

EXERCISES FOR CHAPTER 8

Exercise 8.1. Define briefly, in words:

- Additive effect of an allele.
- Epistasis (in quantitative genetics).

Exercise 8.2. Discuss the parallelisms between the genic factorial model and the factorial method of experimentation.

Exercise 8.3. In Chapter 8 we adopted the factorial approach for one locus with multiple alleles and derived the following formulas

$$\sigma_{\alpha^m}^2 = \sum_{i=1}^m p_i^m (\alpha_i^m)^2 \quad \text{equation (8.18)}$$

$$\sigma_{\alpha^f}^2 = \sum_{j=1}^m p_j^f (\alpha_j^f)^2 \quad \text{equation (8.20)}$$

When $p_i^m = p_i^f$ for $i = 1, \dots, m$,

$$\sigma_{\alpha^m}^2 = \sigma_{\alpha^f}^2 = \sigma_{\alpha}^2 = \sum_{i=1}^m p_i (\alpha_i)^2 \quad \text{equation (8.27)}$$

and

$$\sigma_A^2 = 2\sigma_{\alpha}^2 = 2 \sum_{i=1}^m p_i (\alpha_i)^2 \quad \text{equation (8.29)}$$

The dominance variance is

$$\sigma_D^2 = \sigma_{\delta}^2 = \sum_{i=1}^m \sum_{j=1}^m p_i^m p_j^f \delta_{ij}^2 \quad \text{equation (8.21)}$$

a. Assuming only two alleles, present a diagram similar to that by D.S. Falconer and T.F.C. Mackay, 1996, Introduction to Quantitative Genetics, Fourth edition, Longman, p. 117, (change typographic error on that page from $-2q\alpha$ to $-2p\alpha$) showing the regression of genotypic value on number of A_1 alleles.

From the derivation for the factorial approach, why can we simply label the slope of the line, α , as $\alpha = \alpha_1 - \alpha_2$. Derive algebraically an expression for α_1 and α_2 as a function of α . Show these values on the graph.

b. By use of the equivalences presented or derived in (a), show that our equation (8.29) for two alleles equals Falconer and Mackay's equation [8.3b], namely,

$$\sigma_A^2 = 2 \sum_{i=1}^2 p_i (\alpha_i)^2 = 2pq\alpha^2$$

Exercise 8.4.

a. Calculate all of the additive and dominance effects, and calculate the additive and dominance variances by definition formulas, i.e., from the deviations, for the case of a random-mating population with three alleles A_1, A_2, A_3 for which the allelic frequencies are 0.3, 0.2 and 0.5, respectively, and the genotypic values for $A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3$, and A_3A_3 are 5, 3, 2, 1, 0, 1, respectively.

b. Calculate the additive and dominance variances by use of a computing formula.

Exercise 8.5. Repeat Exercise 8.4 when the genotypic values are 5, 5, 4, 5, 4, and 2, respectively.

Exercise 8.6.

- For the case of complete dominance for a single, two-allelic locus, derive expressions for the additive, dominance, and genotypic variances by use of equations (8.51) and (8.52).
- Calculate the general expression for the ratio of the dominance variance to the additive variance, the ratio of the additive variance to the genotypic variance, and finally the ratio of the dominance variance to the genotypic variance, as a function of the allelic frequency of the unfavorable or decreasing allele.
- Plot each of these ratios as a function of that allelic frequency. Compare them to the graphs in Fig. 8.3 (p. 8.54).
- In experiments workers have generally found that the ratio of σ_D^2/σ_A^2 or σ_D^2/σ_G^2 is low. From this information what might one be able to conclude about the frequency of the favorable allele?

Exercise 8.7. In Figure 8.3 (p. 8.54), we plotted the ratios σ_D^2/σ_A^2 , σ_A^2/σ_G^2 , and σ_D^2/σ_G^2 , as functions of the frequency p of the favorable allele. Considering only σ_A^2/σ_G^2 , it approached one as p approached zero (or zero as q approached zero from Exercise 8.6), and that it approached zero as p approached one (or one as q approached one from Exercise 8.6). Can you explain, in an intuitive sense or in words, why that ratio behaves that way as p approaches:

- One?
- Zero?

Explain the above in terms of the two variances, what each represents, and their relative magnitudes. The diagram depicting the regression of genotypic value on number of A_1 alleles in the genotype is probably most helpful.

- What does this change in the ratio, σ_A^2/σ_G^2 , which equals heritability, $\sigma_A^2/(\sigma_G^2 + \sigma_E^2)$, when $\sigma_E^2 = 0$, have to say about the role of allelic frequencies in determining additive, dominance, and genotypic variances, and their ratios?

Exercise 8.8. Again from Exercise 8.6, part d, we concluded that if σ_D^2/σ_A^2 or σ_D^2/σ_G^2 were low, the frequency of the dominant favorable allele must be relatively low in frequency. This conclusion does not seem to be consistent with other information in that one would think that natural selection would have favored the dominant homozygote and heterozygote thereby effecting a high frequency of the dominant allele. Thus, under these conditions one would expect to find $\sigma_D^2/\sigma_A^2 \gg 1$. This does not appear to be the case biologically.

One explanation might be that complete dominance, which was assumed above, is not at all prevalent for quantitative characters.

- Therefore repeat Exercise 8.6 for $G_{11} - G_{12} = \frac{1}{4}(G_{11} - G_{22})$, i.e., the heterozygote lies half the distance between the favorable homozygote G_{11} and the mid-homozygote $(G_{11} + G_{22})/2$. Give your formulas in terms of $(G_{11} - G_{22})$ and in terms of both p and q . However, plot them against p .
- Compare these results for σ_D^2/σ_A^2 with those for complete dominance. Is our conclusion (part d in Exercise 8.6) regarding the frequency of the dominant favorable allele altered? What conclusions do these results suggest with respect to dominance?

Exercise 8.9. Considering one autosomal locus only, under what conditions can one have a random-mating population with the total genotypic variance greater than zero but no additive variance? What relation exists among the genotypic values if $p_1 = p_2 = 1/2$?

Exercise 8.10. Compute the additive and dominance variances for each locus and the total epistatic variance for a random-mating population in linkage equilibrium with gene frequencies of 1/2 and the following genotypic values:

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	4	4	2
A_1A_2	3	3	1
A_2A_2	2	2	0

If linkage disequilibrium did exist, would the mean of this population be affected? The variances? Explain.

Exercise 8.11. Define rigorously the total additive-by-additive variance for a single random-mating population in linkage equilibrium. From what kind of table are the effects defined? In what way do the assumptions of random mating and linkage equilibrium come into play? Consider only two loci.

Exercise 8.12. With two equally frequent alleles at each of two loci, construct a set of genotypic values for which

- Only additive variance exists.
- Only dominance variance exists.
- Only additive and dominance variances exist.
- Only additive, and additive-by-additive variances exist.
- Additive, dominance, and only additive (at the A locus)-by-dominance (at the B locus) epistatic variances exist.
- Additive, dominance, and only dominance (at the A locus)-by-additive (at the B locus) epistatic variances exist.
- Additive, dominance, and only dominance-by-dominance epistatic variances exist.

Exercise 8.13. Suppose that in a Hardy-Weinberg population in linkage equilibrium with two alleles at each of two loci with all allelic frequencies equal to 1/2, we have the following genotypic values:

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	4	2	3
A_1A_2	4	3	1
A_2A_2	3	2	1

It is interesting to note that the three genotypes of the B locus illustrate negative overdominance [see equation (8.121)], partial dominance, and no dominance in the presence of A_1A_1 , A_1A_2 , and A_2A_2 , respectively. On the other hand, the three genotypes of the A locus illustrate dominance (i.e., dominance of the superior allele), overdominance, complete recessiveness (i.e., complete dominance of the inferior allele) in the presence of B_1B_1 , B_1B_2 , and B_2B_2 , respectively.

- Calculate the following variances by use of the definition formula of a variance i.e., by use of deviations: (That is, calculate deviations of genotypic values from predicted genotypic values for every appropriate cell. Do this by constructing the appropriate table, by calculating the predicted genotypic value in each cell by adding the overall mean, the row effect, and column effect, and then calculating the deviation or residual effect.)

- i. σ_G^2 = total genotypic variance
- ii. $\sigma_{y_a}^2$ = genotypic variance due to locus a
- iii. $\sigma_{y_b}^2$ = genotypic variance due to locus b
- iv. $\sigma_{yy_{ab}}^2$ = interlocus or epistatic variance between loci a and b
- v. $\sigma_{A_a}^2$ = additive variance at locus a
- vi. $\sigma_{D_a}^2$ = dominance variance at locus a
- vii. $\sigma_{AA_{ab}}^2$ = additive-by-additive variance
- viii. $\sigma_{AD_{ab}}^2$ = additive (at locus a)-by-dominance (at locus b) variance

Perform the following checks:

check i. $\sigma_G^2 = \sigma_{y_a}^2 + \sigma_{y_b}^2 + \sigma_{yy_{ab}}^2$

check ii. $\sigma_{y_a}^2 = \sigma_{A_a}^2 + \sigma_{D_a}^2$

b. Calculate the following variances by use of appropriate computing formulas

i. $\sigma_{A_b}^2$ = additive variance at locus b

ii. $\sigma_{D_b}^2$ = dominance variance at locus b

iii. $\sigma_{AA_{ab}}^2$ = additive-by-additive variance (should check with that calculated in a.vii)

iv. $\sigma_{AD_{ab}}^2$ = additive (at locus a)-by-dominance (at locus b) variance (should check with that

calculated in a.viii)

v. $\sigma_{DA_{ab}}^2$ = dominance (at locus a)-by-additive (at locus b)

vi. $\sigma_{DD_{ab}}^2$ = dominance-by-dominance variance

Perform the following checks:

check iii. $\sigma_{y_b}^2 = \sigma_{A_b}^2 + \sigma_{D_b}^2$

check iv. $\sigma_{yy_{ab}}^2 = \sigma_{AA_{ab}}^2 + \sigma_{AD_{ab}}^2 + \sigma_{DA_{ab}}^2 + \sigma_{DD_{ab}}^2$

c. By use of the above variances already computed, calculate each of the following:

i. σ_A^2 = additive variance

ii. σ_D^2 = dominance variance

iii. σ_I^2 = epistatic variance

Exercise 8.14. In equation (8.163), I give formulas for the two-locus epistatic variances, but I did not derive them. What I want you to do is to derive the formula for the additive-by-additive variance only (8.163a). To do so, you need to understand thoroughly the use of orthogonal polynomial coefficients (the X variables in regression) in statistics in partitioning the sum of squares of the dependent variable (the Y or genotypic variable) for quantitative factors. See Steel, R.G.D., J.H. Torrie, and D.A. Dickey, 1997, Principles and Procedures of Statistics. Third edition, McGraw-Hill, NY, Sections 8.3 and 15.7; Snedecor, G.W. and W.G. Cochran, 1967, Statistical Methods, Sixth edition, Iowa State Press, Ames, Section 12.8.

In our genetic problem, the levels of the factors, i.e., the number of A_1 and B_1 alleles in an individual (0, 1, 2), are equally spaced, but the frequencies of the genotypes are not equal, as the levels are in a 3×3 factorial experiment with equal number of replicates for all treatment combinations. Hence, we must derive the appropriate orthogonal polynomial coefficients for unequal frequencies.

To lead into the additive-by-additive variance, which is the problem at hand, we must first investigate the additive-variance case at a single locus. After understanding that simple case, it is a matter of extension to the additive-by-additive variance case. First, confining our attention to a single locus, locus A , we need the orthogonal polynomial coefficients (denoted c_{1i} – analogous to that in statistics for the linear coefficients for the first factor) for the additive variance at the A locus. Once we have the appropriate orthogonal polynomial coefficients, we can calculate the additive contrast for the A locus, which is the covariance, $\text{cov}(c_1, G)$, between the additive or linear orthogonal polynomial coefficients for the A locus and the genotypic values for the A locus (analogous to the linear contrast Q_L in statistics). Second, we calculate the variance, $\sigma_{c_1}^2$, of the independent variable, i.e., the variance of the orthogonal polynomial coefficients. Then, we can calculate the regression coefficient (also often symbolized α_1 in statistics) [see equation (8.63)]

$$\hat{\beta} = \beta = \alpha \left(\begin{array}{l} \text{i.e., average effect of an} \\ \text{allelic substitution} \end{array} \right) = \alpha_1 = \frac{\text{cov}(c_1, G)}{\sigma_{c_1}^2}$$

(I highly recommend that you study pp. 8.30 to 8.34 carefully, but especially Box 8.6.) Then the additive variance is, i.e., the regression coefficient times the corresponding element on the right-hand side of the one normal equation (see the first full paragraph on p. 8.14 for the same idea involving possibly several normal equations)

$$\sigma_{A_a}^2 = \alpha_1 \text{cov}(c_1, G) = \frac{[\text{cov}(c_1, G)]^2}{\sigma_{c_1}^2}$$

(This is analogous to the common formula for the sum of squares for any contrast $SS(Q) = Q^2 / r \sum c^2$.) This is an overview of the approach for the additive variance at the A locus.

To be more specific, by an examination of Box 8.6, one can readily deduce the orthogonal polynomial coefficients for the additive contrast used to calculate the additive variance for locus A . The orthogonal polynomial coefficients for the A locus are simply the deviations of the X values from their mean, $2p$ [equation (8.61)]:

	Frequency	X	$X - \bar{X}$	$c_1 = X - \bar{X}$
A_1A_1	p^2	2	$2 - 2p$	$2q$
A_1A_2	$2pq$	1	$1 - 2p$	$q - p$
A_2A_2	q^2	0	$0 - 2p$	$-2p$

Note that these orthogonal polynomial coefficients weighted by their frequencies do sum to zero as other orthogonal polynomial coefficients do:

$$\begin{aligned} p^2(2q) + 2pq(q - p) + q^2(-2p) &= 2pq(p + q - p - q) \\ &= 0 \end{aligned}$$

Then

$$\begin{aligned}\text{cov}(c_1, G) &= p^2(2q)G_{11} + 2pq(q-p)G_{12} + q^2(-2p)G_{22} \\ &= 2pq[p(G_{11} - G_{12}) + q(G_{12} - G_{22})] \\ &= 2pq\alpha \quad (\text{from Box 8.6})\end{aligned}$$

and

$$\begin{aligned}\sigma_{c_1}^2 &= p^2(2q)^2 + 2pq(q-p)^2 + q^2(-2p)^2 \\ &= 2pq[p^2 + 2pq + q^2] \\ &= 2pq \quad (\text{from Box 8.6})\end{aligned}$$

Hence, the regression coefficient

$$\alpha_1 = \frac{\text{cov}(X, Y)}{\sigma_X^2} = \frac{\text{cov}(c_1, G)}{\sigma_{c_1}^2} = \frac{2pq\alpha}{2pq} = \alpha$$

and the additive variance for the A locus is

$$\begin{aligned}\sigma_A^2 &= \frac{[\text{cov}(X, Y)]^2}{\sigma_X^2} = \frac{[\text{cov}(c_1, G)]^2}{\sigma_{c_1}^2} = \frac{\text{cov}(c_1, G)[\text{cov}(c_1, G)]}{\sigma_{c_1}^2} \\ &= \alpha \text{cov}(c_1, G) \\ &= \frac{2pq\alpha(2pq\alpha)}{2pq} \\ &= 2pq\alpha^2 \quad [\text{equation (8.69)}]\end{aligned}$$

Again, as pointed out above, this is analogous to the familiar expression in statistics for the sum of squares due to a contrast ($r = 1$)

$$SS(Q) = \frac{Q^2}{r \sum c^2} = \frac{Q}{\sum c^2} Q = \alpha Q$$

where $Q = \text{cov}(c, Y)$.

Before we consider the orthogonal polynomial coefficients for the additive variance at the B locus, we must substitute $p = p_{A_1}$ and $q = p_{A_2}$ in all above expressions to identify the A locus. Then upon converting the orthogonal polynomial coefficients for the additive variance at the B locus, simply substitute $p = p_{B_1}$ and $q = p_{B_2}$ in all above expressions.

With this background, you should now be able to undertake the derivation of the additive-by-additive variance (8.163a). Let $p = p_{A_1}$, $q = p_{A_2}$, $r = p_{B_1}$, and $s = p_{B_2}$ to simplify the writing. Hint: In doing any of following algebra, one should *anticipate* each of the answers in the following question from the one-locus case (Box 8.6) and knowledge of the additive-by-additive variance itself (8.163a). It guides one in the simplification process in doing the algebra. Otherwise, one can get hopelessly bogged down!

- Calculate the orthogonal polynomial coefficients for the additive-by-additive variance by setting up the appropriate two-way table and multiplying the corresponding orthogonal polynomial coefficients for the additive variance of the A and B loci—just as one would for the linear-by-linear contrast in statistics.
- Calculate the covariance between those additive-by-additive orthogonal polynomial coefficients and the genotypic values (use the two subscript genotypic designations for the nine genotypes— G_{22} , G_{21} , \dots , G_{00}).
- Calculate the variance of the orthogonal polynomial coefficients.
- Calculate the additive-by-additive regression coefficient, $\alpha_{1,2}$.
- Derive the formula for the additive-by-additive variance (8.163a).

Exercise 8.15.

- a. In Example 8.5, p. 8.105, what is the male or female gametic genetic variance equal to?
- b. In that same Example 8.5, linkage equilibrium, $p_{A_1B_1} = p_{A_1B_2} = p_{A_2B_1} = p_{A_2B_2} = 0.25$, was assumed, as well as random mating. Suppose that linkage disequilibrium, $\Delta_{A_1B_1} = 0.10$, existed, instead of linkage equilibrium, so that $p_{A_1B_1} = p_{A_2B_2} = 0.35$ and $p_{A_1B_2} = p_{A_2B_1} = 0.15$. Independence between loci is also assumed. Assume all other conditions stated or implied in Example 8.5.
- What is the overall mean equal to? Has it changed? If so, why? What two conditions must exist simultaneously?
 - What is the gametic genetic variance equal to?

Exercise 8.16. Suppose a population with the following genotypic values and frequencies existed:

	B_1B_1	B_1B_2	B_2B_2	
A_1A_1	6 $\frac{1}{16}$	6 $\frac{2}{16}$	4 $\frac{1}{16}$	$5\frac{1}{2}$ $\frac{1}{4}$
A_1A_2	4 $\frac{2}{16}$	4 $\frac{4}{16}$	2 $\frac{2}{16}$	$3\frac{1}{2}$ $\frac{1}{2}$
A_2A_2	4 $\frac{1}{16}$	4 $\frac{2}{16}$	2 $\frac{1}{16}$	$3\frac{1}{2}$ $\frac{1}{4}$
	$4\frac{1}{2}$ $\frac{1}{4}$	$4\frac{1}{2}$ $\frac{1}{2}$	$2\frac{1}{2}$ $\frac{1}{4}$	4

Suppose that the trait is flowering time or maturity. Note that locus A is completely dominant for early maturity whereas locus B is completely dominant for late maturity or completely recessive for early maturity.

- Write the linear additive, genic factorial model for this situation. Include only those nonzero terms -- no other terms.
- Calculate the variance of each term in the model, using any method or procedure.
- How can the total genotypic variance be calculated from those variances in (b)? From a purely mathematical point of view, what idea or theorem permits that? State the idea in general. What *minimal* genetical condition(s) must exist to insure the application of the idea?
- If the population above represented the F_2 generation of the cross between an early maturing homozygote $A_2A_2B_2B_2$ and a late maturing homozygote $A_1A_1B_1B_1$, what minimal genetical conditions are required to insure the application of the idea? Contrast this with the above. Discuss.
- If both early and late selection lines were initiated, would the magnitude of the expected response be the same in both directions?
- In the early selection lines, both p_{A_2} and p_{B_2} would be expected to increase in frequency. Would the change in frequency for A_2 , the dominant allele, be greater or less than that for B_2 , the recessive allele? Hint: See D.S. Falconer and T.F.C. Mackay, 1996, Introduction to Quantitative Genetics, Fourth edition, Longman, p. 30.

Exercise 8.17.

- a. State the theorem for the variance of the linear combination of both uncorrelated and correlated random variables.
- b. Discuss the importance of the theorem in statistical genetics.
- c. Give at least eight equations which depend upon its use when the variables are uncorrelated and three equations which depend upon its use when the variables are correlated. For each equation list all underlying conditions or assumptions and specifically identify those assumptions which permit the application of the theorem for either uncorrelated or correlated variables.

Exercise 8.18. Of what value is the genic factorial model in statistical genetics? What are the three most important assumptions underlying the model?

Exercise 8.19. Discuss the question "Is heredity or environment more important?" in the context of things that you might have learned in this course. Discuss the question as though you were explaining it to a nongeneticist. Do not forget to discuss those topics which maybe are self-evident to you. Do not write more than one page!