GENETICS · STUDY GUIDE

Exam 6 Study Guide

This guide covers chapters 11 and 12, the two most mathematical chapters of the course. Chapter 11 measures traits that vary smoothly and splits their variation into genetic and environmental parts. Chapter 12 follows allele frequencies through whole populations and asks what moves them. The formulas are short, but you must know when each one applies. Work one chapter at a time, copy the formula tables by hand, and try every worked example with the steps covered. Then take the ten item self-test. The cram sheet at the end holds every formula on one page.

WHAT THIS GUIDE COVERS

- Chapter 11 · continuous, meristic and threshold traits, the multiple factor hypothesis, mean, variance, covariance, correlation and regression, variance components, H^2 and h^2, twins, the breeder's equation, QTL mapping, GWAS and polygenic scores.
- Chapter 12 · allele and genotype frequencies, Hardy Weinberg and its chi square test, inbreeding and F, mutation, migration, drift and N_e, selection and balance, the neutral theory and molecular clock, isolating barriers and speciation.

CHAPTER 11

Quantitative Genetics

What to know

Most traits vary smoothly. **Continuous traits** take any value, such as height. **Meristic traits** are counted, such as bristle number. **Threshold traits**, such as cleft lip or type 2 diabetes, look present or absent but rest on a continuous liability. The risk to relatives rises with how close they are, how many are affected and how severe the case is.

The **multiple factor hypothesis** says several Mendelian genes, each with a small additive effect, add up. With n genes in an F_2, there are 2n + 1 classes, and $(1/4)^n$ of offspring match one extreme

parent. Class frequencies follow the binomial: $C(2n, k) \times (1/2)^{2n}$. The environment blurs the classes into a bell curve, so you cannot read genotype from phenotype.

The **variance** is the sum of squared deviations divided by n − 1. Its square root is the standard deviation. Independent variances add, which is why we can partition them. The standard error, $s/\sqrt{n}$, describes the precision of a mean, not the spread of individuals. **Covariance** measures joint variation, **r** = $\text{cov}/(s_x s_y)$ standardizes it, and the regression slope $b = \text{cov}/s_x^2$ predicts y from x.

Phenotypic variance splits into $V_P = V_A + V_D + V_I + V_E$. Only the additive part, V_A, makes offspring resemble parents, because parents pass alleles, not genotypes. **Broad sense heritability** is $H^2 = V_G/V_P$. **Narrow sense heritability** is $h^2 = V_A/V_P$, and h^2 can never be larger than H^2.

Estimate heritability four ways. Inbred lines and their F_1 give V_E, so $V_G = V_{P(F2)} - V_E$. The slope of offspring on **midparent** is h^2, and on one parent it is $h^2/2$. Twins give $h^2 \approx 2(r_{MZ} - r_{DZ})$ and $c^2 = 2r_{DZ} - r_{MZ}$. Selection gives realized heritability, R/S.

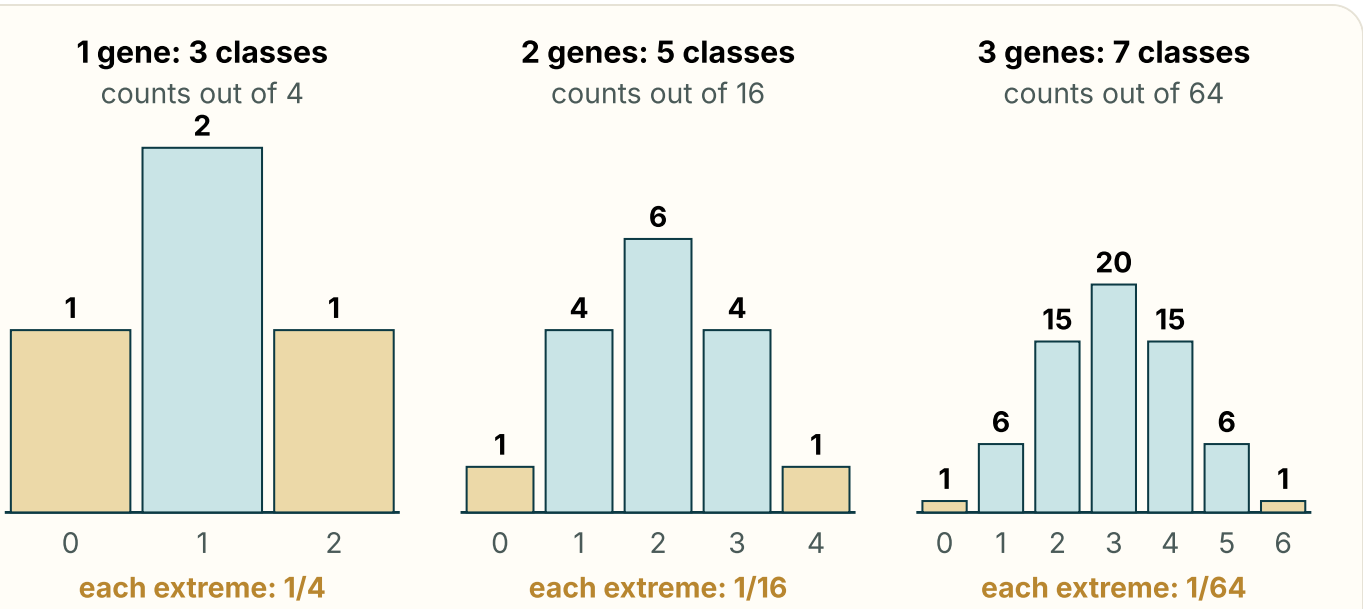
Heritability describes one population in one environment. It is never the genetic share of one person's trait. A highly heritable trait can still change, as average height did with better nutrition. Heritability within groups says nothing about the cause of differences between groups.

The **breeder's equation** is $R = h^2S$. S is the selected parents' mean minus the population mean, and R is the offspring mean minus the population mean. With selection intensity i, $S = i\sigma_P$. Pleiotropy and linkage cause **correlated responses**. Long term selection keeps working until additive variance runs out or natural selection pushes back.

QTL mapping in crosses compares trait means across marker genotypes. The additive effect a is half the difference between homozygote means, and d is the heterozygote mean minus their midpoint. LOD 3 is the usual bar, and resolution is coarse. **GWAS** tests millions of SNPs in populations using linkage disequilibrium, with a threshold of 5×10^{-8}, which is 7.3 on a Manhattan plot. Population stratification can fake signals. A polygenic score sums effect alleles times effect sizes and works best in the ancestry it was built from.

LOCK THESE IN

- n genes: $2n + 1$ classes; $(1/4)^n$ at each extreme.
- Variance divides by $n - 1$. Variances add; standard deviations do not.
- $r = cov/(s_x s_y)$. $b = cov/s_x^2$.
- $H^2 = V_G/V_P$. $h^2 = V_A/V_P$. $h^2 \leq H^2$.
- Midparent slope = h^2. One parent slope = $h^2/2$.
- $h^2 \approx 2(r_{MZ} - r_{DZ})$.
- $R = h^2 S$; realized $h^2 = R/S$; $S = i\sigma_P$.
- GWAS threshold $5 \times 10^{-8} = 7.3$ on the $-\log_{10}$ scale.



x axis: number of additive (capital) alleles. Bars drawn to scale; the tallest bar, 2/4, is one half. More genes give more, smaller steps. Add environmental noise and the steps blur into a bell curve.

Many genes, small effects. With n additive genes the F_2 has $2n + 1$ classes in binomial proportions, and $(1/4)^n$ match each parent.

Formulas for quantitative genetics

Quantity	Formula	Note
Classes and extremes	$2n + 1$ classes; $(1/4)^n$ at one extreme	n additive genes, F_2
Class frequency	$C(2n, k) \times (1/2)^{2n}$	k contributing alleles
Effect per allele	(extreme difference) ÷ 2n	equal additive alleles
Variance	$s^2 = \Sigma(x - \bar{x})^2 \div (n - 1)$	squared units
Standard error	$s \div \sqrt{n}$	precision of the mean
Covariance	$\Sigma(x - \bar{x})(y - \bar{y}) \div (n - 1)$	
Correlation	$r = cov \div (s_x\, s_y)$	−1 to +1
Regression slope	$b = cov \div s_x^2$; $a = \bar{y} - b\,\bar{x}$	midparent slope = h^2
Variance parts	$V_P = V_A + V_D + V_I + V_E$	plus G×E and 2cov(G,E) if present
Heritabilities	$H^2 = V_G/V_P$; $h^2 = V_A/V_P$	
Twins (Falconer)	$h^2 \approx 2(r_{MZ} - r_{DZ})$; $c^2 = 2r_{DZ} - r_{MZ}$	rest is unique environment
Breeder's equation	$R = h^2S$; $S = i\sigma_P$	i: top 50% 0.80; 20% 1.40; 10% 1.76; 1% 2.67
QTL effects	$a = (hom_1 - hom_2)/2$; d = het − midpoint	d = 0 means additive
GWAS threshold	$P = 5 \times 10^{-8}$; $-\log_{10}P = 7.3$	$0.05 \div 10^6$ tests
Polygenic score	Σ (effect alleles × effect size)	ancestry dependent

Worked examples

WORKED EXAMPLE 1 · COUNTING POLYGENES

Two true breeding lines average 30 and 54 cm. The F_1 averages 42 cm. Among 640 F_2 plants, 10 are about 30 cm and 10 about 54 cm. How many genes are there, how much does each contributing allele add, and what fraction of the F_2 should be 46 cm?

STEP 1 · EXTREME FRACTION

10 ÷ 640 = 1/64.

STEP 2 · SOLVE FOR N

$(1/4)^n = 1/64$, so **n = 3 genes**, with 6 allele positions and 7 classes.

STEP 3 · EFFECT PER ALLELE

(54 − 30) ÷ 6 = **4 cm**.

STEP 4 · THE 46 CM CLASS

(46 − 30) ÷ 4 = 4 contributing alleles. $C(6, 4) \times (1/2)^6 = 15/64$, about **23 percent**.

CHECK

The F_1 has 3 of 6 contributing alleles: 30 + 3 × 4 = 42 cm, which matches.

WORKED EXAMPLE 2 · COVARIANCE, CORRELATION AND SLOPE

Midparent values (x) and mean offspring values (y) for five families are: (4, 6), (6, 6), (8, 8), (10, 9), (12, 11). Find the covariance, r and the regression slope.

STEP 1 · MEANS

$\bar{x} = 40/5 = 8$. $\bar{y} = 40/5 = 8$.

STEP 2 · DEVIATIONS

x: −4, −2, 0, +2, +4. y: −2, −2, 0, +1, +3.

STEP 3 · COVARIANCE

Cross products: 8, 4, 0, 2, 12. Sum = 26, so cov = 26 ÷ 4 = **6.5**.

STEP 4 · VARIANCES

$s_x^2 = 40 \div 4 = 10$. $s_y^2 = (4 + 4 + 0 + 1 + 9) \div 4 = 4.5$.

STEP 5 · R AND B

r = 6.5 ÷ (3.162 × 2.121) = **0.97**. b = 6.5 ÷ 10 = **0.65**, so $h^2 \approx 0.65$.

WHY MIDPARENT MATTERS

With a single parent on the x axis, the slope would estimate $h^2/2$, and you would double it.

WORKED EXAMPLE 3 · SPLITTING VARIANCE WITH INBRED LINES

Two inbred lines have variances of 3.0 and 5.0, their F_1 4.0, and the F_2 16.0. Another analysis gives V_A = 6.4 in the F_2. Find V_E, V_G, H^2, h^2 and $V_D + V_I$.

STEP 1 · V_E

The uniform groups have only environmental variance: (3.0 + 5.0 + 4.0) ÷ 3 = **4.0**.

STEP 2 · V_G

16.0 − 4.0 = **12.0**.

STEP 3 · H^2

12.0 ÷ 16.0 = **0.75**.

STEP 4 · H^2

6.4 ÷ 16.0 = **0.40**.

STEP 5 · NONADDITIVE PART

12.0 − 6.4 = **5.6**. Check: 0.40 ≤ 0.75.

WORKED EXAMPLE 4 · THE BREEDER'S EQUATION

A trait has a mean of 60 and a phenotypic standard deviation of 5. A breeder keeps the top 20 percent (i = 1.40). If h^2 = 0.40, predict the next generation mean.

STEP 1 · SELECTION DIFFERENTIAL

$S = i\sigma_P = 1.40 \times 5 = 7.0$.

STEP 2 · RESPONSE

$R = h^2S = 0.40 \times 7.0 = 2.8$.

STEP 3 · NEW MEAN

60 + 2.8 = **62.8**.

CHECK

R must be smaller than S whenever h^2 is below 1: 2.8 < 7.0.

TRAPS PEOPLE FALL INTO

- **Dividing variance by n.** Use n − 1 for a sample.
- **Mixing up standard deviation and standard error.** SD describes individuals. SE describes the mean and shrinks with sample size.
- **Swapping H^2 and h^2.** Only h^2 predicts response to selection in a sexual population.
- **Forgetting to double a one parent slope.** One parent gives $h^2/2$.
- **Reading heritability as a personal share.** It describes variation in a population, not one person.

- **Inferring group differences from heritability.** Two fields with one seed bag show why that fails.
- **Reading a Manhattan plot upside down.** Taller points mean smaller P values.
- **Treating a GWAS hit as the causal variant.** It is usually a neighbor in linkage disequilibrium.

Chapter 11 self-test · ten items

1. In an F_2, 1/256 of plants match each extreme parent. How many additive genes are involved, and how many classes appear?

A. 16 genes and 33 classes

B. 8 genes and 17 classes

C. 4 genes and 9 classes

D. 4 genes and 8 classes

C. 4 genes and 9 classes. $(1/4)^4 = 1/256$, so n = 4. Classes = 2n + 1 = 9.

2. Why is a trait like pyloric stenosis called a threshold trait?

A. It is present or absent, but rests on a continuous liability

B. It shows a 3:1 ratio in every affected family

C. It appears only when two specific recessive genes meet

D. It is caused by one dominant allele with full penetrance

A. It is present or absent, but rests on a continuous liability. Many genes and environmental factors add to liability. The condition appears only when liability passes a threshold.

3. Four values are 3, 5, 7 and 9. What is the sample variance?

A 2.58

B 5.00

C 20.0

D 6.67

D. 6.67. Mean 6; squared deviations 9, 1, 1, 9 sum to 20. Dividing by n − 1 = 3 gives 6.67.

4. Offspring are regressed on one parent, not the midparent, and the slope is 0.22. What is h^2?

A 0.88

B 0.44

C 0.11

D 0.22

B. 0.44. A single parent carries only half of each child's alleles, so the slope estimates $h^2/2$. Double it.

5. Identical twins correlate at 0.70 and fraternal twins at 0.45. What is the shared environment estimate, c^2?

A 0.30

B 0.20

C 0.50

D 0.25

B. 0.20. $c^2 = 2r_{DZ} - r_{MZ} = 0.90 - 0.70 = 0.20$. Here $h^2 = 2(0.25) = 0.50$ and the unique environment is 0.30.

6. Height is highly heritable. Average height in a country rose 10 cm in a century. How do these fit together?

A. Taller people simply had many more children then

B. The genes for height must have changed quickly

C. Heritability does not mean the trait cannot change

D. Heritability must have fallen to zero over that time

C. Heritability does not mean the trait cannot change. Better nutrition shifted the whole population. High heritability describes variation within a population in its current environment.

7. Mice average 40 g. Selected parents average 46 g, and their offspring average 42.4 g. What is the realized heritability?

A. 0.40

B. 2.40

C. 0.60

D. 0.06

A. 0.40. S = 6 g and R = 2.4 g, so h^2 = R/S = 0.40.

8. Broilers selected for fast growth also developed leg problems. What is this called?

A. Regression to the mean

B. A selection plateau

C. Realized heritability

D. A correlated response

D. A correlated response. Genes that affect both growth and legs (pleiotropy), or linked genes, move the second trait too.

9. At a QTL marker, mean values are 30 (M1M1), 28 (M1M2) and 22 (M2M2). What are a and d?

A a = 4 and d = 2

B a = 4 and d = 0

C a = 2 and d = 4

D a = 8 and d = 0

A. a = 4 and d = 2. Midpoint = 26. a = (30 − 22)/2 = 4. d = 28 − 26 = 2, so the M1 linked allele is partly dominant.

10. Why can population stratification create false GWAS hits?

A SNPs in stratified samples cannot be genotyped well

B Cases and controls differ in ancestry as well as disease

C The 5×10^{-8} threshold is too strict for them

D Linkage disequilibrium vanishes in mixed samples

B. Cases and controls differ in ancestry as well as disease. Any SNP that differs between ancestries will look associated with disease if ancestry differs between groups. Principal components correct for it.

Read the full chapter: Chapter 11 · Quantitative Genetics

CHAPTER 12

Population and Evolutionary Genetics

What to know

A **gene pool** is described by genotype frequencies (shares of individuals) and allele frequencies (shares of gene copies). For codominant alleles, p = f(AA) + ½f(Aa), or count alleles over 2N. For X linked loci, count two X per female and one per male. A locus with k alleles has k(k + 1)/2 genotypes. Evolution is a change in allele frequencies.

Hardy Weinberg: in a large, randomly mating population with no mutation, migration or selection, allele frequencies stay constant and genotypes settle at p^2, 2pq and q^2 after one generation. Dominance does not make an allele spread. For a rare recessive, q = √(affected) and carriers 2pq ≈ 2q far outnumber affected people. For X linked recessives, affected males are q and affected females q^2.

Test equilibrium with chi square. For two alleles, df = 3 classes − 1 − 1 estimated allele frequency = **1**, so the critical value is 3.84.

Nonrandom mating changes genotype frequencies but not allele frequencies. The **inbreeding coefficient** F is the chance two alleles are identical by descent. Genotypes become p^2 + Fpq, 2pq(1 − F) and q^2 + Fpq. Find F by path counting: $F = \Sigma(1/2)^n(1 + F_A)$, where n counts people in each loop. First cousins give 1/16. Inbreeding exposes recessive alleles (inbreeding depression), and crossing inbred lines gives heterosis.

Mutation is the source of new alleles but moves frequencies very slowly. **Migration** pulls frequencies toward the source: $\Delta q = m(q_m - q)$. **Genetic drift** is random and strongest when N_e is small. A neutral allele fixes with probability equal to its frequency, and heterozygosity falls by $1/(2N_e)$ each generation. N_e is usually smaller than the census: $4N_mN_f/(N_m + N_f)$ for unequal sexes, and a harmonic mean across generations. Founder effects and bottlenecks are special cases.

For **selection**, weight each genotype frequency by its fitness, divide by $\bar{w}$ and recompute q. Selection against a rare recessive is slow, because nearly all copies hide in carriers. **Heterozygote advantage** keeps both alleles at $\hat{q} = s/(s + t)$, as with the sickle cell allele where malaria is common. **Mutation selection balance** gives $\hat{q} \approx \sqrt{(u/s)}$ for a recessive and u/s for a dominant.

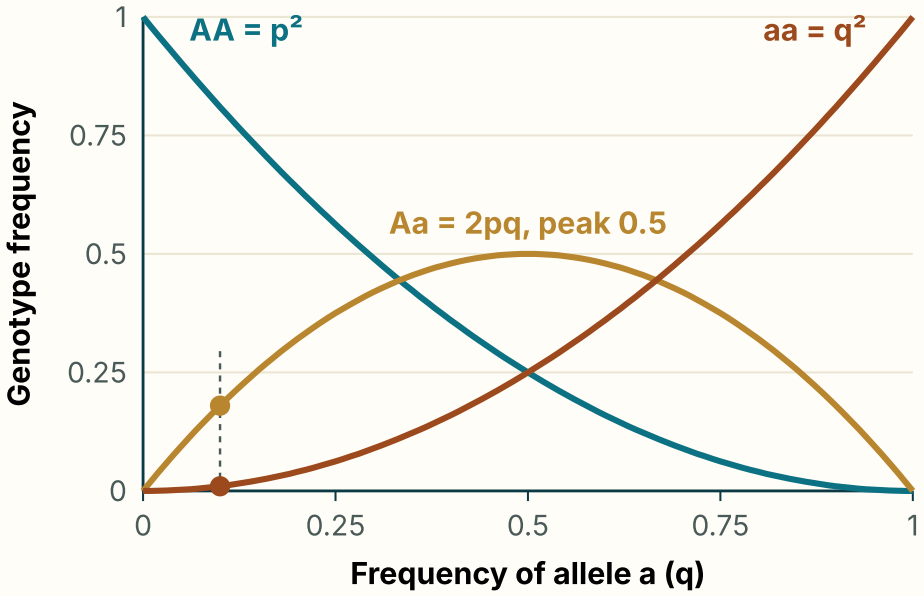
The **neutral theory** says most molecular differences are neutral and fixed by drift, so the neutral substitution rate equals the mutation rate. That gives a **molecular clock**: $K = 2\mu T$, with the 2 because both lineages change. The Jukes Cantor correction fixes saturation in divergent sequences. Constrained sequences evolve slowly. dN/dS below 1 shows purifying selection and above 1 shows positive selection.

Species are kept apart by **prezygotic** barriers (habitat, timing, behavior, mechanics, gametes) and **postzygotic** ones (hybrid inviability, sterility, breakdown). Most speciation is allopatric.

Polyploidy gives fast sympatric speciation in plants. Dobzhansky Muller incompatibilities explain how hybrids can fail without any unfit intermediate. Haldane's rule says the heterogametic hybrid sex suffers first.

LOCK THESE IN

- p = f(AA) + ½f(Aa). p + q = 1.
- Hardy Weinberg: $p^2 + 2pq + q^2 = 1$, reached in one generation of random mating.
- Rare recessive: q = √(affected); carriers ≈ 2q.
- HW chi square with two alleles has df = 1.
- Inbreeding: heterozygotes × (1 − F); each homozygote + Fpq. First cousins F = 1/16.
- Neutral fixation probability = current frequency. Heterozygosity loss per generation = $1/(2N_e)$.
- Heterozygote advantage: $\hat{q} = s/(s + t)$.
- Molecular clock: K = 2μT. Do not forget the 2.



At q = 0.1:
affected aa = 0.01
carriers Aa = 0.18
Carriers outnumber affected 18 to 1.

Rare recessive:
q = √(affected)
carriers ≈ 2q

p + q = 1
$p^2 + 2pq + q^2 = 1$

Hardy Weinberg at every allele frequency. Heterozygotes can never pass 50 percent. When an allele is rare, almost every copy hides in a carrier.

Formulas for population genetics

Quantity	Formula	Note
Allele frequency	$p = (2N_{AA} + N_{Aa}) \div 2N$	or f(AA) + ½f(Aa)
Hardy Weinberg	p^2, 2pq, q^2	three alleles: $(p + q + r)^2$
Rare recessive	q = √(affected); carriers 2pq ≈ 2q	carriers : affected = 2p/q
X linked	affected males q; affected females q^2	count X chromosomes
HW chi square	df = classes − 1 − alleles estimated	two alleles: df 1, critical 3.84
Inbreeding	$p^2 + Fpq$; 2pq(1 − F); $q^2 + Fpq$	$F = 1 - H_o/H_e$
Path counting	$F = \Sigma(1/2)^n(1 + F_A)$	n = people in the loop
Selfing	$F = 1 - (1/2)^t$	t generations
Mutation equilibrium	$\hat{q} = u \div (u + v)$	very slow approach
Migration	$q' = (1 - m)q + mq_m$; gap shrinks by $(1 - m)^t$	
Drift	$H_t = H_0(1 - 1/2N_e)^t$; P(fix) = frequency	new mutation: 1/(2N)
Effective size	$N_e = 4N_mN_f \div (N_m + N_f)$; harmonic mean over time	usually below census
Selection	multiply by w, sum to $\bar{w}$, divide, recompute q	recessive lethal: $t = 1/q_t - 1/q_0$
Balance	$\hat{q} = s/(s + t)$; recessive $\hat{q} \approx \sqrt{(u/s)}$; dominant $\hat{q} \approx u/s$	
Clock and correction	$K = 2\mu T$; $d = -(3/4) \ln(1 - (4/3)p)$	Jukes Cantor

Worked examples

WORKED EXAMPLE 1 · A HARDY WEINBERG CHI SQUARE

A sample of 400 people has 230 AA, 140 Aa and 30 aa. Are they in Hardy Weinberg equilibrium?

STEP 1 · ALLELE FREQUENCIES

p = (2 × 230 + 140) ÷ 800 = 600 ÷ 800 = 0.75. q = 0.25.

STEP 2 · EXPECTED NUMBERS

AA: 0.5625 × 400 = 225. Aa: 0.375 × 400 = 150. aa: 0.0625 × 400 = 25.

STEP 3 · CHI SQUARE

25/225 + 100/150 + 25/25 = 0.111 + 0.667 + 1.000 = **1.78**.

STEP 4 · DECIDE

df = 3 − 1 − 1 = 1, critical 3.84. Since 1.78 < 3.84, **fail to reject**: the sample fits equilibrium.

WORKED EXAMPLE 2 · ONE GENERATION OF SELECTION

Fitnesses are AA = 1, Aa = 1 and aa = 0.6. If q = 0.5, what is q after one generation of selection?

STEP 1 · BEFORE SELECTION

AA 0.25, Aa 0.50, aa 0.25.

STEP 2 · WEIGHT BY FITNESS

0.25, 0.50 and 0.25 × 0.6 = 0.15.

STEP 3 · MEAN FITNESS

$\bar{w}$ = 0.25 + 0.50 + 0.15 = 0.90.

STEP 4 · NEW Q

q′ = (0.15 + ½ × 0.50) ÷ 0.90 = 0.40 ÷ 0.90 = **0.444**.

CHECK

q fell by 0.056. As a recessive allele gets rare, each generation removes less, because most copies hide in carriers.

WORKED EXAMPLE 3 · EFFECTIVE SIZE AND LOST DIVERSITY

A herd has 20 breeding males and 80 breeding females. Find N_e and the fraction of heterozygosity left after 20 generations.

STEP 1 · N_E

4 × 20 × 80 ÷ (20 + 80) = 6,400 ÷ 100 = **64**, well below the census of 100.

STEP 2 · LOSS PER GENERATION

1 ÷ (2 × 64) = 1/128.

STEP 3 · AFTER 20 GENERATIONS

$(1 - 1/128)^{20}$ = **0.855**, so about 15 percent of heterozygosity is lost.

WORKED EXAMPLE 4 · INBREEDING AND A RARE RECESSIVE

A recessive condition has q = 0.02. Compare its frequency among children of unrelated parents with that among children of an uncle and niece (F = 1/8).

STEP 1 · RANDOM MATING

q^2 = 0.0004, or 1 in 2,500.

STEP 2 · WITH INBREEDING

$q^2 + Fpq = 0.0004 + 0.125 \times 0.98 \times 0.02 = 0.0004 + 0.00245 = 0.00285$.

STEP 3 · COMPARE

0.00285 ÷ 0.0004 = **about 7 fold** higher. The rarer the allele, the bigger the jump.

FRAMING

This is a population level risk. Families decide for themselves whether to seek carrier testing or counseling.

TRAPS PEOPLE FALL INTO

- **Using df = 2 for a two allele HW test.** One more df is lost for the allele frequency you estimated. df = 1.
- **Thinking dominant alleles become common.** Hardy Weinberg holds whatever the dominance.
- **Saying inbreeding changes allele frequencies.** It changes genotype frequencies only.
- **Counting n wrong in path counting.** Count the people in the loop, including both parents and the common ancestor, not the inbred child.
- **Using the arithmetic mean for N_e over time.** Use the harmonic mean. One small generation drags it down.
- **Expecting selection to remove a rare recessive fast.** Almost all copies hide in carriers.
- **Leaving out the 2 in K = 2μT.** Both lineages have been changing.
- **Calling a mule's sterility prezygotic.** It acts after the zygote forms, so it is postzygotic.

Chapter 12 self-test · ten items

1. In 200 people, 50 are MM, 100 are MN and 50 are NN. What is the frequency of M?

A 0.33

B 0.75

C 0.50

D 0.25

C. 0.50. (2 × 50 + 100) ÷ 400 = 200 ÷ 400 = 0.50.

2. A recessive condition affects 1 in 10,000 people. About what fraction are carriers?

A About 1 in 100

B About 1 in 25

C About 1 in 5,000

D About 1 in 50

D. About 1 in 50. q = 0.01, p = 0.99, so 2pq = 0.0198, about 1 in 50.

3. About 10 percent of males in a population have an X linked recessive trait. What share of females should show it?

A About 18 percent

B About 5 percent

C About 1 percent

D About 10 percent

C. About 1 percent. Males show it at q = 0.10. Females need two copies: q^2 = 0.01. Carriers would be 2pq = 0.18.

4. What is the inbreeding coefficient of a child of half siblings who share a non inbred mother?

A 1/16

B 1/8

C 1/4

D 1/32

B. 1/8. One common ancestor and a loop of three people: $(1/2)^3 = 1/8$.

5. An island has q = 0.30. Each generation 20 percent of breeders come from a mainland with q = 0.70. What is q next generation?

A 0.34

B 0.50

C 0.44

D 0.38

D. 0.38. $q' = 0.8 \times 0.30 + 0.2 \times 0.70 = 0.24 + 0.14 = 0.38$.

6. A neutral allele has a frequency of 0.15 in a small population. What is its chance of eventually being fixed by drift?

A 0.15

B 0.50

C 0.0225

D 0.85

A. 0.15. For a neutral allele, the probability of fixation equals its current frequency.

7. With AA fitness 0.90, Aa fitness 1 and aa fitness 0.10, what is the equilibrium frequency of a?

A. 0.10

B. 0.90

C. 0.50

D. 0.18

A. 0.10. s = 0.10 and t = 0.90, so $\hat{q} = s/(s + t) = 0.10/1.00 = 0.10$.

8. A recessive lethal has $u = 10^{-5}$ and s = 1. About where does mutation selection balance settle?

A. $q \approx 0.03$

B. $q \approx 0.0001$

C. $q \approx 0.00001$

D. $q \approx 0.003$

D. $q \approx 0.003$. For a recessive, $\hat{q} \approx \sqrt{(u/s)} = \sqrt{10^{-5}} = 0.0032$. A dominant would sit near $u/s = 10^{-5}$.

9. Two species differ at 3 percent of neutral sites. The rate is 1.5×10^{-9} per site per year. When did they split?

A. About 2 million years ago

B. About 5 million years ago

C. About 10 million years ago

D. About 20 million years ago

C. About 10 million years ago. $T = K/(2\mu) = 0.03 \div (3 \times 10^{-9}) = 10^7$ years. Forgetting the 2 gives 20 million.

10. Two populations each fix a new allele that works fine alone. Their hybrids are sterile. Which model explains this?

A. Reinforcement of mating barriers

B. Dobzhansky Muller incompatibility

C. Heterozygote advantage at one locus

D. Haldane's rule of hybrid sexes

B. Dobzhansky Muller incompatibility. Alleles A and B never met in either population. They clash only in the hybrid, so no unfit intermediate was ever needed.

Read the full chapter: Chapter 12 · Population and Evolutionary Genetics

ONE PAGE

Cram sheet

Every formula from chapters 11 and 12 in one place. Cover the right column and see if you can rebuild each row.

Chapter 11 formulas

Name	Formula	Note
Polygenes	$2n + 1$ classes; $(1/4)^n$ extremes; $C(2n, k)(1/2)^{2n}$	additive alleles
Variance and SD	$s^2 = \Sigma(x - \bar{x})^2/(n - 1)$; $s = \sqrt{s^2}$	$SE = s/\sqrt{n}$
Covariance, r, b	cov; $r = cov/(s_x s_y)$; $b = cov/s_x^2$	r^2 = share explained
Variance parts	$V_P = V_A + V_D + V_I + V_E$	V_A is transmitted
Heritability	$H^2 = V_G/V_P$; $h^2 = V_A/V_P$	$h^2 \leq H^2$
Estimates	midparent slope = h^2; one parent = $h^2/2$; twins $2(r_{MZ} - r_{DZ})$	inbred lines give V_E
Selection	$R = h^2S$; $S = i\sigma_P$; realized $h^2 = R/S$	
QTL	a = half the homozygote difference; d = het − midpoint; LOD 3	coarse
GWAS	$5 \times 10^{-8} = 7.3$ on $-\log_{10}$; correct for stratification	polygenic scores ancestry dependent

Chapter 12 formulas

Name	Formula	Note
Frequencies	p = f(AA) + ½f(Aa); k alleles: k(k + 1)/2 genotypes	X: 2 per female, 1 per male
Hardy Weinberg	$p^2 + 2pq + q^2$	one generation of random mating
Recessive and carriers	q = √(affected); 2pq ≈ 2q	X linked: males q, females q^2
HW test	χ^2, df 1 for two alleles	critical 3.84
Inbreeding	$p^2 + Fpq$; 2pq(1 − F); $F = \Sigma(1/2)^n(1 + F_A)$	cousins 1/16; half sibs 1/8; full sibs 1/4
Mutation and migration	$\hat{q} = u/(u + v)$; $\Delta q = m(q_m - q)$	
Drift	H loss $1/(2N_e)$ per generation; P(fix) = frequency	$N_e = 4N_mN_f/(N_m + N_f)$
Selection	weight, sum to $\bar{w}$, divide; s = 1 − w	recessive lethal $t = 1/q_t - 1/q_0$
Balance	s/(s + t); √(u/s); u/s	overdominance; recessive; dominant
Molecular evolution	k = u; K = 2μT; Jukes Cantor; ω = dN/dS	ω < 1 purifying
Speciation	prezygotic vs postzygotic; allopatric, peripatric, parapatric, sympatric; Dobzhansky Muller; Haldane's rule	polyploidy in plants

Dr. Garg Learning · CC BY-SA 4.0