Quantitative Genetic Models
Least Squares Genetic Model

- partition of effects

$$P = G + E + G \times E$$

P is phenotypic effect
G is genetic effect
E is environmental effect
$G \times E$ is interaction effect

- partition of variance components

$$Var(P) = Var(G) + 2Cov(G,E) + Var(E)$$

ignoring $G \times E$

$$Var(P) = Var(G) + Var(E)$$

further ignoring $Cov(G,E)$

Hardy-Weinberg (1908) Principle

change of allele & genotype frequency
over generations

- "ideal" assumptions:
  1 large (infinite) population
  2 random mating
  (no inbreeding or random drift)
  3 no natural or artificial selection
  4 no mutation
  5 closed to migration
  6 discrete generations

- sexual reproduction $\Rightarrow$ 1:1 gamete ratio
  7 separate sexes

Sexual Reproduction to H-W Proportions

genotype	female	male
B	p_{1f}	p_{1m}
b	p_{2f}	p_{2m}

genotype	BB	Bb	bb
frequency	$p_{1f}p_{1m}$	$p_{1f}p_{2m} + p_{2f}p_{1m}$	$p_{2f}p_{2m}$

autosomal locus: same for male, female

$$
\begin{aligned}
p_1 &= p_{BB} + p_{Bb}/2 = p_{1f}p_{1m} + [p_{1f}p_{2m} + p_{2f}p_{1m}]/2 \\
&= [p_{1f}(p_{1m} + p_{2m}) + p_{1m}(p_{1f} + p_{2f})]/2 \\
&= [p_{1f} + p_{1m}]/2
\end{aligned}
$$

next and subsequent generations

genotype	BB	Bb	bb
frequency	p_1^2	$2p_1p_2$	p_2^2

Hardy-Weinberg Equilibrium

- only 1 generation to remove sex differences

- only 2 generations to stabilize proportions

- random mating can stabilize proportions
  subjected to selection
  in 1-2 generations

Hardy-Weinberg Disequilibrium

	B	b	
B	$p_B^2 + D_B$	$p_B p_b - D_B$	$p_B = p_1$
b	$p_b p_B - D_B$	$p_b^2 + D_B$	$p_b = p_2$
	$p_B = p_1$	$p_b = p_1$	1

Test of Hardy-Weinberg Disequilibrium

- chi-square test
  $X^2 = \sum (\text{obs-exp})^2/\text{exp}$

$$X^2 = D_B^2 \left[\frac{N_{BB}}{p_B^2} + \frac{2N_{Bb}}{p_B p_b} + \frac{N_{bb}}{p_b^2} \right]$$

- log likelihood test
  $G = -2 \sum \text{obs} \times \log(\text{exp/obs})$

$$
\begin{aligned}
G = {} & 2[N_{BB} \log\left(1 + \frac{D_B^2}{p_B^2}\right) + N_{Bb} \log\left(\frac{D_B^2}{p_B p_b}\right) \\
& + N_{bb} \log\left(1 - \frac{D_B^2}{p_b^2}\right)]
\end{aligned}
$$

Hardy-Weinberg Equilibrium

gametes	B	b	total
B	p^2	$p(1-p)$	$p_1 = p$
b	$(1-p)p$	$(1-p)^2$	$p_2 = 1 - p$
total	$p_1 = p$	$p_2 = 1 - p$	1

genotype	BB	Bb	bb
frequency	$f_{BB} = (p_1)^2$	$f_{Bb} = 2p_1 p_2$	$f_{bb} = (p_2)^2$

Genetic Variance Components for a Locus

Genotype	value	common	Fisher
BB	G_{11}	$m + a$	$\mu + 2a_1 + d_{11}$
Bb	$G_{12} = G_{21}$	$m + d$	$\mu + a_1 + a_2 + d_{12}$
bb	G_{22}	$m - a$	$\mu + 2a_2 + d_{22}$

$$\mu = \sum_g f_g G_g = \sum_{i,j} p_i p_j G_{ij} = m + (p_1 - p_2)a + 2p_1 p_2 d$$

Additive and Dominance Variance Components Mather-Jinks Partition

$$G_{11} = m + a, \; G_{12} = m + d, \; G_{22} = m - a$$

$$
\begin{aligned}
\sigma_G^2 &= \sum_g f_g (G_g - \mu)^2 = \sum_{i,j} p_i p_j G_{ij}^2 - \mu^2 \\
&= p_1^2 a^2 + 2p_1 p_2 d^2 + p_2^2 (-a)^2 - (\mu - m)^2 \\
&= 2p_1 p_2 [a + (p_2 - p_1)d]^2 + 4p_1^2 p_2^2 d^2 \\
&= \sigma_a^2 + \sigma_d^2
\end{aligned}
$$

additive variance component
$$\sigma_a^2 = 2p_1 p_2 [a + (p_2 - p_1)d]^2$$

dominant variance component $\sigma_d^2 = 4p_1^2 p_2^2 d^2$

Fisher's Least Square Partition of Genetic Variance Components

$$
\begin{aligned}
G_{ij} &= \mu + a_i + a_j + d_{ij} \\
&= \mu + a_1 N_1(ij) + a_2 N_2(ij) + d_{ij}
\end{aligned}
$$

$N_1, N_2 = $ number copies of allele (1=B, 2=b)

parameters

$$
\begin{aligned}
\mu &= \bar{G}_{..} = \sum_{i,j} p_i p_j G_{ij} \\
a_i &= \sum_j p_j G_{ij} - \mu = \bar{G}_{i.} - \bar{G}_{..} \\
a_j &= \bar{G}_{.j} - \bar{G}_{..} \\
d_{ij} &= G_{ij} - a_i - a_j - \mu = G_{ij} - \bar{G}_{i.} - \bar{G}_{.j} + \bar{G}_{..}
\end{aligned}
$$

which satisfy the conditions

$$\sum_i p_i a_i = \sum_{i,j} p_i p_j d_{ij} = \sum_i p_i d_{ij} = \sum_j p_j d_{ij} = 0$$

LS Reference Mean and Additive Effects

$$\begin{aligned}
\mu &= p_1^2 G_{11} + 2p_1 p_2 G_{12} + p_2^2 G_{22} \\
a_1 &= p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})] \\
a_2 &= -p_1[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]
\end{aligned}$$

relation to Mather-Jinks

$$\begin{aligned}
\mu &= m + a(p_1^2 - p_2^2) + 2dp_1 p_2 \\
a_1 &= p_2[a + d(p_2 - p_1)] \\
a_2 &= -p_1[a + d(p_2 - p_1)] \\
a_1 - a_2 &= a + d(p_2 - p_1)
\end{aligned}$$

LS Dominance Effects

$$\begin{aligned}
d_{11} &= G_{11} - \mu - 2a_1 = -p_2^2[2G_{12} - G_{11} - G_{22}] \\
d_{12} &= d_{21} = G_{12} - \mu - a_1 - a_2 \\
&= p_1 p_2[2G_{12} - G_{11} - G_{22}] \\
d_{22} &= G_{22} - \mu - 2a_2 = -p_1^2[2G_{12} - G_{11} - G_{22}]
\end{aligned}$$

LS genetic variance components

$$\begin{aligned}
\sigma_G^2 &= \sum_{i,j} p_i p_j (G_{ij} - \mu)^2 = \sum_{i,j} p_i p_j (a_i + a_j + d_{ij})^2 \\
&= 2\sum_i p_i a_i^2 + \sum_i p_i p_j d_{ij}^2 \\
&= \sigma_a^2 + \sigma_d^2
\end{aligned}$$

Additive Variance

$$\begin{aligned}
\sigma_a^2 &= 2\sum_i p_i a_i^2 = 2[p_1 a_1^2 + p_2 a_2^2] \\
&= 2p_1 p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]^2 \\
&= 2p_1 p_2[p_1(a - d) + p_2(a + d)]^2 \\
&= 2p_1 p_2[a + (p_2 - p_1)d]^2
\end{aligned}$$

Dominance Variance

$$\begin{aligned}
\sigma_d^2 &= \sum_{i,j} p_i p_j d_{ij}^2 = p_1^2 d_{11}^2 + 2p_1 p_2 d_{12}^2 + p_2^2 d_{22}^2 \\
&= p_1^2 p_2^2[2G_{12} - G_{11} - G_{22}]^2 \\
&= 4p_1^2 p_2^2 d^2
\end{aligned}$$

Hardy-Weinberg Disequilibrium

genotype	BB	Bb	bb
frequency	P_{BB}	P_{Bb}	P_{bb}
H-W equilibrium	p_B^2	$2p_B p_b$	p_b^2

with $p_B = P_{BB} + \frac{1}{2}P_{Bb}$ and $p_b = P_{bb} + \frac{1}{2}P_{Bb}$

disequilibrium coefficients

$$\begin{aligned}
P_{BB} &= p_B^2 + D_{BB} \\
P_{Bb} &= 2p_B p_b - 2D_{Bb} \\
P_{bb} &= p_b^2 + D_{bb}
\end{aligned}$$

constraints of gene and genotype frequencies

$$\begin{aligned}
P_{BB} + P_{Bb} + P_{bb} = 1 &\Rightarrow \sum D_{ij} = 0 \\
p_{BB} + p_{Bb}/2 = p_B &\Rightarrow D_{BB} = D_{Bb} = D_{bb}
\end{aligned}$$

D_B measures H-W disequilibrium for locus

$$\begin{aligned}
P_{BB} &= p_B^2 + D_B \\
P_{Bb} &= 2p_B p_b - 2D_B \\
P_{bb} &= p_b^2 + D_B
\end{aligned}$$

Linkage disequilibrium

- Let us consider the distribution of gametes for a two-allele and two-locus model:

Gamete	AB	Ab	aB	ab
Frequency	P_{AB}	P_{Ab}	P_{aB}	P_{ab}
L.-E.	$p_A p_B$	$p_A p_b$	$p_a p_B$	$p_a p_b$

Define disequilibrium coefficients as:

$$\begin{aligned}
P_{AB} &= p_A p_B + D_{AB} \\
P_{Ab} &= p_A p_b + D_{Ab} \\
P_{aB} &= p_a p_B + D_{aB} \\
P_{ab} &= p_a p_b + D_{ab}
\end{aligned}$$

constraints of gene and genotype frequencies

$$D_{AB} + D_{Ab} + D_{aB} + D_{ab} = 0$$

$$p_A = P_{AB} + P_{Ab} = p_A + D_{AB} + D_{Ab}$$

$$\implies D_{AB} = -D_{Ab} \text{ and } D_{ab} = -D_{aB},$$

In general,

$$\sum_i D_{ij} = \sum_j D_{ij} = 0.$$

Thus

$$\begin{aligned}
P_{AB} &= p_A p_B + D_{AB} \\
P_{Ab} &= p_A p_b - D_{AB} \\
P_{aB} &= p_a p_B - D_{AB} \\
P_{ab} &= p_a p_b + D_{AB}
\end{aligned}$$

- Correlation coefficient between two loci:

$$\gamma = \frac{D_{AB}}{\sqrt{p_A p_B p_a p_b}}$$

- Relationship between linkage disequilibrium and recombination frequency: For both backcross and F_2 populations, segregating gametes are produced from F_1

- The linkage disequilibrium in these populations are determined by the gametes of F_1. F_1 produces four gametes in the following proportion:

Gamete	AB	Ab	aB	ab
Frequency	$\frac{1-r}{2}$	$\frac{r}{2}$	$\frac{r}{2}$	$\frac{1-r}{2}$

r = recombination frequency ($0 < r < 0.5$).

$p_A = p_B = p_a = p_b = \frac{1}{2}$
$D_{AB} = P_{AB} - p_A p_B = \frac{1-r}{2} - \frac{1}{4} = \frac{1}{4}(1 - 2r)$
thus

$$\gamma = \frac{D_{AB}}{\sqrt{p_A p_B p_a p_b}} = \frac{(1-2r)/4}{\sqrt{1/16}} = 1 - 2r$$

Linkage disequilibrium

	B	b	
A	$p_A p_B + D_{AB}$	$p_A p_b - D_{AB}$	p_A
a	$p_a p_B - D_{AB}$	$p_a p_b + D_{AB}$	$p_a = 1 - p_A$
	p_B	$p_b = 1 - p_B$	1

correlation between loci

$$X = 1(A), 0(a); \quad Y = 1(B), 0(b)$$

$$E(X) = p_A \times 1 + p_a \times 0 = p_A$$

$$\begin{aligned}
Var(X) &= p_A \times (1 - p_A)^2 + p_a \times (0 - p_A)^2 \\
&= p_A p_a
\end{aligned}$$

$$\begin{aligned}
Cov(X,Y) &= \sum_{x,y} P_{xy} \times (x - E(X))(y - E(Y)) \\
&= (p_A p_B + D_{AB}) \times (1 - p_A)(1 - p_B) + \\
&\quad (p_A p_b - D_{AB}) \times (1 - p_A)(0 - p_B) + \cdots \\
&= D_{AB}
\end{aligned}$$

$$\begin{aligned}
\gamma &= Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \\
&= \frac{D_{AB}}{\sqrt{p_A p_a p_B p_b}}
\end{aligned}$$

Linkage disequilibrium and partition of genetic variance

Consider two loci, no epistasis
locus A: i, j; locus B: k, l

least square genetic model

$$G_{jl}^{ik} = \mu + a^i + a^k + a_j + a_l + d_j^i + d_l^k$$

ik = gamete from parent 1 (superscript)
jl = gamete from parent 2 (subscript)

genotype frequency

P_{jl}^{ik} = frequency of G_{jl}^{ik} in population
Hardy-Weinberg equilibrium: $P_{jl}^{ik} = P_{ik}P_{jl}$
linkage disequilibrium: $P_{ik} = p_i p_k + D_{ik}$
constraints

$$\sum_i p_i a_i = \sum_i p_i d_j^i = \sum_j p_j d_j^i = \sum_i D_{ik} = \sum_k D_{ik} = 0$$

mean of the population

effects

$$G_{jl}^{ik} = \mu + a^i + a^k + a_j + a_l + d_j^i + d_l^k$$

constraints

$$\sum_i p_i a_i = \sum_i p_i d_j^i = \sum_j p_j d_j^i = \sum_i D_{ik} = \sum_k D_{ik} = 0$$

$$\begin{aligned}
\bar{G}_{..} &= \sum_{i,k}\sum_{j,l} P_{jl}^{ik} G_{jl}^{ik} = \sum_{i,k}\sum_{j,l} P_{ik} P_{jl} G_{jl}^{ik} \\
&= \sum_{i,k}\sum_{j,l} (p_i p_k + D_{ik})(p_j p_l + D_{jl}) G_{jl}^{ik} \\
&= \sum_{i,k}\sum_{j,l} p_i p_k p_j p_l G_{jl}^{ik} = \mu = \bar{G}_{..}
\end{aligned}$$

linkage disequilibrium does not affect mean
(not true with epistasis)

variance of the population

$$\begin{aligned}
\sigma_G^2 &= \sum_{i,k}\sum_{j,l} P_{ik} P_{jl} (G_{jl}^{ik} - \mu)^2 \\
&= \sum_{i,k}\sum_{j,l} (p_i p_k + D_{ik})(p_j p_l + D_{jl})(G_{jl}^{ik} - \mu)^2 \\
&= \sum_i p_i (a^i)^2 + \sum_k p_k (a^k)^2 + \sum_j p_j a_j^2 + \sum_l p_l a_l^2 \\
&\quad + 2\sum_{i,k} D_{ik} a^i a^k + 2\sum_{j,l} D_{jl} a_j a_l \\
&\quad + \sum_{i,j} p_i p_j d_j^{i2} + \sum_{k,l} p_k p_l d_l^{k2} + 2\sum_{i,k}\sum_{j,l} D_{ik} D_{jl} d_j^i d_l^k
\end{aligned}$$

linkage disequilibrium

variance contains covariances
between additive effects a^i and a^k, a_j and a_l
between dominance effects d_j^i and d_l^k

H-W equilibrium

no covariance between a^i and d_j^i
inbreeding yields some covariance

Genetic model for BC and F$_2$

genotype effects at loci $B_i, i = 1, 2, \cdots, m$

genotype	$B_i B_i$	$B_i b_i$	$b_i b_i$
effect	a_i	d_i	$-a_i$

P_1 population fixed with alleles B_i
P_2 population fixed with allele b_i
ignore epistasis

- examine trait distribution

- familiarize with various crosses

- underlying model for QTL

- get ready for epistasis

phenotypic value of individual k

P_1 population:

$$y_k = \mu + \sum_{i=1}^m a_i + e_k, \quad e_k \sim N(0, \sigma_e^2)$$

mean and variance of trait in P_1:

$$\mu_{P_1} = \mu + \sum_{i=1}^m a_i, \quad \sigma_{P_1}^2 = \sigma_e^2$$

Similarly for P_2:

$$y_k = \mu - \sum_{i=1}^m a_i + e_k, \quad \mu_{P_2} = \mu - \sum_{i=1}^m a_i, \quad \sigma_{P_2}^2 = \sigma_e^2$$

$F_1 = P_1 \times P_2$ population:

$$y_k = \mu + \sum_{i=1}^m d_i + e_k, \quad \mu_{F_1} = \mu + \sum_{i=1}^m d_i, \quad \sigma_{F_1}^2 = \sigma_e^2$$

Backcross $B_1 = P_1 \times F_1$

$$
\begin{aligned}
y_k &= \mu + \sum_{i=1}^{m} [x_{ik}a_i + (1 - x_{ik})d_i] + e_k \\
&= \mu + \sum_{i=1}^{m} d_i + \sum_{i=1}^{m} x_{ik}(a_i - d_i) + e_k
\end{aligned}
$$

$$
x_{ik} = \begin{cases} 1 & \text{if a } B_i \text{ allele in } k\text{'s } F_1 \text{ gamete} \\ 0 & \text{if a } b_i \text{ allele} \end{cases}
$$

$\mathcal{E}(x_{ik}) = 1/2, \ Var(x_{ik}) = 1/4 \ Cov(x_{ik}, x_{jk}) = (1 - 2r_{ij})/4$

mean and variance

$$
\mu_{B_1} = \mu + \sum_{i=1}^{m} (a_i + d_i)/2
$$

$$
\sigma_{B_1}^2 = \sum_{i=1}^{m} \frac{(a_i - d_i)^2}{4} + \sum_{i \neq j} \frac{(1 - 2r_{ij})(a_i - d_i)(a_j - d_j)}{4} + \sigma_e^2
$$

similarly for $B_2 = P_2 \times F_1$

$F_2 = F_1 \times F_1$ **Population**

$$
\begin{aligned}
y_k &= \mu + \sum_{i=1}^{m} [x_{ik1}x_{ik2}a_i + (1 - x_{ik1})(1 - x_{ik2})(-a_i) \\
&\quad + [x_{ik1}(1 - x_{ik2}) + (1 - x_{ik1})x_{ik2}]d_i] + e_k \\
&= \mu - \sum_{i=1}^{m} a_i + \sum_{i=1}^{m} [x_{ik\cdot}(a_i + d_i) - 2x_{ik1}x_{ik2}d_i] + e_k
\end{aligned}
$$

$$
x_{ik1} = \begin{cases} 1 & \text{if } B_i \text{ allele in first gamete from } F_1 \\ 0 & \text{if } b_i \text{ allele} \end{cases}
$$

$$
\mathcal{E}(x_{ik1}) = \frac{1}{2}, \quad Var(x_{ik1}) = \frac{1}{4}
$$

$$
Cov(x_{ik1}, x_{jk1}) = \frac{1}{4}(1 - 2r_{ij})
$$

and similarly for x_{ik2}

Under the assumption of Hardy-Weinberg equilibrium, x_{ik1} and x_{ik2} are independent. With

---

these conditions, the mean and variance of individuals in F_2 population become

$$
\mu_{F_2} = \mu + \frac{1}{2} \sum_{i=1}^{m} d_i
$$

$$
\begin{aligned}
\sigma_{F_2}^2 &= \frac{1}{2} \sum_{i=1}^{m} a_i^2 + \frac{1}{4} \sum_{i=1}^{m} d_i^2 + \frac{1}{2} \sum_{i \neq j} (1 - 2r_{ij})a_i a_j \\
&\quad + \frac{1}{4} \sum_{i \neq j} (1 - 2r_{ij})^2 d_i d_j + \sigma_e^2.
\end{aligned}
$$

first term: additive variance within loci
second: dominance variance within loci
third: additive covariance between loci
fourth: dominance covariance between loci

Epistasis

section 4.6 of Zeng's notes
postponed to chapter 8

Multiple Interval Mapping