigma^2 = 1)$ ) then we can observe that the phenotype distribution is in fact normally distributed.

**2.1.2.2 Distribution of genotypic and phenotypic values in multiple loci model** [When multiple loci affect a quantitative trait, a polygenic- or an infinitesimal model is applied to model the genetic basis of the trait.] [REMOVE: Now we will consider a multiple-locus model.] When several loci are causal (i.e., they have an effect on a certain trait), it is referred to as a **polygenic model**. When the number of causal loci tend to infinity, it is referred to as an **infinitesimal model**. From a statistical point of view, the genetic values in an infinitesimal model are considered random with a known distribution. [EDIT THIS: IS THE SAME THING REPEATED TWICE?] According to the central limit theorem, the distribution of any sum of a large number of very small effects converges to a normal distribution. Therefore, the genetic values in an infinitesimal model tends to a normal distribution, because of the infinitely large number of causal loci. In the case where a given trait of interest is thought to be influenced by a large number of genetic loci each having a small effect, the sum of the genetics values of all loci together can be approximated by a normal distribution. [EDIT: WHAT ABOUT THE DISTRIBUTION OF THE PHENOTYPIC VALUES?]

### 2.1.3 Genetic parameters

Fisher (1918) and Wright (1921) have introduced fundamental statistical methods in quantitative genetics:

- analysis of variance: the partition of phenotypic variation into heritable (A) and non-heritable components (D, I and E).
- resemblance among relatives: the estimations of the proportion of loci shared by relatives under the infinitesimal model.

**2.1.3.1 Genetic variance:** In the model proposed by Fisher (1918), Cockerham (1954) and Kempthorne (1954), covariance among relatives is described in terms of the additive genetic variance  $V_A$  (variance of additive genetic effects, or additive genetic values), dominance variance  $V_D$  (variance of interaction effects between alleles in the same locus), and epistatic variance  $V_{AA}, V_{AD}, V_{DD}, \dots$  [remove dots (don't connect them :-)] (variance of interaction effects – additive and/or dominance effects – among loci) (Falconer & Mackay 1996; Lynch & Walsh 1998). These partitions are not dependent on numbers of genes or how they interact, but in practice the model is manageable only when the effects are independent from each other, requiring many important assumptions. These include random mating, and hence Hardy-Weinberg equilibrium (i.e. no inbred individuals), linkage equilibrium (independent segregation of loci, which requires many generations to achieve for tightly linked genes) and no selection. [the following equations are not described in the text - first mentioned in line 139 what  $V_{\{P\}}$  is.  $\sigma_{\{P\}}^2$  is never explained in this document]

$$\begin{aligned} V_P &= V_G + V_E \\ &= V_A + V_D + V_I + V_E \end{aligned} \quad (4)$$

$$\begin{aligned} \sigma_P^2 &= \sigma_G^2 + \sigma_E^2 \\ &= \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2 \end{aligned} \quad (5)$$

Many more terms may be included, such as maternal genetic effects, and genotype  $\times$  environment interactions. The model has unlimited opportunities for complexity. This is a strength, in that it is all-accommodating, and a weakness, in that datasets may allow to partition only a few components. [Previous sentence could perhaps be written as: “The advantage of this model is that it is all-accommodating, whereas a disadvantage is that datasets may allow to partition only a few components.”] In practice, assumptions must be made to reduce the complexity of the resemblance among relatives. Usually, the resemblance among relatives is assumed to depend only on additive genetic variance  $V_A$  and dominance variance  $V_D$ , so that the following sources of covariation are neglected:

- Epistatic variance (interaction effects among loci are small compared to additive and dominance effects)
- Environmental variance (effects of **shared environments** are assumed to be small enough [? small enough for what?])

**2.1.3.2 Heritability:** The models and summary statistics defined by Fisher and Wright have remained at the heart of quantitative genetics, not least because they provide ways to make predictions of important quantities, such as

**Broad-sense heritability**, the ratio of total genetic variance  $V_G$  to the overall phenotypic variance  $V_P$ :

$$\begin{aligned} H^2 &= V_G/V_P \\ &= (V_A + V_D + V_I)/V_P \\ H^2 &= \sigma_G^2/\sigma_P^2 \\ &= (\sigma_A^2 + \sigma_D^2 + \sigma_I^2)/\sigma_P^2 \end{aligned} \quad (6)$$

**Narrow-sense heritability**, the ratio of additive genetic variance  $V_A$  to the overall phenotypic variance  $V_P$ :

$$\begin{aligned} h^2 &= V_A/V_P \\ h^2 &= \sigma_A^2/\sigma_P^2 \end{aligned} \tag{7}$$

**2.1.3.3 Genetic correlation:** In a general quantitative genetic model in which, for each individual, two traits ( $P_1$  and  $P_2$ ) are each defined as the sum of a genetic value ( $G_1$  and  $G_2$ ) and an environmental value ( $E_1$  and  $E_2$ ), [OR RATHER USE: In a general quantitative genetic model, where two traits ( $P_1$  and  $P_2$ ) are each defined as the sum of a genetic value ( $G_1$  and  $G_2$ ) and an environmental value ( $E_1$  and  $E_2$ ),]

$$\begin{aligned} P_1 &= G_1 + E_1 \\ P_2 &= G_2 + E_2 \end{aligned}$$

the phenotypic correlation ( $\rho_{P_{12}}$ ) between the traits is defined as:

$$\rho_{P_{12}} = \frac{\sigma_{P_{12}}}{\sqrt{\sigma_{P_1}^2 \sigma_{P_2}^2}} \tag{8}$$

where  $\sigma_{P_{12}}$  is the phenotypic covariance and  $\sigma_{P_1}^2$  and  $\sigma_{P_2}^2$  are the variances of the phenotypic values for the two traits in the population,

and the genetic correlation ( $\rho_{G_{12}}$ ) of the traits is defined as:

$$rho_{G_{12}} = \frac{\sigma_{G_{12}}}{\sqrt{\sigma_{G_1}^2 \sigma_{G_2}^2}} \tag{9}$$

where  $\sigma_{G_{12}}$  is the genetic covariance and  $\sigma_{G_1}^2$  and  $\sigma_{G_2}^2$  are the variances of the genetic values for the two traits in the population. [this is written as one very long sentence from the beginning of the section to the end, but I think it works.]

#### 2.1.4 Basic questions remain

[is this paragraph necessary?] On the premise that many genes and environmental factors interact to impact the trait, it will be difficult to determine the action of individual causal genes. Many basic questions remain: What do the genes do; how do they interact; on what traits does natural selection act; why is there so much genetic variation; and can we expect continued genetic improvement in selection programmes? Ultimately, we want to know at the molecular level not just which genes are involved, whether structural or regulatory, but what specific mutation (nucleotide substitution, deletious [?], copy number variant, etc.) is responsible for genetic effects, and how the causal genes are controlled.

## 2.2 Single locus model for a quantitative trait

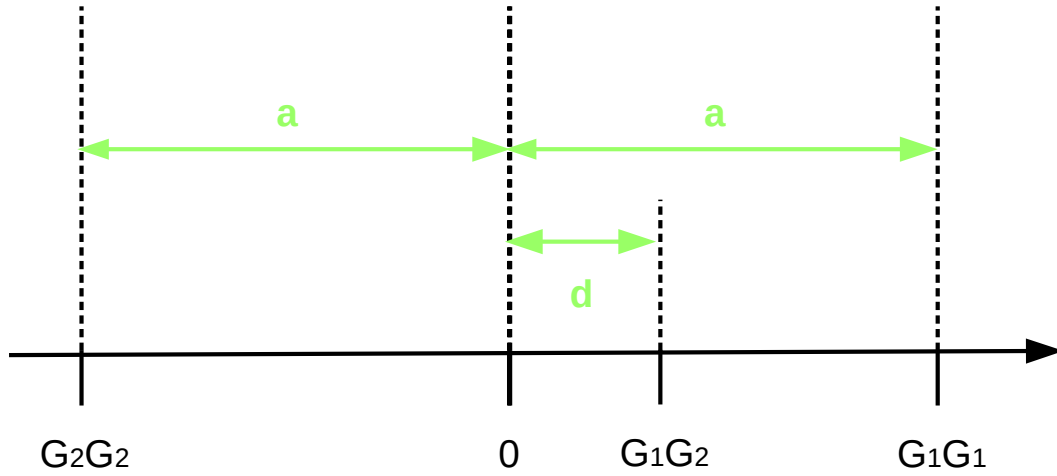
In this section we will be introducing the single locus model for a quantitative trait. Quantitative traits do not take discrete levels, instead they show continuous distributions. Although quantitative traits are most likely influenced by many loci, it helps to first consider the case of only one causal locus in the **single-locus model**. The single-locus model provides the theoretical basis for more complex models, namely the infinitesimal model and genomic models (statistical models describing the effects of marker loci). Furthermore, we must associate the genotypes in the population to the quantitative values of our trait. In the following, population mean, values (phenotypic value  $P$ , genotypic value  $G$ , and additive genetic value  $A$ ) and associated variance ( $V_P$ ,  $V_G$  and  $V_A$ ) will be defined for a single causal locus.

**Table 1:** Values for all Genotypes

Variable	Genotype	Values
$GV_{11}$	$A_1A_1$	$a$
$GV_{12}$	$A_1A_2$	$d$
$GV_{22}$	$A_2A_2$	$-a$

### 2.2.1 Genotypic Values

The values  $G_{ij}$  for each genotype  $A_iA_j$  are assigned as shown in Figure 1.



**Figure 1:** Genotypic Values

The origin (the zero value) for the genotypic values is placed half-way between the two homozygous genotypes  $A_2A_2$  and  $A_1A_1$ . Here we are assuming that  $A_1$  is the favorable allele. This leads to values of  $+a$  for genotype  $A_1A_1$  and of  $-a$  for genotype  $A_2A_2$ , where  $a$  is called additive gene action. The value of genotype  $A_1A_2$  is set to  $d$  and  $d$  is called dominance gene actions. Table 1 summarizes the values for all genotypes.

### 2.2.2 Population Mean

For the complete population, we can compute the **population mean** ( $\mu$ ) of all values at the locus  $G$ . Under the Hardy-Weinberg equilibrium,  $\mu$  corresponds to the expected value in a panmictic population, and is

computed as:

$$\begin{aligned}
\mu &= GV_{11} * f(A_1A_1) + GV_{12} * f(A_1A_2) + GV_{22} * f(A_2A_2) \\
&= a * p^2 + d * 2pq + (-a) * q^2 \\
&= (p - q)a + 2pqd
\end{aligned} \tag{10}$$

[jeg forstår ikke den her p? p og alle de andre skal sikker have . tegn? jeg har nu sat . tegn ind - just ask what I mean] Under the simplifying assumptions of the Hardy-Weinberg equilibrium, the genotype frequency ( $f$ ) of genotypes ( $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_2$ ) depends only on the allele frequency  $p$  of allele  $A_1$ , and the frequency  $q = 1 - p$  of allele  $A_2$ . The population mean then depends on the values of  $a$  and  $d$  and on the allele frequencies  $p$  and  $q$ . The larger the difference between  $p$  and  $q$  the more influence the value  $a$  has on  $\mu$  relatively to  $d$ , because for very different  $p$  and  $q$  the frequency – and contribution – of heterozygotes to  $\mu$  is very small (the product  $2pq$  is low). On the other hand, if  $p = q = 0.5$ , then  $\mu = 0.5d$ . For loci with  $d = 0$ , the population mean  $\mu = (p - q)a$  and hence, if in addition we have  $p = q$ , then  $\mu = 0$ .

### 2.2.3 Additive Genetic Values

The additive genetic value of an individual  $i$  is defined as two times the difference between the mean value of its offspring and the population mean. Applying this definition and using the parameters that we have computed so far leads to the following formulas for the additive genetic value of an individual with a certain genotype.

**2.2.3.1 Additive genetic value for  $A_1A_1$**  Assume that we have a given parent  $S$  with a genotype  $A_1A_1$  and we want to compute its additive genetic value. Let us further suppose that our single parent  $S$  is mated to a potentially infinite number of individuals from the idealized population, then we can deduce the following mean genotypic value for the offspring of parent  $S$ .

	Mates of $S$	
	$f(A_1) = p$	$f(A_2) = q$
Parent $S$		
$f(A_1) = 1$	$f(A_1A_1) = p$	$f(A_1A_2) = q$

Because parent  $S$  has genotype  $A_1A_1$ , the frequency  $f(A_1)$  of a  $A_1$  allele coming from  $S$  is 1 and the frequency  $f(A_2)$  of a  $A_2$  allele is 0. The expected genotypic value ( $\mu_{11}$ ) of the offspring of individual  $S$  can be computed as:

$$\mu_{11} = p * a + q * d \tag{11}$$

We can compute the additive genetic value ( $AV_{11}$ ) for individual  $S$  as shown in equation (12) while using the results given by equations (11) and (10).

$$\begin{aligned}
AV_{11} &= 2 * (\mu_{11} - \mu) \\
&= 2 (pa + qd - [(p - q)a + 2pqd]) \\
&= 2 (pa + qd - (p - q)a - 2pqd) \\
&= 2 (qd + qa - 2pqd) \\
&= 2 (qa + qd(1 - 2p)) \\
&= 2q (a + d(1 - 2p)) \\
&= 2q (a + (q - p)d)
\end{aligned} \tag{12}$$

Additive genetic values for parents with genotypes  $A_2A_2$  and  $A_1A_2$  are derived analogously.

**2.2.3.2 Additive genetic value for  $A_2A_2$**  First, we determine the expected genotypic value for the offspring of a parent  $S$  with genotype  $A_2A_2$

	Mates of parent $S$	
	$f(A_1) = p$	$f(A_2) = q$
Parent $S$		
$f(A_2) = 1$	$f(A_1A_2) = p$	$f(A_2A_2) = q$

The expected genotypic value ( $\mu_{22}$ ) of the offspring of individual  $S$  can be computed as:

$$\mu_{22} = pd - qa \quad (13)$$

The additive genetic value  $AV_{22}$  corresponds to

$$\begin{aligned}
AV_{22} &= 2 * (\mu_{22} - \mu) \\
&= 2 (pd - qa - [(p - q)a + 2pqd]) \\
&= 2 (pd - qa - (p - q)a - 2pqd) \\
&= 2 (pd - pa - 2pqd) \\
&= 2 (-pa + p(1 - 2q)d) \\
&= -2p (a + (q - p)d)
\end{aligned} \quad (14)$$

**2.2.3.3 Additive genetic value for  $A_1A_2$**  The genotype frequencies of the offspring of a parent  $S$  with a genotype  $A_1A_2$  is determined in the following table.

	Mates of parent $S$	
	$f(A_1) = p$	$f(A_2) = q$
Parent $S$		
$f(A_1) = 0.5$	$f(A_1A_1) = 0.5p$	$f(A_1A_2) = 0.5q$
$f(A_2) = 0.5$	$f(A_1A_2) = 0.5p$	$f(A_2A_2) = 0.5q$

The expected mean genotypic value of the offspring of parent  $S$  with genotype  $A_1A_2$  is computed as:

$$\mu_{12} = 0.5pa + 0.5d - 0.5qa = 0.5 [(p - q)a + d] \quad (15)$$

The additive genetic value  $AV_{12}$  corresponds to

$$\begin{aligned}
AV_{12} &= 2 * (\mu_{12} - \mu) \\
&= 2 (0.5(p - q)a + 0.5d - [(p - q)a + 2pqd]) \\
&= 2 (0.5pa - 0.5qa + 0.5d - pa + qa - 2pqd) \\
&= 2 (0.5(q - p)a + (0.5 - 2pq)d) \\
&= (q - p)a + (1 - 4pq)d \\
&= (q - p)a + (p^2 + 2pq + q^2 - 4pq)d \\
&= (q - p)a + (p^2 - 2pq + q^2)d \\
&= (q - p)a + (q - p)^2d \\
&= (q - p) [a + (q - p)d]
\end{aligned} \quad (16)$$



### 2.2.4 Summary of Additive Genetic Values

The term  $a + (q - p)d$  appears in all three additive genetic values. We replace this term by  $\alpha$  and summarize the results in the following table.

Genotype	Additive Genetic Value
$A_1A_1$	$2q\alpha$
$A_1A_2$	$(q - p)\alpha$
$A_2A_2$	$-2p\alpha$

### 2.2.5 Allele Substitution Effect

The difference between genotypes  $A_2A_2$  and  $A_1A_2$  is in the number of  $A_1$ -alleles.  $A_2A_2$  has zero  $A_1$ -alleles and  $A_1A_2$  has one  $A_1$ -allele. The effect of replacing an  $A_2$  allele by an  $A_1$  allele on the additive genetic value corresponds to the difference  $AV_{12} - AV_{22}$ . The computation of this difference between the additive genetic value  $AV_{12}$  and  $AV_{22}$  is:

$$\begin{aligned}
 AV_{12} - AV_{22} &= (q - p)\alpha - (-2p\alpha) \\
 &= (q - p)\alpha + 2p\alpha \\
 &= (q - p + 2p)\alpha \\
 &= (q + p)\alpha \\
 &= \alpha
 \end{aligned} \tag{17}$$

The analogous computation can be done by comparing the additive genetic values  $AV_{11}$  and  $AV_{12}$ :

$$\begin{aligned}
 AV_{11} - AV_{12} &= 2q\alpha - (q - p)\alpha \\
 &= (2q - (q - p))\alpha \\
 &= \alpha
 \end{aligned} \tag{18}$$

Because the differences between additive genetic values computed in (17) and (18) are equal, we can conclude that the additive genetic values show a linear dependence on the number of  $A_1$  alleles. This is the reason why the additive genetic values are also called additive effects, because adding a further  $A_1$  allele instead of a  $A_2$  allele has always the same effect on the additive genetic values, namely just adding the constant allele substitution effect  $\alpha$ .

### 2.2.6 Dominance Deviation

When looking at the difference between the genotypic value  $GV_{ij}$  and the additive genetic value  $AV_{ij}$  for each of the three genotypes, we get the following results:

$$\begin{aligned}
 GV_{11} - AV_{11} &= a - 2q\alpha \\
 &= a - 2q[a + (q - p)d] \\
 &= a - 2qa - 2q(q - p)d \\
 &= a(1 - 2q) - 2q^2d + 2pqd \\
 &= [(p - q)a + 2pqd] - 2q^2d \\
 &= \mu + D_{11}
 \end{aligned} \tag{19}$$

$$\begin{aligned}
GV_{12} - AV_{12} &= d - (q - p)\alpha \\
&= d - (q - p)[a + (q - p)d] \\
&= [(p - q)a + 2pqd] + 2pqd \\
&= \mu + D_{12}
\end{aligned} \tag{20}$$

$$\begin{aligned}
GV_{22} - AV_{22} &= -a - (-2p\alpha) \\
&= -a + 2p[a + (q - p)d] \\
&= [(p - q)a + 2pqd] - 2p^2d \\
&= \mu + D_{22}
\end{aligned}$$

The difference all contain the population mean  $\mu$  plus a certain deviation. This deviation term is called **dominance deviation**. It corresponds to the part of genotypic values which are not accounted for by additive effects – and linear allelic substitution effects. Therefore, it captures the non-linear relationships between genotypic values and the number of  $A_1$  alleles (zero in  $A_2A_2$ , 1 in  $A_1A_2$ , 2 in  $A_1A_1$ ).

### 2.2.7 Summary of Values

The following table summarizes all genotypic values, all additive genetic values and the dominance deviations.

Genotype	Genotypic value	Additive Genetic Value	Dominance Deviation
$A_iA_j$	$GV_{ij}$	$AV_{ij}$	$D_{ij}$
$A_1A_1$	$a$	$2q\alpha$	$-2q^2d$
$A_1A_2$	$d$	$(q - p)\alpha$	$2pqd$
$A_2A_2$	$-a$	$-2p\alpha$	$-2p^2d$

The formulas in the above shown table assume that  $A_1$  is the favorable allele with frequency  $f(A_1) = p$ . The allele frequency of  $A_2$  is  $f(A_2) = q$ . Since we have a bi-allelic locus,  $p + q = 1$ .

Based on the definition of dominance deviation, the genotypic values  $GV_{ij}$  can be decomposed into the following components: population mean ( $\mu$ ), additive genetic value ( $AV_{ij}$ ) and dominance deviation ( $D_{ij}$ ) according to equation (21).

$$GV_{ij} = \mu + AV_{ij} + D_{ij} \tag{21}$$

Taking expected values on both sides of equation (21) and knowing that the population mean  $\mu$  was defined as the expected value of the genotypic values in the population, i.e.  $E[GV] = \mu$ , it follows that the expected values of both the additive genetic values and the dominance deviations must be 0. More formally, we have

$$\begin{aligned}
E[GV] &= E[\mu + AV + D] \\
&= E[\mu] + E[AV] + E[D] \\
&= \mu
\end{aligned} \tag{22}$$

From the last line in equation (22), it follows that  $E[AV] = E[D] = 0$ . This also shows that both additive genetic values and dominance deviations are defined as deviation from the population mean.

### 2.2.8 Variances

The population mean  $\mu$  and the additive genetic values were defined as expected values ( $\mu$ : expected value of genotypes in a given generation; additive genetic value: expected advantage of the offspring of each genotype, relative to  $\mu$ ). Their main purpose is to assess the state of a given population with respect to a certain genetic locus and its effect on a phenotypic trait of interest.

In statistics the measure that is most often used to assess variation in a certain population is called **variance**. For any given discrete random variable  $X$ , the variance is defined as the second central moment of  $X$  which is computed as shown in equation (23).

$$Var[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i) \quad (23)$$

where  $\mathcal{X}$ : set of all possible  $x$ -values  
 $f(x_i)$ : probability that  $x$  assumes the value of  $x_i$   
 $\mu_X$ : expected value  $E[X]$  of  $X$

In this section we will be focusing on separating the obtained variances into different components according to their causative sources. Applying the definition of variance given in equation (23) [check reference] to the genotypic values  $GV_{ij}$ , we obtain the following expression:

$$\begin{aligned} \sigma_G^2 = Var[V] &= (GV_{11} - \mu)^2 * f(A_1A_1) \\ &+ (GV_{12} - \mu)^2 * f(A_1A_2) \\ &+ (GV_{22} - \mu)^2 * f(A_2A_2) \end{aligned} \quad (24)$$

where  $\mu = (p - q)a + 2pqd$  is the population mean.

Based on the decomposition of the genotypic value  $GV_{ij}$  given in (21), the difference between  $GV_{ij}$  and  $\mu$  can be written as the sum of the additive genetic value and the dominance deviation. Then  $\sigma_G^2$  can be written as:

$$\begin{aligned} \sigma_G^2 = Var[GV] &= (AV_{11} + D_{11})^2 * f(A_1A_1) \\ &+ (AV_{12} + D_{12})^2 * f(A_1A_2) \\ &+ (AV_{22} + D_{22})^2 * f(A_2A_2) \end{aligned} \quad (25)$$

Inserting the expressions for the additive genetic values  $AV_{ij}$  and for the dominance deviation  $D_{ij}$  found earlier and simplifying the equation leads to the result in (26).

$$\begin{aligned} \sigma_G^2 &= 2pq\alpha^2 + (2pqd)^2 \\ &= \sigma_A^2 + \sigma_D^2 \end{aligned} \quad (26)$$

The formula in equation (26) shows that  $\sigma_G^2$  consists of two components. The first component  $\sigma_A^2$  is called the **additive genetic variance** and the second component  $\sigma_D^2$  is termed **dominance variance**. Here  $\sigma_A^2$  corresponds to the variance of the additive genetic values. The variance of additive genetic values is also called the additive genetic variance, because as we have already seen the additive genetic values are additive in the number of favorable alleles. In populations where there is no additive genetic variance, individuals all have the same additive genetic value. Therefore, they will produce offspring with the same expected advantage (zero), and selection cannot generate any improvement over generations. Because  $\sigma_D^2$  corresponds to the variance of the dominance deviation effects it is called dominance variance.

## 2.3 Multiple locus model for a quantitative trait

When only a single locus is considered, the genotypic values ( $GV_{ij}$ ) can be decomposed according to equation (21) into population mean, additive genetic value and dominance deviation. When a genotype refers to more than one locus, the genotypic value may contain an additional deviation caused by non-additive combination effects.

### 2.3.1 Epistatic Interaction

Let  $GV_A$  be the genotypic value of locus  $A$  and  $GV_B$  denote the genotypic value of a second locus  $B$ , then the total genotypic value  $GV$  attributed to both loci  $A$  and  $B$  can be written as:

$$GV = GV_A + GV_B + I_{AB} \quad (27)$$

where  $I_{AB}$  is the deviation from additive combination of these genotypic values. When computing the population mean earlier in this chapter, we assumed that  $I$  was zero for all combinations of genotypes. If  $I$  is not zero for any combination of genes at different loci, those genes are said to **interact** with each other or to exhibit **epistasis**. The deviation  $I$  is called interaction deviation or epistatic deviation. If  $I$  is zero, the genes are called to act additively between loci. Hence *additive action* may mean different things. When referring to one locus, it means absence of dominance. When referring to different loci, it means absence of epistasis.

Interaction between loci may occur between pairs or between higher numbers of different loci. The complex nature of higher order interactions, i.e., interactions between higher number of loci does not need to concern us. Because in the total genotypic value  $GV$ , interaction deviations of all sorts are treated together in an overall interaction deviation  $I$ .

Applying the decomposition of the genotypic values  $GV_A$  of locus  $A$  and  $GV_B$  of locus  $B$  as shown in (21) leads to:

$$\begin{aligned} GV &= GV_A + GV_B + I_{AB} \\ &= \mu_A + AV_A + D_A + \mu_B + AV_B + D_B + I_{AB} \end{aligned} \quad (28)$$

Collecting terms in (28) as follows:

$$\begin{aligned} \mu &= \mu_A + \mu_B \\ AV &= AV_A + AV_B \\ D &= D_A + D_B \\ I &= I_{AB} \end{aligned} \quad (29)$$

The decomposition shown in (28) and the collection of variables (see (29)) can be generalized to more than two loci. This leads to the following generalized decomposition of the overall total genotypic value  $GV$  for the case of multiple loci affecting a certain trait of interest.

$$GV = \mu + AV + D + I \quad (30)$$

where  $AV$  is the sum of the additive genetic values attributable to the separate loci and  $D$  is the sum of all dominance deviations.

On the other hand, the dominance deviation measures the effect of a certain genotype occurring in an individual and the epistatic deviation estimates the effects of combining different genotypes at different loci in the genome. But because parents do not pass complete genotypes nor do they pass stretches of DNA with several unlinked loci, but only a random collection of its alleles, it is really the additive genetic value that is of primary importance in assessing the genetic potential of a given selection candidate.

### 2.3.2 Interaction Variance

If genotypes at different loci show epistatic interaction effects as described in section 2.3.1, the interactions give rise to an additional variance component called  $V_I$ , which is the variance of interaction deviations. This new variance component  $V_I$  can be further decomposed into sub-components. The first level of sub-components is according to the number of loci that are considered. Two-way interactions involve two loci, three-way interactions consider three loci and in general  $n$ -way interactions arise from  $n$  different loci. Epistatic interaction can be further decomposed according to whether they involve additive effects, dominance deviations or both, across loci. In general, interaction effects explain only a very small amount of the overall genotypic variation.