The one-locus model

Lecture 3 ▪ Quantitative Genetics

Grum Gebreyesus ▪ ASAP-Bio Knowledge Hub

asap-bio.github.io/lec-qg-one-locus.html

Contents

1 Genotype and allele frequencies

Alleles of a gene vary; there may be two or many allelic forms, and some of these variants affect quantitative traits. Alleles occur in combinations, the genotypes. With two alleles A and a there are three genotypes, AA, Aa and aa, with frequencies f_{AA}, f_{Aa} and f_{aa} summing to one.

Allele frequencies follow from genotype frequencies by counting: each homozygote carries two copies, each heterozygote one.

$$p = \Pr(A) = f_{AA} + \tfrac{1}{2} f_{Aa}, \qquad q = \Pr(a) = f_{aa} + \tfrac{1}{2} f_{Aa}, \qquad p + q = 1$$

2 Hardy-Weinberg proportions

Under random mating, with no selection, mutation or migration, and in a large population, each allele in one generation has an equal chance of being copied into the next. No alleles are added, removed or changed. Genotypes in the offspring are then formed by combining two independently drawn alleles.

Random mating, gamete by gamete

1. A mother of genotype AA (frequency f_{AA}) transmits A with certainty; a mother Aa (frequency f_{Aa}) transmits A with probability $\frac{1}{2}$. Summing over maternal genotypes, the probability that the maternal gamete carries A is $f_{AA} + \frac{1}{2} f_{Aa} = p$.
2. The same holds for the paternal gamete, and under random mating the two draws are *independent*.
3. Therefore $\Pr(AA) = p \times p = p^2$, $\Pr(aa) = q \times q = q^2$, and $\Pr(Aa) = pq + qp = 2pq$, the factor 2 because A may come from either parent.

HARDY-WEINBERG PROPORTIONS

$$f_{AA} = p^2, \qquad f_{Aa} = 2pq, \qquad f_{aa} = q^2$$

These are reached after a **single** generation of random mating, and are the default assumption throughout the course. Note what this means in practice: even if selection has severely distorted genotype frequencies among adults, one generation of random mating among those adults restores Hardy-Weinberg proportions among the zygotes, provided mating itself is random with respect to the locus.

3 The one-locus model

Our first goal is to decompose a phenotype into a genetic part and an environmental part. Consider a locus with two alleles, and write the expected phenotype of each genotype as a conditional expectation, exactly the object defined in the previous lecture:

$$\begin{array}{ll} AA \text{ (freq } p^2) & E(y \mid AA) = x_{AA} \\ Aa \text{ (freq } 2pq) & E(y \mid Aa) = x_{Aa} \\ aa \text{ (freq } q^2) & E(y \mid aa) = x_{aa} \end{array}$$

The population mean is the average of these three, weighted by their Hardy-Weinberg frequencies:

$$\mu = E(y) = p^2 x_{AA} + 2pq\, x_{Aa} + q^2 x_{aa}$$

and the phenotype of an individual is written

$$y = \mu + g + e$$

with g the genotypic deviation and e the environmental deviation, both measured as departures from μ.

4 Average effect of a genotype

Define the genotypic value g of each genotype as its deviation from the population mean:

$$g_{ij} = x_{ij} - \mu \qquad \text{so that} \qquad E(y \mid AA) = \mu + g_{AA}, \quad E(y \mid Aa) = \mu + g_{Aa}, \quad E(y \mid aa) = \mu + g_{aa}$$

By construction these deviations average to zero over the population: $p^2 g_{AA} + 2pq\, g_{Aa} + q^2 g_{aa} = 0$.

5 Average effect of an allele

The following step underpins quantitative genetics. Genotypes are *not* transmitted to offspring, only **alleles** are, so the quantity required is the value associated with an allele rather than with a genotype.

The **average effect of an allele** is the expected deviation from μ of an individual known to carry that allele, with the second allele drawn at random from the population:

$$\alpha_A = E(y \mid G = A-) - \mu = p\, g_{AA} + q\, g_{Aa}$$
$$\alpha_a = E(y \mid G = a-) - \mu = p\, g_{Aa} + q\, g_{aa}$$

The logic is direct: given one allele is A, the other is A with probability p (giving genotype AA) and a with probability q (giving Aa). Weight the genotypic values accordingly.

THE ALLELE SUBSTITUTION EFFECT

$$\beta = \alpha_A - \alpha_a$$

β is the expected change in the trait when one a allele is replaced by one A. It is the single most useful number attached to a locus, and it is what the α_j in the many-locus model of Lecture 1 refers to. Note that average effects satisfy $p\alpha_A + q\alpha_a = 0$.

6 Breeding value and dominance deviation

Since alleles are what get transmitted, the part of an individual's genotypic value that it can pass on is the sum of the average effects of the two alleles it carries. That sum is the **breeding value**.

$$a_{ij} = \alpha_i + \alpha_j \qquad \text{(the additive genetic value)}$$

The breeding value will generally not equal the genotypic value exactly. What is left over is the **dominance deviation**:

$$\delta_{ij} = g_{ij} - (\alpha_i + \alpha_j)$$

giving the decomposition of the genotypic value at one locus:

$$g_{ij} = \underbrace{\alpha_i + \alpha_j}_{\text{additive, transmissible}} + \underbrace{\delta_{ij}}_{\text{dominance, not transmissible}}$$

SIGNIFICANCE OF THE DECOMPOSITION

A parent passes on *one allele*, not a genotype, so the dominance deviation, which arises from the specific *combination* of two alleles, is broken up at meiosis and not transmitted. This is precisely why selection responds to the additive component and not to dominance, and why the breeder's equation later contains σ^2_A rather than σ^2_G.

Putting it together, the genetic model for one locus is

$$y = \mu + g + e = \mu + a + \delta + e$$

7 Additive gene action as a special case

Consider the case in which there is no dominance. Write the genotypic values in the symmetric form $x_{aa} = x - s$, $x_{Aa} = x$, $x_{AA} = x + s$, so that the heterozygote sits exactly midway between the homozygotes. Then $\mu = x + (p - q)s$ and the table becomes:

Genotype	x_{ij}	g_{ij}	$\alpha_i + \alpha_j$	δ_{ij}
aa	$x - s$	$-2ps$	$-2ps$	0
Aa	x	$(p - q)s$	$(p - q)s$	0
AA	$x + s$	$2qs$	$2qs$	0

Under purely additive gene action every dominance deviation is zero, the genotypic value equals the breeding value, and $\beta = \alpha_A - \alpha_a = s$ independently of allele frequency. This is the reference case against which dominance and overdominance are judged.

8 Variances: gametic, additive and dominance

8.1 Gametic variance

The gametic variance is the variance of the effect transmitted by a single allele. A gamete carries A with probability p (contributing α_A) and a with probability q (contributing α_a). Since $p\alpha_A + q\alpha_a = 0$, the mean is zero and

$$\sigma^2_{\text{gametic}} = p\,\alpha_A^2 + q\,\alpha_a^2 = pq\beta^2$$

8.2 Additive genetic variance

An individual carries two independently drawn gametes, so the additive variance is twice the gametic variance:

$$\sigma^2_A = 2pq\beta^2$$

8.3 Dominance variance

The dominance variance is the variance of the deviations δ_{ij}, which have mean zero:

$$\sigma^2_D = p^2\delta^2_{AA} + 2pq\,\delta^2_{Aa} + q^2\delta^2_{aa}$$

and, writing d for the dominance effect (the deviation of the heterozygote from the midpoint of the homozygotes), this evaluates to the compact form

$$\sigma^2_D = (2pq\,d)^2$$

THE VARIANCE DECOMPOSITION

$$\sigma^2_G = \sigma^2_A + \sigma^2_D \qquad \text{and} \qquad \sigma^2_y = \sigma^2_A + \sigma^2_D + \sigma^2_E$$

The additive and dominance components are **uncorrelated** by construction, which is what allows the variances to be added without a covariance term, exactly the condition flagged in the variance section of Lecture 2. Note also that $\sigma^2_D = \sigma^2_G - \sigma^2_A$ holds *for one locus only*; with several loci the difference would also contain epistatic variance.

9 Everything depends on allele frequency

The following dependence is easily overlooked. From the results derived above:

- μ depends on the allele frequencies.
- The α's depend on the allele frequencies.
- Therefore the breeding values, the dominance deviations, and σ^2_A and σ^2_D all depend on the allele frequencies too.

The *genotypic values* x_{AA}, x_{Aa}, x_{aa} are properties of the biology and do not change. Everything else on that list is a property of the **population**. The same locus, with the same biology, has a different additive variance in two populations that differ in allele frequency, and $\sigma^2_A = 2pq\beta^2$ necessarily goes to zero as $p \to 0$ or $p \to 1$. A locus fixed for one allele contributes nothing to genetic variance, however large its biological effect.

PARTICULARLY IMPORTANT

1. Hardy-Weinberg proportions are the default assumption.
2. The genetic model is $y = \mu + g + e = \mu + a + \delta + e$.
3. Only the additive part a is transmitted to offspring.
4. All genetic parameters are population parameters, not constants of the gene.

10 Exercises

Exercise 3.1

Sickle-cell anaemia. The locus has two alleles C and c. Individuals cc have anaemia; Cc individuals have partial protection against malaria. In parts of West Africa the frequency of c is $p_c = 0.2$.

1. What is the expected genotype distribution at birth?
2. cc individuals have far fewer children than CC or Cc, and Cc have more children than CC, so allele frequencies among *adults* depart from Hardy-Weinberg proportions. Assuming people marry and have children independently of their C/c genotype, how far will the zygote distribution of the next generation deviate from Hardy-Weinberg proportions?

1. With $q = p_c = 0.2$ and $p = p_C = 0.8$: $CC = p^2 = 0.64$, $Cc = 2pq = 0.32$, $cc = q^2 = 0.04$.
2. Differential fertility among genotypes is a form of natural selection, and selection is a violation of the Hardy-Weinberg conditions, so the *adult* population is not in HW proportions. But the question asks about the *zygotes* of the next generation. Since mating is random with respect to the C/c genotype, the zygotes are formed by independently sampling two alleles from the adult gamete pool, so Hardy-Weinberg proportions are **restored in a single generation**, at the new allele frequency.
The distinction is that selection changes the allele *frequencies*, while random mating re-establishes Hardy-Weinberg *proportions* at the new frequencies.

Exercise 3.2

Crossing two populations. One locus with alleles A and a. In population 1 the frequency of A is $p_1 = 0.7$; in population 2 it is $p_2 = 0.4$.

1. Compute the expected genotype frequencies in each subpopulation, assuming HW proportions.
2. Crossbred offspring are produced with one randomly chosen parent from each population. What is the frequency of A in the crossbreds, and what is their expected genotype distribution? Are the crossbreds in Hardy-Weinberg proportions? Why or why not?
3. The crossbreds are mated randomly among themselves to produce F2 animals. What are the allele and genotype frequencies among the F2?

1. Population 1: $AA = 0.49$, $Aa = 0.42$, $aa = 0.09$. Population 2: $AA = 0.16$, $Aa = 0.48$, $aa = 0.36$.
2. Allele frequency in the crossbreds is the average of the parental frequencies, $\bar{p} = (0.7+0.4)/2 = 0.55$. But the genotype frequencies are *not* $\bar{p}^2, 2\bar{p}\bar{q}, \bar{q}^2$. Each crossbred receives one gamete from each population, so

$$\begin{aligned}\Pr(AA) &= p_1 p_2 = 0.7 \times 0.4 = 0.28\\ \Pr(Aa) &= p_1 q_2 + q_1 p_2 = 0.7 \times 0.6 + 0.3 \times 0.4 = 0.54\\ \Pr(aa) &= q_1 q_2 = 0.3 \times 0.6 = 0.18\end{aligned}$$

HW would predict $0.3025, 0.495, 0.2025$. The crossbreds have *more* heterozygotes than HW (0.54 against 0.495), so they are **not** in Hardy-Weinberg proportions. The reason: the two gametes are drawn from populations with *different* allele frequencies, so the two draws are not identically distributed, which is one of the HW conditions.
3. Now mating is random within a single population. Allele frequency stays at $\bar{p} = 0.55$ (no selection), and one generation of random mating restores HW proportions: $AA = 0.3025$, $Aa = 0.495$, $aa = 0.2025$. This is the classic F1-to-F2 loss of heterozygosity.

Exercise 3.3

Compute the means. Set $p = 0.7$. For each of the three models of gene action below, compute μ, α_A, α_a and β.

Model	x_{aa}	x_{Aa}	x_{AA}
Additive	1	2	3
Dominant	1	2	2
Overdominant	1	2	1

Use $\mu = p^2 x_{AA} + 2pq\, x_{Aa} + q^2 x_{aa}$, $g_{ij} = x_{ij} - \mu$, $\alpha_A = p\, g_{AA} + q\, g_{Aa}$, $\alpha_a = p\, g_{Aa} + q\, g_{aa}$, with $p = 0.7, q = 0.3$.
Additive: $\mu = 0.09(1) + 0.42(2) + 0.49(3) = 2.4$. Then $g_{aa} = -1.4$, $g_{Aa} = -0.4$, $g_{AA} = 0.6$. $\alpha_a = 0.3(-1.4) + 0.7(-0.4) = -0.7$; $\alpha_A = 0.3(-0.4) + 0.7(0.6) = 0.3$; $\beta = \alpha_A - \alpha_a = 1.0$.
Dominant: $\mu = 0.09(1) + 0.42(2) + 0.49(2) = 1.91$. $g_{aa} = -0.91$, $g_{Aa} = 0.09$, $g_{AA} = 0.09$. $\alpha_a = 0.3(-0.91) + 0.7(0.09) = -0.21$; $\alpha_A = 0.3(0.09) + 0.7(0.09) = 0.09$; $\beta = 0.30$.
Overdominant: $\mu = 0.09(1)+0.42(2)+0.49(1) = 1.42$. $g_{aa} = -0.42$, $g_{Aa} = 0.58$, $g_{AA} = -0.42$. $\alpha_a = 0.3(-0.42) + 0.7(0.58) = 0.28$; $\alpha_A = 0.3(0.58) + 0.7(-0.42) = -0.12$; $\beta = -0.40$.
Under overdominance β is *negative* at $p = 0.7$, although the A allele is not inferior in any biological sense. The substitution effect is a population quantity and changes sign with allele frequency.

Exercise 3.4

Compute and plot the variances. For the same three scenarios, compute V_G, V_A and V_D, reusing the μ, α_A, α_a you already have. Then plot all three as functions of p over $(0, 1)$ using the

course R code (`var.R`). Interpret the shapes: why are the maxima and minima where they are, and how would you identify the allele frequencies at which $V_A = 0$?

Use $V_G = p^2 g_{AA}^2 + 2pq\, g_{Aa}^2 + q^2 g_{aa}^2$, $V_A = 2pq\beta^2$, and $V_D = V_G - V_A$ (valid for a single locus only; with several loci the remainder would also contain epistatic variance).
Reference implementation:

```
 mu <- function(x_aa,x_Aa,x_AA,p){ q <- 1-p; q^2*x_aa + 2*p*q*x_Aa +
p^2*x_AA }
Vg <- function(x_aa,x_Aa,x_AA,p){ q <- 1-p; m <- mu(x_aa,x_Aa,x_AA,p)
q^2*x_aa^2 + 2*p*q*x_Aa^2 + p^2*x_AA^2 - m^2 }
alphaA <- function(x_aa,x_Aa,x_AA,p){ q <- 1-p; p*x_AA + q*x_Aa -
mu(x_aa,x_Aa,x_AA,p) }
alphaa <- function(x_aa,x_Aa,x_AA,p){ q <- 1-p; p*x_Aa + q*x_aa -
mu(x_aa,x_Aa,x_AA,p) }
Va <- function(x_aa,x_Aa,x_AA,p){ q <- 1-p
beta <- alphaA(x_aa,x_Aa,x_AA,p) - alphaa(x_aa,x_Aa,x_AA,p); 2*p*q*beta^2 }
p <- seq(0.01, 0.99, 0.01)
```

Interpretation. All variances vanish at $p = 0$ and $p = 1$, because a fixed locus is not variable and therefore contributes no variance whatever its biological effect. In the additive case β is constant, so $V_A = 2pq\beta^2$ is a symmetric parabola peaking at $p = 0.5$. Under dominance the peak shifts away from 0.5, because β itself varies with p. Under overdominance β passes through zero at an intermediate frequency, so $V_A = 0$ there while V_G and V_D remain positive, a locus with substantial genetic variance but *no* additive variance, and therefore no response to selection.
To find where $V_A = 0$, solve $\beta(p) = \alpha_A - \alpha_a = 0$ for p, since $2pq > 0$ strictly inside the interval.