Building Blocks of Population Genetics
Jack Dekkers

Genotype frequencies, Allele frequencies, Haplotype frequencies, Linkage Disequilibrium, Linkage

Single locus $\rightarrow$ allele (or gene) frequencies $\rightarrow$ genotype frequencies

Consider a single locus in a random mating outbred population. The locus has alleles $A_1$ and $A_2$ with allele (or gene) frequencies $p$ and $q$.

Under random mating (Hardy Weinberg Equilibrium), the allele received from one parent is independent of the allele received from the other parent, resulting in the following relationship between allele and genotype frequencies:

Table 1: Genotype probabilities, single locus two-allele case

<table>
<thead>
<tr>
<th>Paternal allele</th>
<th>Maternal allele</th>
<th>Marginal prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr(A_1) = p$</td>
<td>$Pr(A_2) = q$</td>
<td></td>
</tr>
<tr>
<td>$p^2$</td>
<td>$pq$</td>
<td>$p^2 + pq = p(p + q) = p$</td>
</tr>
<tr>
<td>$pq$</td>
<td>$q^2$</td>
<td>$pq + q^2 = q(p + q) = q$</td>
</tr>
</tbody>
</table>

This results in the HWE genotype frequencies: $p^2$, $2pq$, $q^2$

With multiple loci we also need to consider haplotypes and their frequencies, and relationships between allele, haplotype, and genotype frequencies.

Haplotype = the combination of alleles at $>1$ locus that an individual inherited from a parent. E.g. an individual with (unordered) genotype $A_1A_2$ and $B_1B_2$ at loci A and B, can have the following combinations of haplotype pairs (separated by /):

$A_1B_1/A_2B_2$ $\rightarrow$ alleles $A_1$ and $B_1$ received from one parent and $A_2$ and $B_2$ from the other

$A_1B_2/A_2B_1$ $\rightarrow$ alleles $A_1$ and $B_2$ received from one parent and $A_2$ and $B_1$ from the other

Haplotype frequency = frequency of a given haplotype in a population
With two loci with two alleles, there are 4 possible haplotypes, 16 ordered genotypes (ordered based on haplotypes), and 9 unordered genotypes (see tables 2,3)

**Table 2: Haplotype frequencies and genotype frequencies under random mating (HWE)**

<table>
<thead>
<tr>
<th>Haplotype - freq</th>
<th>Maternal haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_1B_1$</td>
</tr>
<tr>
<td>$A_1B_1$</td>
<td>$r$</td>
</tr>
<tr>
<td>$A_1B_2$</td>
<td>$s$</td>
</tr>
<tr>
<td>$A_2B_1$</td>
<td>$t$</td>
</tr>
<tr>
<td>$A_2B_2$</td>
<td>$u$</td>
</tr>
</tbody>
</table>

**Table 3: Unordered and ordered genotypes and their frequencies under random mating**

<table>
<thead>
<tr>
<th>Unordered genotypes</th>
<th>Frequency</th>
<th>Possible ordered genotypes and their frequencies (from Table 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1B_1B_1$</td>
<td>$r^2$</td>
<td>$A_1B_1/A_1B_1$ $r^2$</td>
</tr>
<tr>
<td>$A_1A_1B_1B_2$</td>
<td>$2rs$</td>
<td>$A_1B_1/A_1B_2$ rs $A_1B_1/A_2B_1$ sr</td>
</tr>
<tr>
<td>$A_1A_1B_2B_1$</td>
<td>$s^2$</td>
<td>$A_1B_2/A_1B_1$ $s^2$</td>
</tr>
<tr>
<td>$A_1A_2B_1B_1$</td>
<td>$2rt$</td>
<td>$A_2B_1/A_2B_1$ rt $A_2B_1/A_1B_1$ tr</td>
</tr>
<tr>
<td>$A_1A_2B_2B_1$</td>
<td>$2ru+2st$</td>
<td>$A_1B_2/A_2B_2$ ru $A_1B_2/A_2B_1$ st $A_2B_1/A_2B_2$ tr $A_2B_1/A_1B_1$ ur</td>
</tr>
<tr>
<td>$A_1A_2B_2B_2$</td>
<td>$t^2$</td>
<td>$A_2B_2/A_2B_1$ $t^2$</td>
</tr>
<tr>
<td>$A_2A_1B_1B_2$</td>
<td>$2tu$</td>
<td>$A_2B_1/A_2B_2$ tu $A_2B_1/A_2B_1$ ut</td>
</tr>
<tr>
<td>$A_2A_2B_1B_2$</td>
<td>$u^2$</td>
<td>$A_2B_2/A_2B_1$ $u^2$</td>
</tr>
</tbody>
</table>

The unordered genotype is what is obtained from genotyping, i.e. the genotype at each locus

**What is the relationship between haplotype frequencies and the frequencies of alleles that make up each haplotype?** This depends on whether the alleles at the two loci are **dependent** or **independent**:

- **Independence** of alleles at A and B
  - **Linkage Equilibrium** haplotype frequencies (product of allele freqs)
- **Dependence** of alleles at A and B
  - **Linkage Disequilibrium**

**HWE** multi-locus genotype frequencies (product of haplotype freqs)
Haplotype probabilities (= frequencies) - two-allele case:
What is the probability of a progeny to receive from a parents: allele $A_i$ at locus $A$ and allele $B_j$ at locus $B$?

**i) if the alleles at the two loci are independent from each other**

- joint probability = product of marginal probabilities

<table>
<thead>
<tr>
<th>Locus B</th>
<th>Locus A – allele frequencies</th>
<th>Marginal prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr(B_1) = p_B$</td>
<td>$Pr(A_1) = p_A$</td>
<td>$Pr(A_2) = q_A$</td>
</tr>
<tr>
<td>$Pr(B_2) = q_B$</td>
<td>$Pr(A_1 B_2) = p_A q_B$</td>
<td>$Pr(A_2 B_1) = q_A p_B$</td>
</tr>
</tbody>
</table>

Marginal prob

- $p_A p_B + p_A q_B = p_A (p_B + q_B) = p_A$
- $q_A p_B + q_A q_B = q_A (p_B + q_B) = q_A$

**ii) What if the alleles at the two loci are NOT independent?**

- joint probabilities deviate from product of marginal probabilities (by $+D$)

<table>
<thead>
<tr>
<th>Locus B</th>
<th>Locus A – allele frequencies</th>
<th>Marginal prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr(B_1) = p_B$</td>
<td>$Pr(A_1 B_1) = r$</td>
<td>$Pr(A_2 B_1) = t$</td>
</tr>
<tr>
<td>$Pr(B_2) = q_B$</td>
<td>$Pr(A_1 B_2) = s$</td>
<td>$Pr(A_2 B_2) = u$</td>
</tr>
</tbody>
</table>

Marginal prob

- $p_A p_B + D + p_A q_B - D = p_A (p_B + q_B) = p_A$
- $q_A p_B - D + q_A q_B + D = q_A (p_B + q_B) = q_A$

If alleles are dependent ➔ loci are in **Linkage Disequilibrium**

or in **Gametic Phase Disequilibrium**

The term ‘linkage’ in Linkage disequilibrium is actually not quite correct and a bit misleading because disequilibrium can occur between unlinked loci, although it is more likely to be present (and persist) between linked loci (see later). Thus, ‘Gametic phase’ disequilibrium is a better term; gametic phase refers to the haploid phase of chromosomes and disequilibrium refers to dependence between alleles that make up the haplotypes that are present in the current generation and which originated from the haploid gametes produced by their parents.

$$D = \text{measure of disequilibrium} = D = r - p_A p_B$$

$$D = \frac{Pr(A_1 B_1) - Pr(A_1)Pr(B_1)}{Pr(A_1)Pr(B_1)}$$

The value obtained for $|D|$ is the same irrespective of the haplotype used.
You can also calculate $D$ as: 

\[
D = \frac{1}{2} \left[ \Pr \left( \frac{A_1 B_1}{A_2 B_2} \right) - \Pr \left( \frac{A_1 B_2}{A_2 B_1} \right) \right] = ru - st
\]

**Other measures of LD:**

$D' = D$ standardized to make it less dependent on allele frequencies

\[
D' = \frac{D}{D_{\text{max}}}
\]

where $D_{\text{max}} = \min(p_A p_B, q_A q_B)$ if $D < 0$

$D_{\text{max}} = \min(p_A q_B, q_A p_B)$ if $D > 0$

$r^2 = \text{squared correlation between allele at locus } A \text{ and allele at locus } B$

- also measures ability ($R^2$) to predict allele at locus $A$ from allele at locus $B$

To derive $r^2$: Let $X = 1$ when allele $A_1$ present, $X = 0$ if $A_2$ present (= Bernoulli var.)

$Y = 1$ when allele $B_1$ present, $Y = 0$ if $B_2$ present (= Bernoulli var.)

Then: \( \text{cov}(X, Y) = E(XY) - E(X) E(Y) \)

\[
= r - p_A p_B = D
\]

\[
\Rightarrow \text{Corr.} = r_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{D}{\sqrt{(p_A q_A)(p_B q_B)}}
\]

\[
\Rightarrow r^2 = r_{XY}^2 = \frac{D^2}{p_A q_A p_B q_B}
\]

(Note: this $r$ is different than $r$ in the table above)

$|D'|$ and $r^2$ range between 0 and 1

$|D'|$ is strongly inflated if one haplotype has a very low frequency

$r^2$ is the preferred measure of LD for most uses

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_1$</td>
<td>$Y = 1$</td>
<td>$\Pr(A_1 B_1) = r$</td>
</tr>
<tr>
<td>$XY = 1$</td>
<td>$XY = 0$</td>
<td></td>
</tr>
<tr>
<td>$B_2$</td>
<td>$Y = 0$</td>
<td>$\Pr(A_1 B_2) = s$</td>
</tr>
<tr>
<td>$XY = 0$</td>
<td>$XY = 0$</td>
<td></td>
</tr>
</tbody>
</table>
Mechanisms that generate Linkage Disequilibrium (LD)

A variety of mechanisms generate linkage disequilibrium, and several of these can operate simultaneously. They can be separated into:

1. **Recurrent factors** – operate to create LD each generation
   a. **Drift** (inbreeding) in small populations – by chance or sampling, haplotypes passed on to the next generation are not in LE frequencies
   b. **Recurrent migration** – continuous mixing of populations in which haplotypes occur in different frequencies (e.g. \( \Pr(A_1B_1) = 1 \) for pop. 1 and =0 for pop. 2)
   c. **Selection** – certain haplotypes may be selected upon and increase in frequency
      – selection creates LD between loci that are selected upon (= Bulmer effect)
      – selection with epistasis (certain combinations of alleles are favorable)
        also creates LD between loci involved.

2. **Punctual factors** – operate only sporadically over time to create LD
   a. **Mutation** – occurs in a specific haplotype, which is then the only haplotype that contains that mutation, resulting it to be in LD with the mutation.
   b. **One-time admixture/migration/crossing** (e.g. producing \( F_1/F_2 \)) – results in mixing populations with different haplotype frequencies
   c. **Population bottleneck / founder effects** – severe drift from 1-time sampling effects

---

![Processes that create LD](image1)

**LD is continuously eroded by recombination**

\[ \text{LD} = rac{1}{1+4N_e c} \]

- \( c \) = recombination rate
- meiosis
- non-recombinant
- recombinant

\[ f_{ij}(1-c), \quad f_{ij}(1-c), \quad f_{jc}, \quad f_{jc}, \quad f_{jc} \]
**LD is continuously eroded by recombination:** how does $D$ change over time?

Let $r_0 =$ frequency of $A_1B_1$ haplotypes in generation 0 $\rightarrow D_0 = r_0 - p_Ap_B$

What is the frequency of $A_1B_1$ haplotypes in generation 1?

In the following derivation, we will consider parental origin of haplotypes and will let $\bullet$ indicate ‘any’ allele, so $A_1B_1/\bullet\bullet$ indicates an individual that received the $A_1B_1$ from its father and any haplotype ($A_1B_1$ or $A_1B_2$ or $A_2B_1$ or $A_2B_2$) from its mother.

There are four ways that parents from generation 0 can generate gametes that carry the $A_1B_1$ haplotype and that will produce generation 1:

1. non-recombinant $A_1B_1$ haplotype produced by a $A_1B_1/\bullet\bullet$ parent
2. non-recombinant $A_1B_1$ haplotype produced by a $\bullet\bullet/A_1B_1$ parent
3. recombinant $A_1B_1$ haplotype produced by a $A_1B_1/A_1B_1$ parent
4. recombinant $A_1B_1$ haplotype produced by a $\bullet\bullet/A_1B_1$ parent

For case 1, the frequency of $A_1B_1/\bullet\bullet$ parents is $r_0$. The frequency of non-recombinant $A_1B_1$ haplotypes produced by these parents is $\frac{1}{2} (1-c)$. Thus, the frequency of $A_1B_1$ haplotype produced by $A_1B_1/\bullet\bullet$ parents = Prob(1.) = $\frac{1}{2}(1-c)r_0$.

Case 2 results in the same frequency: Prob(2.) = $\frac{1}{2}(1-c)r_0$.

For case 3, the frequency of $\bullet\bullet/A_1B_1$ parents is $p_Ap_B$. The frequency of recombinant $A_1B_1$ haplotypes produced by these parents is $\frac{1}{2}c$, so the overall frequency is $\frac{1}{2}cp_Ap_B$.

Case 4 results in the same frequency: Prob(4.) = $\frac{1}{2}cp_Ap_B$.

Thus, the overall frequency of $A_1B_1$ gametes produced by generation 0 is the sum of these four mutually exclusive cases:

- $r_1 = r_0(1-c) + p_Ap_Bc$
- $D_1 = r_1p_Ap_B = r_0(1-c) + p_Ap_Bc - p_Ap_B = r_0(1-c) - p_Ap_B(1-c) = (r_0 + p_Ap_B)(1-c) = D_0(1-c)$
- $D_2 = D_1(1-c) = D_0(1-c)^2$
- $D_t = D_0(1-c)^t \rightarrow D_\infty = 0$

Erosion of LD by recombination occurs faster when loci are further apart. LD is halved each generation if loci are unlinked ($c = \frac{1}{2}$).

Since $\mathbf{r}^2 = \frac{D^2}{p_Aq_Ap_Bq_B}$, LD measured by $r^2$ will decline at a rate of $(1-c)^2$ per generation:

$$r_t^2 = r_0^2(1-c)^{2t}$$
Balance between drift and recombination: in small(er) closed populations

- LD is continuously created by drift (sampling) (small effective population size, \(N_e\))
- LD is continuously eroded by recombination – faster at longer distances

This results in a balance/equilibrium of average LD at a given distance: \(E(r^2_{\infty,c}) = \frac{1}{1 + 4N_e c}\) (Sved 1971)

Most outbred domesticated plant and animal populations have small(er) (historical) effective population size and drift-recombination balance is expected to be the main contributor to LD ➔ LD expected to be sizeable at short distances, but small at longer distances.

Most human populations have large (historical) \(N_e\) ➔ \(E(r^2_{\infty,c}) = \frac{1}{1 + 4N_e c}\) is smaller at given distance.
Building Blocks of Quantitative Genetics

Jack Dekkers

1 locus case:
Genotype: \( A_2A_2 \quad A_1A_2 \quad A_1A_1 \)

Genotypic value: \( \mu-a \quad \mu \quad \mu+d \quad \mu+a \) \( \mu = \) mid-homozygote value

Input matrix for epistatic effects:

<table>
<thead>
<tr>
<th></th>
<th>( A_1A_1 )</th>
<th>( A_1A_2 )</th>
<th>( A_2A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_1B_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( B_1B_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( B_2B_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\( \mu = \) mid-homozygote value

\( \mu = \mu_1 + \mu_2 \)

Single-locus frequencies, genotypic values (deviated from \( \mu \)), and expectation

\[
\begin{array}{cccc}
T & \text{Frequency, } \text{Pr}(T) & \text{Genotypic value, } G_T & \text{Pr}(T) \times G_T \\
A_1A_1 & p^2 & a & p^2a \\
A_1A_2 & 2pq & d & 2pqd \\
A_2A_2 & q^2 & a & q^2a \\
\end{array}
\]

Population mean = \( \mu + E(G_T) = M = \mu + p^2a + 2pqd + -q^2a = \mu + a(p - q) + 2pqd \)

Extension to two loci (without epistasis):

The genotypic value of an individual is the sum of the genotypic values at each locus:

\( G_T = \mu + G_A + G_B \) \( G_i = \) genotypic value locus \( i \), as defined for 1-locus case

Now the homozygote “midpoint” \( \mu \) is midway between the best and worst double homozygote (\( A_1A_1B_1B_1 \) and \( A_2A_2B_2B_2 \)) = \((17+3)/2 = 10 \) in the example below.

Pop. mean = \( \mu + E(G_T) = M = \mu + E(G_A + G_B) = E(G_A) \quad + \quad E(G_B) \)

\( = \mu + \{a_A(p_A - q_A) + 2p_Aq_Ad_A\} + \{a_B(p_B - q_B) + 2p_Bq_Bd_B\} \)

Spreadsheet Genotypic_values_models_v10.xls

Input parameters:

Output:

### 2-locus genotypic values and frequencies (random mating)

<table>
<thead>
<tr>
<th>B locus genotype</th>
<th>( \mu + G_T )</th>
<th>( G_A )</th>
<th>( G_B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_1B_1 )</td>
<td>3</td>
<td>0.09</td>
<td>0.0324</td>
</tr>
<tr>
<td>( B_1B_2 )</td>
<td>-1</td>
<td>0.42</td>
<td>0.1512</td>
</tr>
<tr>
<td>( B_2B_2 )</td>
<td>-3</td>
<td>0.49</td>
<td>0.1764</td>
</tr>
</tbody>
</table>

Population mean \( M = 10.14 \)

New \( \mu = 10.14 \)

\( h_A = 8.38 \)

\( h_B = 11.76 \)

Re-calculated 1-locus additive, dominance, and genotypic values with epistasis

<table>
<thead>
<tr>
<th>( G_A )</th>
<th>( G_B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>-4</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>-3</td>
<td>-3</td>
</tr>
</tbody>
</table>
Extension to many loci: \( G_T = \Sigma G_i \) summation is over all loci

Homozygote “midpoint” \( \mu \) is average of the best and the worst multi-homozygote

Population mean \( = \mu + \mathbb{E}(G_T) = M \)

\[
M = \mu + \Sigma\{a_i(p_i - q_i) + 2p_i q_i d_i\} = \Sigma a_i(p_i - q_i) + 2\Sigma p_i q_i d_i
\]

**Allele-based models for additive effects**

In practice, we are interested in selecting the ‘best’ individuals to be used as parents to breed the next generation; we want to select individuals whose progeny have the highest expected phenotype, i.e. whose progeny have the highest expected genotypic value.

To identify these individuals we need to know how the genotypic value of progeny relates to the genotypic value of their parents.

Models described in terms of \( a \) and \( d \) are for genotypic values for whole genotypes. But individuals pass on alleles NOT genotypes

The breeding value of an individual is defined to quantify an individual’s value as a parent. It is related to the expected genotypic or phenotypic values of that individual’s progeny.

An individual’s **breeding value** = 2 x expected deviation of the mean phenotype of an individual’s progeny from population mean \( (M) \) when mated at random to other individuals from the population.

\[
A_i = 2\ E(P_{\text{progeny}}-M)
\]
In general: an individual’s breeding value is the sum of the average effects of the alleles that the individual carries: \( A_{ij} = \alpha_i + \alpha_j \)

**Average effect** \( \alpha_i \) = Average deviation from the population mean of individuals who received allele \( i \) (i.e. \( A_1 \) or \( A_2 \)) from one parent and the other allele at random (i.e. \( A_1 \) with freq. \( p \) and \( A_2 \) with freq. \( q \))

<table>
<thead>
<tr>
<th>“Allele ( i )”</th>
<th>“Other Allele” (=random) ( A_1 ) ( Pr(A_1) = p ) ( A_2 ) ( Pr(A_2) = q )</th>
<th>Mean ( G_T ) of resulting individuals</th>
<th>Mean of ( G_T ) deviated from Population mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( G_{A1A1} = a ) ( G_{A1A2} = d )</td>
<td>( pa + qd )</td>
<td>( \alpha_1 = pa + qd - M )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( =q[a + (q-p)d] = q\alpha )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( G_{A2A1} = d ) ( G_{A2A2} = -a )</td>
<td>( pd - qa )</td>
<td>( \alpha_2 = pd - qa - M )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( =-p[a + (q-p)d] = -p\alpha )</td>
</tr>
</tbody>
</table>

Another important concept/quantity is the average allele substitution effect (\( \alpha \)):

**Average allele substitution effect** = \( \alpha = \alpha_1 - \alpha_2 = a + (q - p)d \)

= average effect on the genotypic value of substituting a random \( A_2 \) allele for an \( A_1 \) allele

**One locus multiple alleles**: an individual’s breeding value is the sum of the average effects of the alleles that the individual carries: \( A_{ij} = \alpha_i + \alpha_j \)

**For two loci**: individual with alleles \( i \) and \( j \) at locus A and alleles \( k \) and \( l \) at locus B:
\( A_{ijkl} = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bk} + \alpha_{Bl} \) with each \( \alpha_{ni} \) derived as above for 1-locus case *(no epistasis)*

**Many loci**: \( A_{ijkl} = \sum_{locus l=1}^{n} \sum_{allele i=1}^{2} \alpha_{li} \) sum average effects over all \( n \) loci and the individual’s two alleles at each locus

**Spreadsheet ‘Genotypic_value_models.v10.xls’**

**Allele-based model for genotypic values**

<table>
<thead>
<tr>
<th>Locus A</th>
<th>Average allele effects</th>
<th>Substitution effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( 4.62 )</td>
<td>( -5.76 )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( 6.86 )</td>
<td>( 2.24 )</td>
</tr>
</tbody>
</table>

**A locus genotype**

<table>
<thead>
<tr>
<th>( A_1A_1 )</th>
<th>( A_1A_2 )</th>
<th>( A_2A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b + G )</td>
<td>( 0.24 )</td>
<td>( -0.78 )</td>
</tr>
<tr>
<td>Additive</td>
<td>( -0.64 )</td>
<td>( 0.96 )</td>
</tr>
<tr>
<td>Epistasis</td>
<td>( -- )</td>
<td>( -- )</td>
</tr>
</tbody>
</table>

**B locus genotype**

<table>
<thead>
<tr>
<th>( B_1B_1 )</th>
<th>( B_1B_2 )</th>
<th>( B_2B_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2.88 )</td>
<td>( 0.34 )</td>
<td>( -0.46 )</td>
</tr>
<tr>
<td>( 6.52 )</td>
<td>( 1.94 )</td>
<td>( -3.28 )</td>
</tr>
<tr>
<td>( 2.92 )</td>
<td>( -0.54 )</td>
<td>( -1.86 )</td>
</tr>
<tr>
<td>( -0.42 )</td>
<td>( 0.00 )</td>
<td>( 0.00 )</td>
</tr>
<tr>
<td>( -1.06 )</td>
<td>( 0.00 )</td>
<td>( 0.00 )</td>
</tr>
</tbody>
</table>

**Single locus genotypic and breeding values deviated from population mean (M)**

- \( A \) locus G
- B locus G
- A locus Br.val.
- B locus Br.val.
Alternate derivation of allele substitution effect based on
Linear regression on number of ‘1’ alleles

Allele substitution effects can also be derived by analyzing phenotype (or the genotypic value) by a linear regression model on the number of ‘1’ alleles that an individual carries (as we have done in some of the homeworks):

Linear regression of $Y$ on $X$: 
\[
Y = M + \hat{b}_Y (X - \bar{X}) + \epsilon
\]
\[
\hat{Y} = M + \hat{b}_Y (X - \bar{X})
\]

In this case, $X = \# 1$ alleles: $X_T = \{2, 1, 0\}$ for $T = \{A_1A_1, A_1A_2, A_2A_2\}$

<table>
<thead>
<tr>
<th>Genotype $T$</th>
<th>$G_T$</th>
<th># $A_1$</th>
<th>$X_T$</th>
<th>Frequency $f$</th>
<th>$f^2 X_T^2$</th>
<th>$F^2 G_T X_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>a</td>
<td>2</td>
<td>$p^2$</td>
<td>4$p^2$</td>
<td>2$p^2 a$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>d</td>
<td>1</td>
<td>2$p q$</td>
<td>2$p q$</td>
<td>2$p q d$</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>-a</td>
<td>0</td>
<td>$q^2$</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td></td>
<td></td>
<td>4$p^2 + 2pq$</td>
<td>2$p^2 a + 2pq d$</td>
<td></td>
</tr>
</tbody>
</table>

\[
\hat{b} = \frac{\text{cov}(G_T, X_T)}{\text{var}(X_T)} = \frac{E(G_T X_T) - E(G_T) E(X_T)}{E(X_T^2) - E(X_T)^2} = \frac{\left[2p^2 a + 2pq d\right] - \left[a(p - q) + 2pq d\right] \left[2p^2 + 2pq\right]}{\left[4p^2 + 2pq\right] - \left[2p^2 + 2pq\right]^2}
\]

Interpretation of regression coefficient $\hat{b}$:
- When $X_T$ increases by 1, $\hat{Y}$ increases by $a + (q - p)d = \alpha = \text{allele substitution effect}$
- When $X_T$ increases by 1, an allele substitution has occurred $\Rightarrow \hat{b} = \alpha$

**Allele-based models for dominance and epistatic effects**

When $d$ is not 0, breeding values will not explain everything about the genotypic value:

**Single locus** example: $p=0.6; a=+4; d=1; M=0$ $\Rightarrow \alpha = a + (q - p)d = +3.8$
- $\alpha_1 = q\alpha = 0.4*3.8 = +1.52$
- $\alpha_2 = -p\alpha = -0.6*3.8 = -2.28$

<table>
<thead>
<tr>
<th>Genotype $T$</th>
<th>Frequency</th>
<th>Genotypic value $(G)$ deviated from $M$</th>
<th>Breeding value $A$</th>
<th>Dominance deviation $\delta = G - A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>0.36</td>
<td>+2.72</td>
<td>$2\alpha_1 = +3.04$</td>
<td>-0.32</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>0.48</td>
<td>-0.28</td>
<td>$\alpha_1 + \alpha_2 = -0.76$</td>
<td>+0.48</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>0.16</td>
<td>-5.28</td>
<td>$2\alpha_2 = -4.56$</td>
<td>-0.72</td>
</tr>
</tbody>
</table>

These differences between the single-locus genotypic and breeding values are called ‘dominance deviations’.
Based on this, the genotypic value at a single locus of an individual that has alleles $i$ and $j$ at that locus (deviated from the population mean, $M$) can be written as:

$$G_{ij} = \alpha_i + \alpha_j + \delta_{ij} = A_{ij} + \delta_{ij}$$

$\alpha_i = \text{average effect of allele } i$  
$\alpha_j = \text{average effect of allele } j$  
$\delta_{ij} = \text{dominance deviation effect of the interaction of alleles } i \text{ and } j$

Dominance deviations can also be derived as a function of allele frequencies and $d$

<table>
<thead>
<tr>
<th>Genotype $T$</th>
<th>$G_T$ (deviated from $\mu$)</th>
<th>Average allele effects</th>
<th>Dom.dev. $\delta_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$a$</td>
<td>$\alpha_1 = q\alpha$</td>
<td>$\alpha_1 = q\alpha$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2 = p\alpha$</td>
<td>$\delta_{ij} = -2q^2d$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$d$</td>
<td>$\alpha_1 = q\alpha$</td>
<td>$\alpha_2 = -p\alpha$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$-a$</td>
<td>$\alpha_2 = -p\alpha$</td>
<td>$\delta_{ij} = -2p^2d$</td>
</tr>
</tbody>
</table>

Graphical representation of average effects and dominance deviations

Dominance deviations are the residuals from the regression of genotypic values on the number of $A_1$ alleles.

The regression line represents the breeding values.
**Extension to two loci (no epistasis)**

With two loci, the genotypic value is the sum of the individual’s genotypic value at each locus:

\[ G_T = G_A + G_B \]

\[ G_i = \text{genotypic value locus } i, \text{ as defined for 1-locus case} \]

And the genotypic value at each locus can be partitioned into additive and dominance effects:

\[ G_{Aij} = \alpha_{Ai} + \alpha_{Aj} + \delta_{Aij} \]

and

\[ G_{Bij} = \alpha_{Bi} + \alpha_{Bj} + \delta_{Bij} \]

Thus the overall genotypic value can be written as the sum of average allele effects and dominance deviations as:

\[ G_T = \alpha_{Ai} + \alpha_{Aj} + \delta_{Aij} + \alpha_{Bi} + \alpha_{Bj} + \delta_{Bij} \]

The sum of average allele effects define the **breeding value**: \( A_T = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bi} + \alpha_{Bj} \)

The sum of dominance deviations define the **dominance effect**: \( D_T = \delta_{Aij} + \delta_{Bij} \)

Thus the genotypic value can be written as the breeding value and its dominance effect:

\[ G_T = A_T + D_T \]

---

**Input matrix for epistatic effects**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>A1A1</th>
<th>A1A2</th>
<th>A2A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1B1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B1B2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B2B2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Allele-based model for genotypic values**

<table>
<thead>
<tr>
<th>Average allele effects</th>
<th>Substitution effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus A</td>
<td></td>
</tr>
<tr>
<td>A1 = 1.44</td>
<td>A2 = -2.16</td>
</tr>
<tr>
<td>Locus B</td>
<td></td>
</tr>
<tr>
<td>B1 = 1.82</td>
<td>B2 = -0.78</td>
</tr>
</tbody>
</table>

All values are now deviated from the population mean, M.

**Extension to multiple loci (no epistasis):**

\[ G_T = A_T + D_T \]

\[ A_T = \sum_{\text{locus } i=1}^{n} \sum_{\text{allele } i=1}^{2} \alpha_{li} \]

\[ D_T = \sum_{\text{locus } i=1}^{n} \delta_{lij} \]
**Epistatic Deviations**

When epistatic effects are present, the genotypic value of an individual cannot be written as a simple sum of the genotypic value at each locus but an effect of the interaction between loci needs to be added: For two loci:  
\[ G_T = G_A + G_B + G_{AxB} \]

Similarly, the genotypic value of an individual can also not be written as the sum of a breeding value and a dominance deviation but an epistatic deviation effect \( I_T \) needs to be added:  
\[ G_T = A_T + D_T + I_T \]

Epistatic deviation effects for each individual (or multi-locus genotype) can be calculated by subtraction, after the additive and dominance deviation effects have been computed as described before:  
\[ I_T = G_T - A_T - D_T \]
Genetic Variance Components

Single-locus model

<table>
<thead>
<tr>
<th>Genotype</th>
<th>$A_2A_2$</th>
<th>$A_1A_2$</th>
<th>$A_1A_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypic value</td>
<td>$-a$</td>
<td>$0$</td>
<td>$d$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T$</th>
<th>Frequency</th>
<th>Genotypic value</th>
<th>$Pr(T)$</th>
<th>$(G_T)^2$</th>
<th>$Pr(T)$</th>
<th>Breeding value</th>
<th>Dominance dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$p^2$</td>
<td>$a$</td>
<td>$p^2a^2$</td>
<td>$a^2$</td>
<td>$2q\alpha$</td>
<td>$-2q^2d$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$2pq$</td>
<td>$d$</td>
<td>$2pqd$</td>
<td>$2pqd^2$</td>
<td>$(q-p)\alpha$</td>
<td>$2pqd$</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$q^2$</td>
<td>$-a$</td>
<td>$-q^2a$</td>
<td>$q^2a^2$</td>
<td>$-2p\alpha$</td>
<td>$-2p^2d$</td>
<td></td>
</tr>
</tbody>
</table>

Genetic model for genotypic values: $G_T = A_T + D_T$ = Breeding value + Dominance dev.

Variance of genotypic values in a population = (Total) Genetic variance = $V_G$

$V_G = \text{var}(G_T) = p^2a^2 + 2pqd^2 + q^2a^2 - E(G_T)^2 = 2pq[a + (q-p)d]^2 + (2pqd)^2$

Using $\alpha = a + (q-p)d$ = allele substitution effect: $V_G = 2pq \alpha^2 + (2pqd)^2$

Additive genetic variance = variance of breeding values in a population = $V_A$

$V_A = \text{var}(A_T) = p^2(2q\alpha)^2 + 2pq[2(q-p)\alpha]^2 + q^2(-2p\alpha)^2 - 0^2 = 2pq\alpha^2$ (Note that $E(A_T)$=0)

Dominance variance = variance of Dominance deviations in a population = $V_D$

Using the table on p1, the variance of dominance deviations in the population is:

$V_D = \text{var}(D_T) = \text{var}(\delta_{ij}) = p^2(-2q^2d)^2 + 2pq(2pqd)^2 + q^2(-2p^2d)^2 - 0^2 = (2pqd)^2$ (Note that $E(D_T)$=0)

$\Rightarrow$ Genotypic variance = $V_G = 2pq\alpha^2 + (2pqd)^2 = V_A + V_D$

= Additive Variance + Dominance Variance

Note: $\text{cov}(A_T,D_T) = 0$; i.e. breeding values and dominance deviations are independent

Extension to two loci – first without epistasis:

Genotypic value = $G_T = G_A + G_B$ $G_i$ = genotypic value locus $i$

$V_G = \text{var}(G_T) = \text{var}(G_A + G_B) = \text{var}(G_A) + \text{var}(G_B) + 2\text{cov}(G_A,G_B)$

$= \text{var}(G_A) + \text{var}(G_B) + 0$ \hspace{1cm} \text{cov}=0 \text{ if loci are in LE}

$= 2p_Aq_A\alpha_A^2 + (2p_Aq_Ad_A)^2 + 2p_Bq_B\alpha_B^2 + (2p_Bq_Bd_B)^2$

$= \{2p_Aq_A\alpha_A^2 + 2p_Bq_B\alpha_B^2\} + \{(2p_Aq_Ad_A)^2 + (2p_Bq_Bd_B)^2\}$

$= \{ V_{A_A} + V_{A_B} \} + \{ V_{D_A} + V_{D_B} \}$

$= V_A + V_D$
Homozygote midpoint = 10

\[
a_A = 4, \quad a_B = 3
\]
\[
d_A = 2, \quad d_B = -1
\]
\[
p_A = 0.6, \quad p_B = 0.3
\]
\[
q_A = 0.4, \quad q_B = 0.7
\]

Linkage Disequilibrium D = 0
Recomb. Rate = 0.2

Input matrix for epistatic effects

<table>
<thead>
<tr>
<th></th>
<th>A_1A_1</th>
<th>A_1A_2</th>
<th>A_2A_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_1B_1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B_1B_2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B_2B_2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

ALLELE-BASED MODEL FOR GENOTYPIC VALUES

<table>
<thead>
<tr>
<th>Locus</th>
<th>Average allele effects</th>
<th>Substitution effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A_1 = 1.44</td>
<td>A_2 = -2.16</td>
</tr>
<tr>
<td></td>
<td>A = 3.6</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>B_1 = 1.82</td>
<td>B_2 = -0.78</td>
</tr>
<tr>
<td></td>
<td>B = 2.6</td>
<td></td>
</tr>
</tbody>
</table>

All values are now deviated from the population mean, M.

Population variances

<table>
<thead>
<tr>
<th></th>
<th>A locus</th>
<th>B locus</th>
<th>Population</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Genetic</td>
<td>7.142</td>
<td>3.016</td>
<td>10.158</td>
<td>100.0%</td>
</tr>
<tr>
<td>Additive effects</td>
<td>6.221</td>
<td>2.839</td>
<td>9.060</td>
<td>89.2%</td>
</tr>
<tr>
<td>Breeding values</td>
<td>6.221</td>
<td>2.839</td>
<td>9.06</td>
<td></td>
</tr>
<tr>
<td>Dominance</td>
<td>0.922</td>
<td>0.176</td>
<td>1.098</td>
<td>10.8%</td>
</tr>
<tr>
<td>Epistasis</td>
<td>--</td>
<td>--</td>
<td>0.000</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

“Additive effects” refer to breeding values computed as the sum of average allele effects.
“Breeding values” are computed based on the expected progeny means

Extended to >2 loci, this gives:

\[ V_G = \sum V_{Gi} = \sum \left\{ 2p_i q_i \alpha_i^2 + (2p_i q_i d_i)^2 \right\} = \sum V_{Ai} + \sum V_{Di} = V_A + V_D \]

with:

\[ V_A = \sum V_{Ai} = \sum 2p_i q_i \alpha_i^2 \quad \text{and} \quad V_D = \sum V_{Di} = \sum (2p_i q_i d_i)^2 \]

⇒ the genetic, additive, and dominance variances for a quantitative trait are the simple sum of the genetic, additive, and dominance variances at each locus that affect the trait.
With Epistatic effects = Interactions between the effects that loci have on phenotype

Two locus example: \[ G_T = G_A + G_B + G_{AB} \]

Genotype-based model

\[ G_T = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bi} + \alpha_{Bj} + \delta_{Ai} + \delta_{Bi} + I_{AB} \]

Allele-based model

\[ G_T = A + D + I \]

Epistatic variance = variance of epistatic deviations in a population = \( V_I = \text{var}(I_{AB}) \)

→ Complete partitioning of genetic variance: \( V_G = V_A + V_D + V_I \)

Note: all cov’s = 0

→ Epistatic variance can be obtained by difference: \( V_I = V_G - V_A - V_D \) see spreadsheet for ex.

### Population matrix for epistatic effects

<table>
<thead>
<tr>
<th>B locus</th>
<th>A locus</th>
<th>( A_1A_1 )</th>
<th>( A_1A_2 )</th>
<th>( A_2A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_B_</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B_B_</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B_B_</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Population variances

<table>
<thead>
<tr>
<th>Population variances</th>
<th>A locus</th>
<th>B locus</th>
<th>Population</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Genetic</td>
<td>7.595</td>
<td>4.009</td>
<td>11.774</td>
<td>100.0%</td>
</tr>
<tr>
<td>Additive effects</td>
<td>6.793</td>
<td>3.591</td>
<td>10.384</td>
<td>88.2%</td>
</tr>
<tr>
<td>Breeding values</td>
<td>6.793</td>
<td>3.591</td>
<td>10.41367</td>
<td>88.4%</td>
</tr>
<tr>
<td>Dominance</td>
<td>0.004</td>
<td>0.413</td>
<td>1.220</td>
<td>10.8%</td>
</tr>
<tr>
<td>Epistasis</td>
<td>--</td>
<td>--</td>
<td>0.110</td>
<td>7.2%</td>
</tr>
</tbody>
</table>

### Population means

<table>
<thead>
<tr>
<th>Genotypic values</th>
<th>A locus</th>
<th>B locus</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Additive, dominance, and epistatic effects are independent (no covariances)

In a typical population, most genetic variance is additive — see also Hill et al. PLOS Genetics (2008)
Impact of Linkage disequilibrium on genetic variances

Consider 2 linked loci, A and B:
Each individual has a paternal and a maternal gamete

The breeding value is the sum of the average effects of the paternal and maternal alleles:

\[ A = \alpha_A^p + \alpha_B^p + \alpha_A^m + \alpha_B^m \]

If the loci are in LD \( \Rightarrow \) allele states (0/1) at two loci on the same gamete are not independent
\( \Rightarrow \) they have a non-zero covariance \( (r^2 = \text{squared correl. of allele states (0/1)} > 0) \)

Then, the variance caused by the additive effect of the paternal (or maternal) gamete is

\[ \text{var}(\alpha_A^p + \alpha_B^p) = \text{var}(\alpha_A^p) + \text{var}(\alpha_B^p) + 2\text{cov}(\alpha_A^p, \alpha_B^p) \]

\[ = \frac{1}{2}V_A^A + \frac{1}{2}V_A^B + 2D_{AB}\alpha_A^p\alpha_B^p \]

\text{Additive genetic variance} = \text{var}(\alpha_A^p + \alpha_B^p)

\text{Dominance genetic variance} = \text{var}(\delta_{Ap}^p + \delta_{Bp}^p)

\[ = V_{D_A} + V_{D_B} + 8D_{AB}\alpha_A^p\alpha_B^p \]
\( (D^2 \text{ because dominance is based on combinations of paternal and maternal alleles}) \)

<table>
<thead>
<tr>
<th>Homozygote midpoint</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_A = 4 )</td>
<td>( a_B = 3 )</td>
</tr>
<tr>
<td>( d_A = 2 )</td>
<td>( d_B = -1 )</td>
</tr>
<tr>
<td>( p_A = 0.6 )</td>
<td>( p_B = 0.3 )</td>
</tr>
<tr>
<td>( q_A = 0.4 )</td>
<td>( q_B = 0.7 )</td>
</tr>
</tbody>
</table>

| Linkage Disequilibrium \( D \) | 0 |
| Recomb. Rate \( = \) | 0.2 |

Note that individual locus variances are not affected by LD but across locus variances are
(= because of non-zero covariances).

Whether \( V_A \) increases or decreases depends on whether the favorable alleles are in repulsion or coupling phase
Whether \( V_D \) increases or decreases depends on whether \( d \) has the same sign for both loci.
Phenotypic and Environmental Effects and Variances

Models for Environmental Effects

Phenotype for a quantitative trait is determined by genetic and environmental factors:

\[ P = \mu + G + E \]

\( \mu \) includes the mean and \textit{systematic (environmental) effects}

\( G \) = genotypic value

\( E \) = \textit{Random environmental effects}

Partitioning of phenotypic variance

Phenotypic variance = var. of phenotypes in a pop. \textit{after removal/adjustment for syst. effects}

\[ V_P = \text{var}(P-\mu) = \text{var}(G+E) = \text{var}(G) + \text{var}(E) + 2\text{cov}(G,E) \]

If genotypes are distributed at random relative to random environmental effects \( \Rightarrow \text{cov}(G,E) = 0 \)

\[ \Rightarrow V_P = V_G + V_E = V_A + V_D + V_I + V_E \]

Relative importance of the genetic component

Genetic variance as a fraction of the phenotypic variance:

\textit{Broad sense heritability} \[ H^2 = \frac{V_G}{V_P} \] = proportion of phen. var. in a pop. that is \textit{genetic}

\textit{Narrow sense heritability} \[ h^2 = \frac{V_A}{V_P} \] = proportion of phen. var. that is \textit{additive genetic}
MODELS FOR TRAITS WITH REPEATED MEASURES

**General** environmental effects = \( Eg \) = Effects that are common to each measurement

**Special** environmental effects = \( Es \) = Effects that are specific to a given measurement

\[ P_{ij} = G_i + Eg_i + Es_{ij} \]

\( P_{ij} \) = \( j^{th} \) measurement of phenotype on \( i^{th} \) individual

\( G_i \) and \( Eg_i \) are common to all measurements on individual \( i \)

\( Es_{ij} \) = special envir. effect for \( j^{th} \) measurement on \( i^{th} \) individual

This also allows random environmental variance to be separated into variances due to General versus Special environmental effects:

\[ V_P = V_G + V_{Eg} + V_{Es} \]

Note: \( \text{cov}(Eg,Es) = 0 \)

**Repeatability** = \( r \) = correlation between repeated measures on the same individual

(Assume that \( V_{Es} \) and therefore \( V_P \) is the same for each measurement)

\[
    r = \frac{\text{cov}(P_{ij}, P_{ik})}{\sqrt{\text{var}(P_{ij})\text{var}(P_{ik})}} = \frac{\text{cov}(P_{ij}, P_{ik})}{\sqrt{V_P V_P}} = \frac{\text{cov}(P_{ij}, P_{ik})}{V_P}
\]

\[
    \text{cov}(P_{ij}, P_{ik}) = \text{cov}(G_i + Eg_i + Es_{ij}, G_i + Eg_i + Es_{ik})
\]

\[= \text{cov}(G_i,G_i) + \text{cov}(Eg_i,Eg_i) + \text{Cov}(Es_{ij},Es_{ik}) \]

\[= V_G + V_{Eg} + 0 \quad \text{special env. effects are independent} \]

\( \Rightarrow \) repeatability = \( r = (V_G + V_{Eg}) / V_P \) = prop. of \( V_P \) that is due to effects that are consistent across measurements (\( G + Eg \))

\( \Rightarrow \) \( 1 - r = V_{Es} / V_P \) = prop. of \( V_P \) that is due to effects that differ between measurements (\( Es \))
CORRELATED TRAITS

Phenotypic correlation = correlation between phenotypes on traits 1 and 2 on same individual
  = caused by genetics and environment

Phenotype trait 1

\[ P_1 = A_1 + D_1 + I_1 + E_1 \]

(Additive) genetic correlation = \( r_A = \text{Corr}(A_1, A_2) \)

Environmental correlation

\[ r_E = \frac{\text{Cov}(E_1, E_2)}{\sigma_{E_1} \sigma_{E_2}} \]

Phenotype trait 2

\[ P_2 = A_2 + D_2 + I_2 + E_2 \]

\[ r_p = \frac{\text{Cov}(P_1, P_2)}{\sigma_{P_1} \sigma_{P_2}} \]

\[ r_A = \frac{\text{Cov}(A_1, A_2)}{\sigma_{A_1} \sigma_{A_2}} \]

\[ r_E = \frac{\text{Cov}(E_1, E_2)}{\sigma_{E_1} \sigma_{E_2}} \]

Genetic correlation – caused by - pleiotropic genes = genes with effect on both traits
  - linkage – a gene that affects trait 1 is in LD with a gene that affects trait 2
    → transient correlation – disappears with loss of LD
  - quantifies the overall effect on both traits, across all loci
    → \( r_A = 0 \) does not imply that there are no pleiotropic genes

Environmental correlation – caused by random environmental factors that affect both traits
  – measures the overall effect of all environmental factors

Some quantitative genetic math to show relationships among correlations:

\[ \text{Cov}(P_1, P_2) = \text{Cov}(A_1 + E_1, A_2 + E_2) = \text{Cov}(A_1, A_2) + \text{Cov}(E_1, E_2) \]

\[ r_p \sigma_{P_1} \sigma_{P_2} = r_A \sigma_{A_1} \sigma_{A_2} + r_E \sigma_{E_1} \sigma_{E_2} \]

\[ r_p \sigma_{P_1} \sigma_{P_2} = r_A h_1 \sigma_{A_1} h_2 \sigma_{A_2} + r_E \sigma_{e_1} \sigma_{e_2} \]

\[ e^2 = 1 - h^2 = \text{prop.of phen.var. that is not add.genetic} \]

\[ r_p = r_A h_1 h_2 + r_E e_1 e_2 \]
GENETIC RELATIONSHIPS AND INBREEDING

Are two alleles the same? **Identity By State (IBS) versus Identity By Descent (IBD)**

- **IBS**: if we can genotype individuals o and o’ for this locus (QTL), then we can directly determine whether the alleles the two individuals carry are indeed the same – if they are the same, this is referred to as the alleles being **IBS**.

- **IBD**: if we cannot genotype the locus (i.e. the usual case), then we cannot determine IBS directly but, if o and o’ have a common ancestor, then we can determine the probability that the two alleles are identical because they may have originated from a common ancestor.

**IBD probabilities from pedigree:**

\[
\text{Prob(op is IBD to o’p)} = P(op = o’p) = \text{probability that alleles op and o’p originated from the same allele of the common ancestor}
\]

**Example IBD probabilities, coefficients of coancestry and additive and dominance coefficients**

<table>
<thead>
<tr>
<th>Individual o – o’</th>
<th>IBD probabilities for pairs of alleles</th>
<th>Coancestry coefficient</th>
<th>Additive relationship coefficient</th>
<th>Dominance relationship coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sire(o) – Offspring(o’)</td>
<td>op–o’p</td>
<td>½</td>
<td>½</td>
<td>½</td>
</tr>
<tr>
<td>Dam – Offspring</td>
<td>om–o’m</td>
<td>0</td>
<td>½</td>
<td>½</td>
</tr>
<tr>
<td>Paternal half-sibs</td>
<td>op–o’m</td>
<td>½</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Full sibs</td>
<td>om–o’p</td>
<td>½</td>
<td>½</td>
<td>0</td>
</tr>
<tr>
<td>Identical twins</td>
<td>op–o’p</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Coefficient of coancestry** (also coeff. of kinship or consanguinity) between o and o’ (See also Ch 5 p85)

\[ f_{oo'} = \text{probability that an allele drawn at random from o is IBD to an allele drawn random from o’} \]

\[ r_{oo'} = 2f_{oo'} = \text{coefficient of relationship = additive genetic relationship coefficient} \]

**NOTE**: \( f_{oo'} \) is also equal to the coefficient of inbreeding of a progeny produced by o and o’

= probability that an individual’s alleles are IBD

**From IBD probabilities to covariances of between relatives:**

\[
\text{Cov}(G_o, G_{o'}) = r_{oo'} V_A + u_{oo'} V_D
\]

This equation applies to each locus that affects the trait but also to total genetic value; summing variances over loci \( V_A = \Sigma V_{Al} \), this equation also applies to multiple loci.
Thus, the genetic covariance (resemblance) between relatives is a function of their genetic relationships and genetic variance components.
- the additive genetic cov. between relatives = genetic relationship x add. genetic var. = \( r_{oo'} V_A \)
- relatives ‘share’ a portion \( r_{oo'} \) of their additive genetic variance because they share a portion \( r_{oo'} \) of their alleles
- the dominance genetic cov. betw. relatives = dom. relationship x dom. genetic var. = \( u_{oo'} V_D \)

**Alternative derivation of additive covariances** based on quantitative genetics algebra

**Model of phenotype:**

\[ P = A + E \]

\( E \) includes D, I, environment

**Offspring phenotype:**

\[ P_o = A_o + E_o = \frac{1}{2} A_s + \frac{1}{2} A_d + R A_s + R A_d + E_o \]

\( \frac{1}{2} \) * breeding value of parents

- Breeding value = 2*E(P_o-M)  \( \text{(by definition)} \)
- Includes some dominance and epistatic effects

\( R A_s, R A_d \) = random assortment / Mendelian sampling terms
- sampling of 1 of 2 parent alleles at each locus during meiosis
- by definition independent from other terms: Cov(A_s, RA_s) = 0

**Without inbreeding:**

\[ \text{Var}(RA_s) = \frac{1}{4} V_A \]

\[ \text{Var}(RA_d) = \frac{1}{4} V_A \]

(see derivation below)

**With inbreeding:**

\[ \text{Var}(RA_s) = \frac{1}{4}(1-F_s)V_A \]

\[ \text{Var}(RA_d) = \frac{1}{4}(1-F_d)V_A \]

Thus:

\[ \text{Var}(A_o) = \text{Var}(\frac{1}{2} A_s + \frac{1}{2} A_d + RA_s + RA_d) = \]

\[ \frac{1}{4} V_A + \frac{1}{4} V_A + \frac{1}{4} V_A + \frac{1}{4} V_A = V_A \text{ (no inbreeding or selection)} \]

**Single locus derivation of Var(RA)**

<table>
<thead>
<tr>
<th>Parent Genotype</th>
<th>Frequency</th>
<th>Genotypic value of parent ( [\alpha=a+(q-p)d] )</th>
<th>Offspring mean phenotype ( \alpha' ) *breeding value parent</th>
<th>Transmitted allele</th>
<th>Frequency</th>
<th>Offspring mean phenotype</th>
<th>Mendelian sampling term (RA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1A1</td>
<td>( p^2 )</td>
<td>a ( 2q(\alpha-qd) )</td>
<td>( q \alpha )</td>
<td>A1</td>
<td>1</td>
<td>( q \alpha )</td>
<td>0</td>
</tr>
<tr>
<td>A1A2</td>
<td>2pq</td>
<td>d ( (q-p)\alpha+2qd )</td>
<td>( \frac{1}{2}(q-p)\alpha )</td>
<td>( A1 )</td>
<td>( \frac{1}{2} )</td>
<td>( q \alpha )</td>
<td>( \frac{1}{2} \alpha )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( A2 )</td>
<td>( \frac{1}{2} )</td>
<td>( -p \alpha )</td>
<td>( -\frac{1}{2} \alpha )</td>
</tr>
<tr>
<td>A2A2</td>
<td>( q^2 )</td>
<td>-a (-2p(\alpha+pd) )</td>
<td>( -p \alpha )</td>
<td>A2</td>
<td>1</td>
<td>( -p \alpha )</td>
<td>0</td>
</tr>
</tbody>
</table>

\( E(RA_s) = p^2(0) + 2pq\frac{1}{2}(\frac{1}{2} \alpha) + 2pq\frac{1}{2}(-\frac{1}{2} \alpha) + q^2(0) = 0 \)

**Without inbreeding:**

\[ \text{Var}(RA_s) = p^2(0)^2 + 2pq\frac{1}{2}(\frac{1}{2} \alpha)^2 + 2pq\frac{1}{2}(-\frac{1}{2} \alpha)^2 + q^2(0)^2 \]

\[ = \frac{1}{2}pq \alpha^2 = \frac{1}{4} V_A \]

\( V_A = 2pq \alpha^2 \)

**With inbreeding:**

\( F_s = \text{Pr(two alleles in s are ibd)} \Rightarrow RA_s = 0 \)

\( 1-F_s = \text{Pr(two alleles in s not ibd)} \Rightarrow RA_s = \text{as in Table above} \)

\[ \text{Var}(RA_s) = F_s(0) + (1-F_s)\frac{1}{4} V_A = \frac{1}{4}(1-F_s)V_A \]