

EXERCISES FOR CHAPTER 2

Exercise 2.1. Define the word population as you think it should be defined and discuss your understanding of the use of that word in population genetics. Do you think the term is equivalent to a sample as the term is used in statistics? Would the term ever be equivalent to a population as used in statistics? If so, when?

Exercise 2.2. Suppose that there are n loci with m_1, m_2, \dots, m_n alleles at each locus.

a. Derive a general formula for calculating the number of different genotypes with distinctly different breeding behaviors in a population if two loci, say 1 and 2, have a recombination value less than one-half ($\rho_{1\dots} < 0.5$) and all other pairs of loci are independent. In considering the number of different genotypes with different breeding behaviors, coupling and repulsion phases for loci 1 and 2 must be considered different.

b. Derive a general formula for calculating the number of different genotypes with distinctly different breeding behaviors in a population if two loci, say 1 and 2, have a recombination value less than one-half and one of the two loci, say 2, and another locus, say 3, also have a recombination value less than one-half. All other pairs of loci are independent including loci 1 and 3.

c. Extend the analysis in part (b) by assuming the additional condition of linkage between loci 1 and 3. How would the general formula derived in part (b) be extended? Assuming $n=3$ (all loci linked—none independent), $m_1 = 3, m_2 = 4, m_3 = 5$, check that the number of genotypes is equal to that given in equation (2.13).

Exercise 2.3. One occasionally hears from geneticists that if two genes are located 50 or more centimorgans apart on a linkage map, the genes at the two loci would assort independently and a recombination value of independence would be observed. Criticize and discuss the statement.

Exercise 2.4. A cross of A_1B_2/A_2B_1 with A_2B_2/A_2B_2 produced 60 nonrecombinant and 40 recombinant offspring. Is there evidence of linkage? Perform a χ^2 test. If so, estimate the recombination value and its standard error. (After D. L. Hartl, 1980, Principles of Population Genetics. Chapter 1, Problem 6, p. 68.)

Exercise 2.5. Suppose that a female *Drosophila* fruit fly with the genotype $\frac{+ cv +}{ec + ct}$ was crossed to *ec cv ct* males (all genes are sex-linked and are designated *ec* for echinus, *cv* for crossveinless, and *ct* for cut), and the following results were obtained (data from E.W. Sinnott, L.C. Dunn, and Th. Dobzhansky, 1950, Principles of Genetics, McGraw-Hill, p. 218):

Parental	crossveinless (+ <i>cv</i> +)	2207	} 0.814592
combinations	echinus cut (<i>ec</i> + <i>ct</i>)	2125	
Crossovers between <i>ec</i> and <i>cv</i> only	echinus crossveinless (<i>ec cv</i> +)	273	} 0.101166
	cut (+ + <i>ct</i>)	265	
Crossovers between <i>cv</i> and <i>ct</i> only	echinus (<i>ec</i> + +)	217	} 0.082738
	crossveinless cut (+ <i>cv ct</i>)	223	
Double crossovers	wild type (+ + +)	5	} 0.001504
	echinus crossveinless cut (<i>ec cv ct</i>)	3	
		5,318	

Find estimates of the three parameters, $\rho_{1..}$, $\rho_{.1}$, and $\rho_{\overline{11}}$, and the amount of interference as discussed on pp. 2.36 to 2.39 and defined on p. 2.71. Contrast this measure with the usual measure of interference in general genetics.

Exercise 2.6. For three linked loci ρ_{11} denotes the probability of recombination in both the first and second segments. For closely linked loci it is known that $\rho_{11} < \rho_{1.}\rho_{.1}$. This is true because if recombination has occurred in the first segment, recombination is less likely to occur in the second segment. Would the probability of no recombination in both segments ρ_{00} be greater or less than $\rho_{0.}\rho_{.0}$?

Exercise 2.7. Suppose that for an individual with genotype $A_0B_0C_0D_0/A_1B_1C_1D_1$ the true recombination values are 0.04 between loci *A* and *B*, 0.10 between *B* and *C*, and 0.16 between *C* and *D*. Under the assumption of independence of recombination values between all loci (no interference), calculate the expected gametic frequencies for the eight complementary gametic groups and 16 different gametes. Also compute the values of the seven parameters required to describe the gametic output from the above individual.

Exercise 2.8.

- What does $\rho_{\overline{111}}$ denote?
- What is it equal to in terms of complementary gametic group probabilities?
- What would it be equal to if there were no interference between segments?
- What would (c) be equal to if $\rho_{1..} = 0.05$, $\rho_{.1.} = 0.11$, and $\rho_{.1} = 0.17$?

Exercise 2.9. First, write the recombination parameter $\rho_{\overline{1..111}}$ (seven segments or eight loci) in terms of *total* complementary gametic group probabilities, and then in terms of *elementary* complementary gametic group probabilities. Do not bother to enumerate all of them, but clearly illustrate the pattern. How many elementary complementary gametic group probabilities are there?

Exercise 2.10.

- What is the meaning of $\rho_{.11.1}$?
- What is the meaning of the same expression with a single bar over all of the ones, $\overline{\rho_{.11.1}}$? Explain clearly what $\overline{\rho_{.11.1}}$ means in terms of number of crossovers in each of the three segments individually and collectively.

Exercise 2.11.

- State the rules for writing one-segment, two-segment, three-segment, four-segment and five-segment recombination parameters themselves for n loci. Secondly, set forth the rules for defining those same kinds of recombination parameters for n loci in terms of a certain minimum number of total complementary gametic group probabilities. Give one example in each category for $n = 7$. State the total number of possible recombination parameters in each category. Also state, in general, how these total complementary gametic group probabilities may be expressed in terms of elementary complementary gametic group probabilities.
- State the rule for writing any k -segment recombination parameter itself for n loci, and further generalize the above rules by giving a single rule for defining any recombination parameter involving k segments as a sum of the smallest number of total complementary gametic group probabilities. Write the rules sufficiently clearly, so that anyone could write any recombination parameter itself and its equivalent in terms of total complementary gametic group probabilities by following your rules.

Exercise 2.12. The number of recombination parameters required to specify the gametic frequencies for two, three, and four linked loci are one, three, and seven, respectively. What would be the number for five linked loci? Briefly outline or describe what each would be. Give the appropriate symbol for each.

Exercise 2.13.

- How many recombination parameters are required to specify the gametic frequencies for six linked loci? Briefly outline or describe what each would be. Give the appropriate symbol for each.
- Express the parameters $\rho_{1\dots}$ and $\overline{\rho_{.11.1}}$ as the sum of the probabilities of both the total complementary gametic group probabilities (the minimum number) and the elementary complementary gametic groups. Also express only $\rho_{1\dots}$ as a function of gametic frequencies.
- Then express the linkage values λ_{ij} (in locus notation) = $\lambda_{1\dots} = 1 - 2\rho_{1\dots}$ and $\lambda_{jlmn} = \overline{\lambda_{.11.1}} = 1 - 2\overline{\rho_{.11.1}}$ as functions of gametic frequencies by use of formula (2.72G) (see Table 2.2A, p. 2.56). For your convenience the 64 possible gametes and the corresponding subsets (P) with respect to paternal genes (0 's) are listed below. Opposite each kind of gamete, indicate what $c(M * P) = c(ij \cap P)$ and $c(M * P) = c(jlmn \cap P)$ is for each of the two linkage values and what the sign would be for each corresponding gametic frequency. For only $\lambda_{ij} = \lambda_{i\dots}$, number the negative gametes from 1 to 32 and identify each of the same gametes in (b) above by writing the corresponding number (1 to 32) adjacent to it. Note that they are the same set of gametes.

Kind of gamete	P	$c(ij \cap P)$	$c(jlmn \cap P)$	Kind of gamete	P	$c(ij \cap P)$	$c(jlmn \cap P)$
(a)	{ 111111 }	{ }					
(b)	011111	i		(e)	000011	$ijkl$	
	101111	j			000101	$ijkm$	
	110111	k			000110	$ijkn$	
	111011	l			001001	$ijlm$	
	111101	m			001010	$ijln$	
	111110	n			001100	$ijmn$	
(c)	001111	ij			010001	$iklm$	
	010111	ik			010010	$ikln$	
	011011	il			010100	$ikmn$	
	011101	im			011000	$ilmn$	
	011110	in			100001	$jklm$	
	100111	jk			100010	$jkln$	
	101011	jl			100100	$jkmn$	
	101101	jm			101000	$jlmn$	
	101110	jn			110000	$klmn$	
	110011	kl		(f)	000001	$ijklm$	
	110101	km			000010	$ijkln$	
	110110	kn			000100	$ijkmn$	
	111001	lm			001000	$ijlmn$	
	111010	ln			010000	$iklmn$	
111100	mn		100000		$jklmn$		
(d)	000111	ijk		(g)	{ 000000 }	$ijklmn$	
	001011	ijl					
	001101	ijm					
	001110	ijn					
	010011	ikl					
	010101	ikm					
	010110	ikn					
	011001	ilm					
	011010	iln					
	011100	imn					
	100011	jkl					
	100101	jkm					
	100110	jkn					
	101001	jlm					
	101010	jln					
	101100	jmn					
	110001	klm					
	110010	kln					
110100	kmn						
111000	lmn						

$$(a) \binom{6}{0} = 1, (b) \binom{6}{1} = 6, (c) \binom{6}{2} = 15, (d) \binom{6}{3} = 20, (e) \binom{6}{4} = 15, (f) \binom{6}{5} = 6, (g) \binom{6}{6} = 1$$

Exercise 2.14. What is one of the advantages of using the linkage value λ instead of the recombination value ρ when expressing each as a function of the genetic frequencies?

Exercise 2.15. Distinguish between the deterministic and the stochastic models in population genetics.

Exercise 2.16.

a. Show that equation (2.95), $\text{cov}(X, Y) = E(XY) - \mu_X \mu_Y$, and equation (2.95a),

$\text{cov}(X, Y) = E[XE(Y|X)] - \mu_X \mu_Y$ [or equation (2.95b), $\text{cov}(X, Y) = E[E(X|Y)Y] - \mu_X \mu_Y$] are equivalent ways of calculating the covariance by applying (2.95b) to (2.89).

b. Since $\text{cov}(X, Y) = E(XY) - \mu_X \mu_Y = E[XE(Y|X)] - \mu_X \mu_Y$, it must likewise be true that in linear or straight-line regression $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum(X_i - \bar{X})(\hat{Y}_i - \bar{Y})$, where \hat{Y} equals the point on the regression line. Prove algebraically that $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum(X_i - \bar{X})(\hat{Y}_i - \bar{Y})$ where

$$\hat{Y}_i = a + bX_i = \bar{Y} + b(X_i - \bar{X}), \quad b = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}, \quad \text{and} \quad \sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum(X_i - \bar{X})Y_i$$

Exercise 2.17. Let us assume three linked loci with $\rho_{1,2} = 0.30$ and $\rho_{2,3} = 0.40$.

a. Find ρ_{13}^- under the assumption of no interference.

b. Derive a general formula for ρ_{13}^- in terms of the coefficient of coincidence $c = \frac{\rho_{12}\rho_{23}}{\rho_{13}}$ and in terms of interference $i = 1 - c$.

c. Find ρ_{13}^- under the assumption that the coefficient of coincidence is 0.90 or the interference is 0.10.

What is the general effect upon the recombination parameter as the interference increases?

Exercise 2.18. In equation (2.104), I assumed that $\mu_d = 0$, but I did not prove it. Prove that the mean of all deviations of genotypic frequencies for nonrandom mating from the genotypic frequencies for random mating equals zero. We assume a single, multiple allelic locus.