

Means and Variances

Gustavo de los Campos (gustavoc@msu.edu)

September 8th, 2025

I. Single-locus model

This section presents the math for deriving the mean and genetic variance in a single locus model under Hardy-Weinberg equilibrium. The notation follows closely the one used Chapter 7 of Falconer & Mackay (1996).

Genotypes

Consider a single locus with two alleles (A_1/A_2) and let p denote the frequency of allele A_1 , that is $P(A_1) = p$, implying, $P(A_2) = 1 - p = q$.

For this locus we have three possible genotypes: A_1A_1 , A_1A_2 , and A_2A_2 . Table 1 gives the expected genotype frequencies under Hardy-Weinberg Equilibrium (HWE).

Table 1: Genotypes and expected frequencies under Hardy-Weinberg Equilibrium

Genotype	HWE Frequency
A_1A_1	p^2
A_1A_2	$2pq$
A_2A_2	q^2

where $q = p(A_2) = 1 - p$.

Genetic values

The genetic value of an individual (G) is the expected phenotype given the genotype $G = E[Y|Genotype]$.

Since we have three possible genotypes, for a single-locus model there are three possible genetic values (or means):

- $G(A_1A_1)$,
- $G(A_1A_2)$, and
- $G(A_2A_2)$.

Following the parameterization used in Falconer & Mackay (1996), for the single locus model we represent the three possible genetic values using

- $G(A_1A_1) = \mu + a$,
- $G(A_1A_2) = \mu + d$, and
- $G(A_2A_2) = \mu - a$.

These three means can be simplified by subtracting μ such that

- $G(A_1A_1) = a$,
- $G(A_1A_2) = d$, and
- $G(A_2A_2) = -a$.

With this parameterizations, we model the three means (all represented as deviations from μ) using two parameters (a , and d).

The population mean under Hardy-Weinberg equilibrium

To compute the population mean, we update **Table 1** adding a column for the genetic values. After some simplifications, including using $p^2 - q^2 = p - q$, we get $E(G) = a(p - q) + 2pqd$.

Table 2: Computing the population mean

Genotype	HWE Frequency	G	Freq \times G
A_1A_1	p^2	a	p^2a
A_1A_2	$2pq$	d	$2pqd$
A_2A_2	q^2	-a	$-q^2a$
Expected value			$a(p - q) + 2pqd$

Additive action: A model with $d = 0$ is a strictly additive model in which the genetic value of the heterozygous is half-way in between the genetic value of the two homozygous.

Genetic variance

Recall that $\sigma_G^2 = Var(G) = E(G^2) - E(G)^2$. We already have an expression for $E(G)$; to derive $E(G^2)$, we can update Table 3 by adding a column for G^2 , multiplying by frequency, and taking column-wise sums to obtain the expected values.

Table 3: Computing $E[G^2]$

Genotype	HWE Frequency	G	$Freq \times G$	$Freq \times G^2$
A_1A_1	p^2	-a	$-p^2a$	p^2a^2
A_1A_2	$2pq$	d	$2pqd$	$2pqd^2$
A_2A_2	q^2	a	q^2a	q^2a^2
Expected Value			$a(p - q) + 2pqd$	$a^2(p^2 + q^2) + 2pqd^2$

Using $E[G^2] = a^2(p^2 + q^2) + 2pqd^2$ and $E(G) = a(p - q) + 2pqd$ in $Var(G) = E(G^2) - E(G)^2$ leads, after simplifications, to the following expression for the genetic variance of a single locus:

$$\sigma_G^2 = 2pq(a + d[q - p])^2 + (2pqd)^2.$$

Acknowledgments: I Alexa Lupi (PhD) for reading and commenting an early version of this handout.

References

Falconer, D.S. and Mackay, T.F.C. (1996) Introduction to Quantitative Genetics. 4th Edition, Addison Wesley Longman, Harlow.

Regression approach

Let's consider estimating $\{\mu, a, d\}$ using a regression model on two variables $X \in \{-1, 0, 1\}$ and $H = \{1 \text{ if } A_1 A_2 ; 0 \text{ otherwise}\}$