

SOLUTIONS TO EXERCISES FOR CHAPTER 2

**Exercise 2.1.**

In population genetics the term population is often used to simply define a collection of a finite or infinite group of individuals, often interbreeding, of the same species. In this case, no numerical variable or attribute is identified; the term population is simply a description of the individuals involved. On the other hand, if a variable or attribute is identified, then the term population is most commonly used as it is in statistics. In statistics the term population refers to the group of all possible values, infinite or finite in size, and their attributes such as their mean and variance about which we want to draw conclusions about their possible values by drawing a representative sample from the population. The population is most commonly infinite in size. For example, the cross between two populations or the  $F_2$  resulting from the cross between two inbred lines is conceptually infinite in size. However, in other cases of an evolving population the population may be finite in size. Occasionally the term population is used incorrectly to refer to a sample from a population. See my notes, Chapter 2, pp. 1 and 2, where I attempt to define the word population. D. L. Hartl, 1980, Principles of Population Genetics, Sinauer Associates, p. 62, gives a good discussion of the difficulty of defining a population.

**Exercise 2.2.**

a. The general formula for the total number of genotypes for some loci being independent and others linked is given in equation (2.13b). Its general derivation is given in Chapter 2, pp. 2.6 to 2.10, and combines equations (2.11) for independent loci and (2.13) for all loci linked on a single chromosome. The details of the derivation will not be repeated here. The general equation for two linked loci and independence between all other pairs of loci can be written from equation (2.13b) as

$$\left( \begin{array}{l} \text{total} \\ \text{number of} \\ \text{genotypes} \end{array} \right) = \prod_{i=1}^s \left[ \frac{\left( \prod_{j=1}^{n_i} m_{ij} \right) \left( \prod_{j=1}^{n_i} m_{ij} + 1 \right)}{2} \right]$$

$$= \frac{m_{11}m_{12}(m_{11}m_{12} + 1)}{2} \prod_{i=2}^s \frac{m_i(m_i + 1)}{2}$$

where the two linked loci are assumed to be on chromosome 1 and each of the other loci are assumed to be on separate chromosomes.

b. From a strict theoretical viewpoint, the question itself is in error. The basic concept of loci 1 and 3 being independent, when loci 1 and 2, and loci 2 and 3 are linked, can be said to be in error. This can be shown by considering equation (2.56) which assumes no interference between adjacent segments. Equation (2.56a) shows that even if each of two adjacent segments each exhibit a recombination value 0.40 ( $\rho_{1.} = \rho_{.1} = 0.40$ ), the recombination value for the larger segment is 0.48—not 0.50. It is only as  $\rho_{1.} \rightarrow 0.5$  and  $\rho_{.1} \rightarrow 0.5$  that  $\rho_{\overline{11}} \rightarrow 0.5$  when no interference exists. Hence, we see the fallacy of the question itself from a theoretical point of view.

However, from a practical point-of-view one may choose to ignore the slight difference in breeding behavior between coupling and repulsion heterozygotes for loci that are widely separated on the same chromosome, for their breeding behaviors are not that different. The answer to this question per se is not discussed in my notes. The total number of genotypes is equal to the total number of genotypes with all loci independent plus the number of additional genotypes due to difference in phase for every pair of linked loci. For

every double heterozygote there are two kinds of heterozygotes, coupling  $A_1B_1/A_2B_2$  and repulsion  $A_1B_2/A_2B_1$ , so this is the reason for calculating the number of double heterozygotes and increasing the total number of genotypes by that number. Thus,

$$\begin{aligned} \left( \begin{array}{l} \text{total number} \\ \text{of genotypes} \end{array} \right) &= \left( \begin{array}{l} \text{number of genotypes} \\ \text{with independence} \end{array} \right) + \left( \begin{array}{l} \text{number of additional genotypes due} \\ \text{to difference in phase at loci 1 and 2} \end{array} \right) \\ &\quad + \left( \begin{array}{l} \text{number of additional genotypes due} \\ \text{to difference in phase at loci 2 and 3} \end{array} \right) \\ &= \prod_{i=1}^n \left[ \frac{m_i(m_i+1)}{2} \right] + \prod_{i=1}^2 \left[ \frac{m_i(m_i-1)}{2} \right] \prod_{i=3}^n \left[ \frac{m_i(m_i+1)}{2} \right] \\ &\quad + \prod_{i=2}^3 \left[ \frac{m_i(m_i-1)}{2} \right] \prod_{i=1, i \neq 2,3}^n \left[ \frac{m_i(m_i+1)}{2} \right] \end{aligned}$$

For the values of the number of loci and number of alleles in the example enumerated in part (c) below we substitute those values in the above expression to obtain the total number of genotypes:

$$\begin{aligned} \left( \begin{array}{l} \text{total number} \\ \text{of genotypes} \end{array} \right) &= \left( \frac{3(3+1)}{2} \right) \left( \frac{4(4+1)}{2} \right) \left( \frac{5(5+1)}{2} \right) + \left( \frac{3(3-1)}{2} \right) \left( \frac{4(4-1)}{2} \right) \left( \frac{5(5+1)}{2} \right) + \left( \frac{4(4-1)}{2} \right) \left( \frac{5(5-1)}{2} \right) \left( \frac{3(3+1)}{2} \right) \\ &= 6 \cdot 10 \cdot 15 + 3 \cdot 6 \cdot 15 + 6 \cdot 10 \cdot 6 \\ &= 900 + 270 + 360 \\ &= 1530 \end{aligned}$$

c. If we were to consider the phase for three linked loci, we would need to calculate the additional number of double heterozygotes for the additional pair of linked loci 1 and 3. The additional term for considering the phase of the double heterozygotes for loci 1 and 3 is

$$\left( \begin{array}{l} \text{number of additional genotypes due} \\ \text{to difference in phase at loci 1 and 3} \end{array} \right) = \prod_{i=1, i \neq 2}^3 \left[ \frac{m_i(m_i-1)}{2} \right] \prod_{i=2, i \neq 3}^n \left[ \frac{m_i(m_i+1)}{2} \right]$$

Then we add that quantity to the three major terms in part (b) above. Thus, we are effectively doubling the number of double heterozygotes for loci 1 and 3. The sum of the four major terms is equal to that obtained from equation (2.13b) for one chromosome with three linked loci and all other loci independent. One does not have to deal separately with the triple heterozygotes. The four different phases of the triple heterozygotes are embodied in the three different two-locus combinations.

We give a simple example for three loci. Let  $n = 3$ ,  $m_1 = 3$ ,  $m_2 = 4$ ,  $m_3 = 5$ . Hence, we have

<u>Locus</u>	<u>Number of alleles</u>	<u>Number of homozygotes</u>	<u>Number of heterozygotes</u>	<u>Number of genotypes</u>
1	3	3	3	6
2	4	4	6	10
3	5	5	10	15

From a tree diagram shown below of the number of homozygotes and heterozygotes for each locus, we calculate the number of genotypes for each combination of homozygous and heterozygous states at loci 1, 2, and 3. The total of 900 equals the total number of genotypes under independence.

			Additional number of genotypes					
			Case 4:					
			Case 1:	Case 2:	Case 3:	1 and 2	Case 5:	
			Only	Only	Only	and	All	
			1 and 2	2 and 3	1 and 3	2 and 3	pairs	
			linked	linked	linked	linked	linked	
1	Locus 2	3	Branch number	All loci indep				
		hom 5	1	60				
	hom 4	het 10	2	120				
hom 3		hom 5	3	90				
	het 6	het 10	4	180		180	180	180
		hom 5	5	60				
	hom 4	het 10	6	120			120	120
het 3		hom 5	7	90	90		90	90
	het 6	het 10	8	180	180	180	180	360
				900	270	360	300	630
				930				930

It is observed that for each case with only two loci linked the total number of genotypes with different breeding behaviors is equal to the total number of genotypes with independence plus the additional number of double heterozygotes at the corresponding loci. That additional number includes those double heterozygotes per se which are homozygous for the other locus and also those triple heterozygotes which are the same double heterozygotes but are also heterozygous for the other locus.

In our numerical example of only three loci which are all linked, the total number of genotypes (900+ 930 =1,830) must be equal to that given by equation (2.13), namely,

$$\frac{(3 \cdot 4 \cdot 5)(3 \cdot 4 \cdot 5 + 1)}{2} = 30(31) = 1,830$$

### Exercise 2.3.

The statement is not quite true, for if two loci are 50 centimorgans apart, the recombination value is less than 0.50 because of the occurrence of double crossovers. Centimorgans are directly proportional to recombination values for values less than 0.10 to 0.15, i.e., two loci separated by 10 centimorgans or map units would show a recombination value of 0.10, but loci separated by 50 centimorgans would show a recombination value less than 0.50. This can be deduced as follows by considering three ordered loci,  $A$ ,  $B$ , and  $C$ : Recombination in the overall or  $AC$  segment is equal to the probability of crossing over in only the first or  $AB$  segment plus the probability of crossing over in only the second or  $BC$  segment, but in terms of recombination values it is equal to the recombination value in the first or  $AB$  segment plus the recombination value in the second or  $BC$  segment minus two times the probability of crossing over in both of the two segments, i.e., using equation (2.56)

$$\begin{aligned} \rho_{11} &= \rho_{10} + \rho_{01} = \rho_1 + \rho_1 - 2\rho_{11} && \text{substituting equation (2.100A)} \\ &= \rho_1 + \rho_1 - 2c\rho_1\rho_1 \end{aligned}$$

For recombination in the overall segment to be equal to the recombination in the first segment plus recombination in the second segment, the last term must be equal to zero which would occur only if the coefficient of coincidence  $c$  were equal to zero, i.e., complete interference  $i = 1 - c = 1 - 0 = 1$  existed, so no double crossingover occurred. However, we know that double crossingover does occur so recombination in the larger segment is not equal to the sum of the recombination values or map units of the smaller segments.

For example, suppose that one has two equal segments each with recombination of  $\rho_{1.} = \rho_{.1} = 0.25$  and that no double crossingover occurs within each segment, so that the map distance of each segment is 25 map units. Then the recombination in the larger segment would be, assuming no interference between segments.

$$\begin{aligned}\rho_{11} &= \rho_{1.} + \rho_{.1} - 2\rho_{1.}\rho_{.1} \\ &= 0.25 + 0.25 - 2(0.25)(0.25) \\ &= 0.50 - 2(0.0625) \\ &= 0.3750\end{aligned}$$

Of course, interference between the two segments does exist, increasing the recombination value somewhat for the larger segment, but the recombination value would not be equal to 0.50, unless complete interference between the two segments existed.

#### Exercise 2.4.

	Nonrecombination	Recombination	
Obs.	60	40	
Exp.	50	50	
$\chi^2$	$\frac{(60-50)^2}{50}$	$\frac{(40-50)^2}{50}$	
	= 2	= 2	= 4.00 > 3.84

reject null hypothesis of no linkage at  $\alpha = 0.05$

The estimated recombination value is

$$\hat{\rho}_1 = \frac{40}{100} = 0.4$$

The standard error of the estimated recombination value is

$$\hat{\sigma}_{\hat{\rho}_1} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.6)(0.4)}{100}} = \sqrt{0.0024} = 0.0490 \quad (\text{see Mather, K., 1951, The Measurement of Linkage in Heredity})$$

#### Exercise 2.5.

First, we label the information, using the notation of Chapter 2 (we place hats on the parameters to denote the corresponding estimate or statistic):

		Recombination in segment <i>cv-ct</i>		
		No	Yes	
Recombination in segment <i>ec-cv</i>	No	$\hat{\rho}_{00}$ = 0.814592	$\hat{\rho}_{01}$ = 0.082738	$\hat{\rho}_{0.} = 0.897330$
	Yes	$\hat{\rho}_{10}$ = 0.101166	$\hat{\rho}_{11}$ = 0.001504	$\hat{\rho}_{.1} = 0.102670$
		$\hat{\rho}_{.0}$ = 0.915758	$\hat{\rho}_{.1}$ = 0.084242	1.000000

Estimates of the recombination parameters are [see equation (2.53)]

$$\hat{\rho}_1 = \hat{\rho}_{10} + \hat{\rho}_{11} = 0.101166 + 0.001504 = 0.102670$$

$$\hat{\rho}_{.1} = \hat{\rho}_{01} + \hat{\rho}_{11} = 0.082738 + 0.001504 = 0.084242$$

$$\hat{\rho}_{11} = \hat{\rho}_{01} + \hat{\rho}_{10} = 0.082738 + 0.101166 = 0.183904$$

Interference is discussed on pp. 2.36 to 2.39. It is defined herein differently from that in general genetics as a deviation from independence of recombination in the two segments [ see equation (2.100)]. The absolute value of the measure of interference can be calculated from any one of the four cells [see equation (2.100)].

Choosing the  $\rho_{11}$  cell, we obtain:

$$\text{interference} = \hat{\rho}_{11} - \hat{\rho}_1 \hat{\rho}_{.1} = 0.001504 - 0.102670(0.084242) = 0.001504 - 0.0086491 = -0.00714512614$$

Here we use the difference between the observed frequency of double recombinants and its expected frequency, assuming independence.

However, in general genetics we commonly take the ratio of the same two quantities. It is called the coefficient of coincidence, namely,

$$\hat{\rho}_{11} / \hat{\rho}_1 \hat{\rho}_{.1} = 0.001504 / 0.102670(0.084242) = 0.001504 / 0.008649 = 0.173893 .$$

In the practice of general genetics, the coefficient of coincidence is calculated equivalently as the ratio of the number of observed double recombinants to the expected number of double recombinants, assuming independence, namely,

$$8 / [0.102670(0.084242)5,318] = 0.173927 \text{ (within rounding errors of } 0.173893 \text{)}. \text{ Then interference is defined as one minus the coefficient of coincidence, namely [see equation (2.100B)],}$$

$$i = 1 - c = 1 - 0.173893 = 0.826107$$

Interference increases up to one as the frequency of double recombinants decreases to zero.

### Exercise 2.6.

The probability of no recombination in both segments would also be less than the product of no recombination in the two adjacent segments., i.e.,  $\rho_{00} < \rho_{0.}\rho_{.0}$  . If one cell on the main diagonal of  $2 \times 2$  table is less than the product of their marginal probabilities, the other cell on the main diagonal must also be less than their marginal probabilities. Both off-diagonal cells must increase so that the marginal probabilities remain the same. See equation (2.100).

### Exercise 2.7.

The frequencies of the eight complementary gametic groups under the assumption of no interference are [see equation (2.55) for three loci]

$$\rho_{000} = \rho_{0.}\rho_{.0}\rho_{..0} = (0.96)(0.90)(0.84) = 0.72576$$

$$\rho_{001} = \rho_{0.}\rho_{.0}\rho_{..1} = (0.96)(0.90)(0.16) = 0.13824$$

$$\rho_{010} = \rho_{0.}\rho_{.1}\rho_{..0} = (0.96)(0.10)(0.84) = 0.08064$$

$$\rho_{011} = \rho_{0.}\rho_{.1}\rho_{..1} = (0.96)(0.10)(0.16) = 0.01536$$

$$\rho_{100} = \rho_{1.}\rho_{.0}\rho_{..0} = (0.04)(0.90)(0.84) = 0.03024$$

$$\rho_{101} = \rho_{1.}\rho_{.0}\rho_{..1} = (0.04)(0.90)(0.16) = 0.00576$$

$$\rho_{110} = \rho_{1.}\rho_{.1}\rho_{..0} = (0.04)(0.10)(0.84) = 0.00336$$

$$\rho_{111} = \rho_{1.}\rho_{.1}\rho_{..1} = (0.04)(0.10)(0.16) = 0.00064$$

$$\text{Total } 1.00000$$

The frequencies of the 16 different kinds of gametes are [see equation (2.47) for three loci]

$A_0B_0C_0D_0$	$p_{0000} = 0.36288$	$A_0B_1C_1D_1$	$p_{0111} = 0.01512$
$A_1B_1C_1D_1$	$p_{1111} = 0.36288$	$A_1B_0C_0D_0$	$p_{1000} = 0.01512$
$A_0B_0C_0D_1$	$p_{0001} = 0.06912$	$A_0B_1C_1D_0$	$p_{0110} = 0.00288$
$A_1B_1C_1D_0$	$p_{1110} = 0.06912$	$A_1B_0C_0D_1$	$p_{1001} = 0.00288$
$A_0B_0C_1D_1$	$p_{0011} = 0.04032$	$A_1B_0C_1D_1$	$p_{1011} = 0.00168$
$A_1B_1C_0D_0$	$p_{1100} = 0.04032$	$A_0B_1C_0D_0$	$p_{0100} = 0.00168$
$A_0B_0C_1D_0$	$p_{0010} = 0.00768$	$A_0B_1C_0D_1$	$p_{0101} = 0.00032$
$A_1B_1C_0D_1$	$p_{1101} = 0.00768$	$A_1B_0C_1D_0$	$p_{1010} = 0.00032$

The values of the recombination parameters are [see equation (2.67)]

$$\rho_{1..} = \rho_{100} + \rho_{101} + \rho_{110} + \rho_{111} = 0.03024 + 0.00576 + 0.00336 + 0.00064 = 0.04000$$

$$\rho_{.1.} = \rho_{010} + \rho_{011} + \rho_{110} + \rho_{111} = 0.08064 + 0.01536 + 0.00336 + 0.00064 = 0.10000$$

$$\rho_{..1} = \rho_{001} + \rho_{011} + \rho_{101} + \rho_{111} = 0.13824 + 0.01536 + 0.00576 + 0.00064 = 0.16000$$

$$\rho_{\overline{11.}} = \rho_{010} + \rho_{011} + \rho_{100} + \rho_{101} = 0.08064 + 0.01536 + 0.03024 + 0.00576 = 0.13200$$

$$\rho_{\overline{1.1}} = \rho_{100} + \rho_{110} + \rho_{001} + \rho_{011} = \quad \text{and so on} \quad = 0.18720$$

$$\rho_{\overline{.11}} = \rho_{010} + \rho_{110} + \rho_{001} + \rho_{101} = \quad = 0.22800$$

$$\rho_{\overline{111}} = \rho_{100} + \rho_{010} + \rho_{001} + \rho_{111} = \quad = 0.24976$$

The numerical values are obtained by substituting in the above values for the complementary gametic groups.

### Exercise 2.8.

a.  $\rho_{\overline{111}}$  denotes the recombination in the compound segment consisting of three adjacent segments (see p. 2.43).

b. In terms of complementary gametic group probabilities, it is equal to

$$\rho_{\overline{111}} = \rho_{100} + \rho_{010} + \rho_{001} + \rho_{111} \quad [\text{see equation (2.67)}]$$

c. If there was no interference, it would be equal to

$$\rho_{\overline{111}} = \rho_{1..}\rho_{.0.}\rho_{..0} + \rho_{0..}\rho_{.1.}\rho_{..0} + \rho_{0..}\rho_{.0.}\rho_{..1} + \rho_{1..}\rho_{.1.}\rho_{..1} \quad [\text{see equation (2.55) for analogy with three loci instead of four loci}]$$

d. If recombination in the three separate segments were  $\rho_{1..} = 0.05$ ,  $\rho_{.1.} = 0.11$ , and  $\rho_{..1} = 0.17$ , then recombination in the overall compound segment would be

$$\begin{aligned} \rho_{\overline{111}} &= 0.05(0.89)(0.83) + 0.95(0.11)(0.83) + 0.95(0.89)(0.17) + 0.05(0.11)(0.17) \\ &= 0.036935 + 0.086735 + 0.143735 + 0.000935 \\ &= 0.26834 \end{aligned}$$

### Exercise 2.9.

The recombination parameter in terms of total complementary gametic group probabilities is

$$\begin{aligned} \rho_{\overline{1..111}} &= \rho_{.1..000} + \rho_{.0..100} + \rho_{.0..010} + \rho_{.0..001} \\ &\quad + \rho_{.1..110} + \rho_{.1..101} + \rho_{.1..011} + \rho_{.0..111} \quad [\text{see equation (2.71b)}] \end{aligned}$$

where the first total complementary gametic group probability  $\rho_{.1..000}$  is equal to the sum of the following eight

$(2^{n-k-1} = 2^{8-4-1} = 2^3 = 8)$  elementary complementary gametic group probabilities with the sequences of 000, 001, 010, 011, 100, 101, 110, and 111 replacing the dots in the successive total complementary gametic group probabilities:

$$\begin{aligned} \rho_{.1..000} &= \rho_{0100000} + \rho_{0101000} + \rho_{0110000} + \rho_{0111000} \\ &\quad + \rho_{1100000} + \rho_{1101000} + \rho_{1110000} + \rho_{1111000} \end{aligned}$$

There are 64 elementary complementary gametic groups for this recombination parameter, namely,

$$2^{k-1}2^{n-k-1} = 2^{4-1}2^{8-4-1} = 8 \cdot 8 = 64 \quad (\text{see p. 2.51}).$$

### Exercise 2.10.

a.  $\rho_{.11.1.}$  is a total complementary gametic group probability involving seven loci  $A, B, C, D, E, F,$  and  $G$  or six segments [see equation (2.63) for a four-locus analogy]. It denotes the probability of recombination in the second ( $BC$ ), third ( $CD$ ), and fifth ( $EF$ ) segments, regardless of the presence or absence of recombination in the other segments.

b.  $\overline{\rho_{.11.1.}}$  is one of 63 ( $= 2^{n-1} - 1 = 2^{7-1} - 1$ ) [equation (2.72)] recombination parameters. It denotes the probability of recombination or observable exchange in the noncontiguous segment consisting of segments  $BD$  and  $EF$ . It is equal to

$$\overline{\rho_{.11.1.}} = \rho_{.10.0.} + \rho_{.01.0.} + \rho_{.00.1.} + \rho_{.11.1.} \quad [\text{see equations (2.67) and (2.71b)}]$$

This means that an observable exchange occurs when an odd number of crossovers occurs in any one of the three single segments  $BC, CD,$  and  $EF$  conjunctive with an even number of crossovers in the remaining three segments or an odd number of crossovers in all three segments  $BC, CD,$  and  $EF$ .

### Exercise 2.11.

a. One-segment recombination parameters for  $n$  loci:

Each one-segment recombination parameter may be written by placing a single subscript 1 in one of the  $n - 1$  subscript positions on  $\rho$  corresponding to the segment whose recombination value is being specified, and placing  $n - 2$  dots elsewhere.

Any one-segment recombination parameter is simply equal to one total complementary gametic group probability written exactly as the recombination parameter itself.

For example, for  $n = 7$

recombination parameter = minimum number of total complementary gametic group probabilities

$$\rho_{.1...} = \rho_{.1...}$$

The total number of one-segment recombination parameters is  $\binom{n-1}{1} = n - 1$ .

Two-segment recombination parameters for  $n$  loci:

Each two-segment recombination parameter may be written by writing two subscript 1s on  $\rho$  in the subscript positions corresponding to the two segments whose recombination value is being specified, and placing  $n - 3$  dots elsewhere. Then put a continuous bar above the two 1s to connect them (this distinguishes the recombination parameter itself from the corresponding total complementary gametic group probability).

Any two-segment recombination parameter is equal to the sum of *two* total complementary gametic group probabilities each of which has a one and a zero in the positions corresponding to the two segments, with dots elsewhere.

For example, for  $n = 7$

$$\overline{\rho_{..1.1.}} = \rho_{.1.1.0.} + \rho_{.0.1.1.}$$

The total number of two-segment recombination parameters is  $\binom{n-1}{2}$ .

Three-segment recombination parameters for  $n$  loci:

Write three subscript 1s on  $\rho$  in the subscript positions corresponding to the three segments whose recombination value is being specified and place  $n - 4$  dots elsewhere. Then put a continuous bar above the three 1s, connecting the two extreme ones.

Each of these three-segment recombination parameters is equal to the sum of *four* total complementary gametic group probabilities—three of which are composed of one 1 and two 0s (the three combinations of one 1 and two 0s in three positions), and the other composed of three 1s in the positions corresponding to the three segments concerned. Note that the number of 1s present in any total complementary gametic group probability is always an odd number, 1 or 3. The expression for the sum of the four total complementary gametic group probabilities can be written as follows: With respect to the subscripts on the recombination parameter itself hold the dots constant, and replace the 1s by all possible combinations of 1s and 0s, adding a term for the total complementary gametic group probability whenever the sum of the 1s is *odd*.

For example, for  $n = 7$

$$\overline{\rho_{.1.11.}} = \rho_{.1.00.} + \rho_{.0.10.} + \rho_{.0.01.} + \rho_{.1.11.}$$

The total number of three-segment recombination parameters is  $\binom{n-1}{3}$ .

Four-segment recombination parameters for  $n$  loci:

Again, write  $\rho$  with  $n - 1$  subscripts of which four are 1s corresponding to the four chromosomal segments composing the compound segment for which the amount of recombination is being defined, and  $n - 5$  are dots with a bar above connecting the two extreme 1s.

Each of these four-segment recombination parameters is equal to the sum of *eight* total complementary gametic group probabilities—four of which are composed of one 1 and three 0s (the four combinations of one 1 and three 0s in four positions), and four others composed of three 1s and one 0 in the positions corresponding to the four segments concerned. That is, with respect to the subscripts on the recombination parameter itself hold the dots constant, and replace the 1s by all possible combinations of 1s and 0s, adding a term for the total complementary gametic group probability whenever the sum of the 1s is *odd*.

For example, for  $n = 7$

$$\overline{\rho_{.1.111.}} = \rho_{.1.000.} + \rho_{.0.100.} + \rho_{.0.010.} + \rho_{.0.001.} + \rho_{.1.110.} + \rho_{.1.101.} + \rho_{.1.011.} + \rho_{.0.111.}$$

The total number of four-segment recombination parameters is  $\binom{n-1}{4}$ .

Five-segment recombination parameters for  $n$  loci:

Five-segment recombination parameters are written as  $\rho$  followed by  $n - 1$  subscripts of which five are 1s corresponding to the five chromosomal segments for which the parameter is defined, and  $n - 6$  are dots. Again a bar is placed above the 1s.

Each parameter may be expressed in terms of *sixteen* total complementary gametic group probabilities—five of which are composed of one 1 and four 0s [ $\binom{5}{1} = 5$  is the number of combinations of 5 things taken one at a time, or the number of ways that one 1 can be placed in 5 positions, or the number of combinations of 5 things of which one is of one kind (a single 1) and four are alike of another kind (0s)], 10 of which are composed of three 1s and two 0s [ $\binom{5}{3} = 10$  is the number of combinations of 5 things—3 of one kind and 2 of another kind], and finally one of which is composed of five 1s in the positions corresponding to the five segments concerned [ $\binom{5}{5} = 1$  is the number of combinations of 5 things taken 5 at a time]. That is, with respect to the subscripts on the recombination parameter itself hold the dots constant, and replace the 1s by all possible combinations of 1s and 0s, adding a term for the total complementary gametic group probability whenever the sum of the 1s is *odd*.

For example, for  $n = 7$

$$\begin{aligned} \overline{\rho_{1.1111}} &= \rho_{1.0000} + \rho_{0.1000} + \rho_{0.0100} + \rho_{0.0010} + \rho_{0.0001} && \left. \vphantom{\overline{\rho_{1.1111}}} \right\} \binom{5}{1} = 5 \\ &+ \rho_{1.1100} + \rho_{1.1010} + \rho_{1.1001} + \rho_{1.0110} + \rho_{1.0101} + \rho_{1.0011} && \left. \vphantom{\overline{\rho_{1.1111}}} \right\} \binom{5}{3} = 10 \\ &+ \rho_{0.1110} + \rho_{0.1101} + \rho_{0.1011} + \rho_{0.0111} && \\ &+ \rho_{1.1111} && \left. \vphantom{\overline{\rho_{1.1111}}} \right\} \binom{5}{5} = 1 \end{aligned}$$

The total number of five-segment recombination parameters for  $n$  loci is  $\binom{n-1}{5}$ .

Any total complementary gametic group probability can be expressed as a sum of the elementary complementary gametic group probabilities (which is any  $\rho$  followed by 0s and 1s only—no dots) by holding the 0s and 1s constant in the total complementary gametic group probability and replacing all dots by all possible combinations of 0s and 1s.

b. The  $k$ -segment recombination parameters for  $n$  loci:

Any  $k$ -segment recombination parameter itself can be written as  $\rho$  followed by  $n - 1$  subscripts of which  $k$  are 1s in positions corresponding to the  $k$  chromosomal segments for which the parameter is defined, and  $n - k - 1$  are dots in the remaining positions. Then a continuous bar is placed above the two or more 1s, connecting the two extreme ones.

Any  $k$ -segment recombination parameter is equal to the sum of  $2^{k-1}$  total complementary gametic group probabilities. These probabilities are composed of all possible combinations consisting of an *odd* number of 1s (e.g., 1, 111, 11111, etc.) in the  $k$  positions and 0s in the remaining  $k$  positions with dots retained in all original positions. Alternatively, the  $k$ -segment recombination parameter is equal to the sum of  $\binom{k}{1} = k$  total complementary gametic group probabilities with one 1 and  $(k - 1)$  0s in the subscript positions corresponding to the  $k$  segments concerned with dots in the  $(n - k - 1)$  positions, plus  $\binom{k}{3}$  total complementary gametic group probabilities with three 1s and  $(k - 3)$  0s in the subscript positions corresponding to the  $k$  segments concerned with dots in the other  $n - k - 1$  other positions, plus  $\binom{k}{5}$  total complementary gametic group probabilities with five 1s and  $(k - 5)$  0s in the subscript positions corresponding to the  $k$ -segments concerned with dots in the  $(n - k - 1)$  other positions, and so on. That is, the minimum number of total complementary gametic group probabilities may be expressed as follows:

$$2^{k-1} = \sum_{i \in S} \binom{k}{i}$$

where  $S = \{1, 3, 5, 7, \dots, k\}$  for odd  $k$

$S = \{1, 3, 5, 7, \dots, k - 1\}$  for even  $k$

The above details for any  $k$ -segment recombination parameter are presented in tabular form in Chapter 2, p. 2.51, for  $k = 1, 2, \dots, n - 1$ .

**Exercise 2.12.**

Fifteen recombination parameters are required [see equation (2.72)]. We assume the order of genes A, B, C, D, and E in the following.

- $\rho_{1\dots}$  = recombination in segment AB
- $\rho_{.1\dots}$  = recombination in segment BC
- $\rho_{..1\dots}$  = recombination in segment CD
- $\rho_{\dots 1}$  = recombination in segment DE
- $\rho_{\overline{11}\dots}$  = recombination in segments AB and BC combined
- $\rho_{\overline{1.1}\dots}$  = recombination in segments AB and CD combined
- $\rho_{\overline{1..1}\dots}$  = recombination in segments AB and DE combined
- $\rho_{\overline{.11}\dots}$  = recombination in segments BC and CD combined
- $\rho_{\overline{.1.1}\dots}$  = recombination in segments BC and DE combined
- $\rho_{\overline{\dots 11}}$  = recombination in segments CD and DE combined
- $\rho_{\overline{111}\dots}$  = recombination in segment AB, BC, and CD combined
- $\rho_{\overline{11.1}\dots}$  = recombination in segment AB, BC, and DE combined
- $\rho_{\overline{1..11}\dots}$  = recombination in segment AB, CD, and DE combined
- $\rho_{\overline{.111}\dots}$  = recombination in segment BC, CD, and DE combined
- $\rho_{\overline{1111}\dots}$  = recombination in segment AB, BC, CD, and DE combined

**Exercise 2.13.**

a. The number of recombination parameters for six loci is  $2^{6-1} - 1 = 31$ . We assume that the order of the loci is A, B, C, D, E, F. The recombination parameters are: {The first five  $\left[\binom{5}{1} = 5\right]$  below are one-segment recombination parameters; the next ten  $\left[\binom{5}{2} = 10\right]$  are two-segment recombination parameters, the next ten  $\left[\binom{5}{3} = 10\right]$  are three-segment recombination parameters, the next five  $\left[\binom{5}{4} = 5\right]$  are four-segment recombination parameters, and the last one  $\left[\binom{5}{5} = 1\right]$  is a five-segment recombination parameter. See p. 2.48.}

The symbol for each recombination parameter and the segments associated with each parameter are presented:

- |                   |                             |   |  |
|-------------------|-----------------------------|---|--|
| $\rho_{1\dots}$   | recombination in segment AB | } | probability of an odd number of crossovers in a single segment,<br>ignoring other segments |
| $\rho_{.1\dots}$  | recombination in segment BC |   |  |
| $\rho_{..1\dots}$ | recombination in segment CD |   |  |
| $\rho_{\dots 1}$  | recombination in segment DE |   |  |
| $\rho_{\dots 1}$  | recombination in segment EF |   |  |

$\rho_{\overline{11...}}$	recombination in segment AB, BC or $AB \cup BC$ or AC	}	probability of an odd number of crossovers in segments composed of two smaller segments, ignoring other segments
$\rho_{\overline{1.1..}}$	recombination in segment AB, CD or $AB \cup CD$		
$\rho_{\overline{1..1.}}$	recombination in segment AB, DE or $AB \cup DE$		
$\rho_{\overline{1...1}}$	recombination in segment AB, EF or $AB \cup EF$		
$\rho_{\overline{.11..}}$	recombination in segment BC, CD or $BC \cup CD$ or BD		
$\rho_{\overline{.1.1.}}$	recombination in segment BC, DE or $BC \cup DE$		
$\rho_{\overline{.1..1}}$	recombination in segment BC, EF or $BC \cup EF$		
$\rho_{\overline{..11.}}$	recombination in segments CD and DE or $CD \cup DE$ or CE		
$\rho_{\overline{..1.1}}$	recombination in segments CD and EF or $CD \cup EF$		
$\rho_{\overline{...11}}$	recombination in segments DE and EF or $DE \cup EF$ or DF		

$\rho_{\overline{111..}}$	recombination in segments AB, BC, and CD or $AB \cup BC \cup CD$ or AD	}	probability of an odd number of crossovers in segments composed of three smaller segments, ignoring other segments
$\rho_{\overline{11.1.}}$	recombination in segment AB, BC, DE or $AB \cup BC \cup DE$		
$\rho_{\overline{11..1}}$	recombination in segment AB, BC, EF or $AB \cup BC \cup EF$		
$\rho_{\overline{1.11.}}$	recombination in segment AB, CD, DE or $AB \cup CD \cup DE$		
$\rho_{\overline{1.1.1}}$	recombination in segment AB, CD, EF or $AB \cup CD \cup EF$		
$\rho_{\overline{1..11}}$	recombination in segment AB, DE, EF or $AB \cup DE \cup EF$		
$\rho_{\overline{.111.}}$	recombination in segment BC, CD, DE or $BC \cup CD \cup DE$ or BE		
$\rho_{\overline{.11.1}}$	recombination in segment BC, CD, EF or $BC \cup CD \cup EF$		
$\rho_{\overline{.1.11}}$	recombination in segment BC, DE, EF or $BC \cup DE \cup EF$		
$\rho_{\overline{..111}}$	recombination in segment CD, DE, EF or $CD \cup DE \cup EF$ or CE		

$\rho_{\overline{1111.}}$	recombination in segment AB, BC, CD, DE or $AB \cup BC \cup CD \cup DE$ or AE	}	similarly for four segments
$\rho_{\overline{111.1}}$	recombination in segment AB, BC, CD, EF or $AB \cup BC \cup CD \cup EF$		
$\rho_{\overline{11.11}}$	recombination in segment AB, BC, DE, EF or $AB \cup BC \cup DE \cup EF$		
$\rho_{\overline{1.111}}$	recombination in segment AB, CD, DE, EF or $AB \cup CD \cup DE \cup EF$		
$\rho_{\overline{.1111}}$	recombination in segment BC, CD, DE, EF or $BC \cup CD \cup DE \cup EF$ or BF		

$\rho_{\overline{11111}}$  recombination in segment AB, BC, CD, DE, EF or  $AB \cup BC \cup CD \cup DE \cup EF$  or AF

b. For the recombination parameter  $\rho_{1\dots}$ , the smallest number of total complementary gametic group probabilities is one and is the same as the recombination parameter itself. The recombination parameter is equal to the sum of sixteen elementary gametic group probabilities ( $2^{n-k-1} = 2^{6-1-1} = 2^4 = 16$ ), namely,

$$\begin{aligned}\rho_{1\dots} &= \rho_{10\dots} + \rho_{11\dots} \\ &= (\rho_{100\dots} + \rho_{101\dots}) + (\rho_{110\dots} + \rho_{111\dots}) \\ &= (\rho_{1000\dots} + \rho_{1001\dots}) + (\rho_{1010\dots} + \rho_{1011\dots}) + (\rho_{1100\dots} + \rho_{1101\dots}) + (\rho_{1110\dots} + \rho_{1111\dots}) \\ &= (\rho_{10000} + \rho_{10001}) + (\rho_{10010} + \rho_{10011}) + (\rho_{10100} + \rho_{10101}) + (\rho_{10110} + \rho_{10111}) \\ &\quad + (\rho_{11000} + \rho_{11001}) + (\rho_{11010} + \rho_{11011}) + (\rho_{11100} + \rho_{11101}) + (\rho_{11110} + \rho_{11111})\end{aligned}$$

For the recombination parameter  $\rho_{\overline{.11.1}}$ , the smallest number of total complementary gametic group probabilities is four ( $2^{k-1} = 2^{3-1} = 2^2 = 4$ ). The probabilities are

$$\rho_{\overline{.11.1}} = \rho_{.10.0} + \rho_{.01.0} + \rho_{.00.1} + \rho_{.11.1}$$

Each of those four probabilities is composed of four ( $2^{n-k-1} = 2^{6-3-1} = 2^2 = 4$ ) elementary complementary gametic group probabilities. Hence, the total number is  $4 \times 4 = 16$ , is the same for all recombination parameters, and is always equal to one-half of the total number of elementary gametic group probabilities. The sixteen elementary probabilities are

$$\begin{aligned}\rho_{\overline{.11.1}} &= \rho_{.10.0} + \rho_{.01.0} + \rho_{.00.1} + \rho_{.11.1} \\ &= \rho_{01000} + \rho_{01010} + \rho_{11000} + \rho_{11010} \\ &\quad + \rho_{00100} + \rho_{00110} + \rho_{10100} + \rho_{10110} \\ &\quad + \rho_{00001} + \rho_{00011} + \rho_{10001} + \rho_{10011} \\ &\quad + \rho_{01101} + \rho_{01111} + \rho_{11101} + \rho_{11111}\end{aligned}$$

The recombination parameter  $\rho_{1\dots}$  expressed in terms of gametic frequencies is [the number of each gametic frequency is use below in part (c)]

$$\begin{aligned}\rho_{1\dots} &= \underset{1}{p_{011111}} + \underset{2}{p_{100000}} + \underset{3}{p_{011110}} + \underset{4}{p_{100001}} + \underset{5}{p_{011100}} + \underset{6}{p_{100011}} + \underset{7}{p_{011101}} + \underset{8}{p_{100010}} \\ &\quad + \underset{9}{p_{011000}} + \underset{10}{p_{100111}} + \underset{11}{p_{011001}} + \underset{12}{p_{100110}} + \underset{13}{p_{011011}} + \underset{14}{p_{100100}} + \underset{15}{p_{011010}} + \underset{16}{p_{100101}} \\ &\quad + \underset{17}{p_{010000}} + \underset{18}{p_{101111}} + \underset{19}{p_{010001}} + \underset{20}{p_{101110}} + \underset{21}{p_{010011}} + \underset{22}{p_{101100}} + \underset{23}{p_{010010}} + \underset{24}{p_{101101}} \\ &\quad + \underset{25}{p_{010111}} + \underset{26}{p_{101000}} + \underset{27}{p_{010110}} + \underset{28}{p_{101001}} + \underset{29}{p_{010100}} + \underset{30}{p_{101011}} + \underset{31}{p_{010101}} + \underset{32}{p_{101010}}\end{aligned}$$

c. The linkage values  $\lambda_{ij} = \lambda_{1\dots}$  and  $\lambda_{jlmn} = \lambda_{\overline{.11.1}}$  are expressed in terms of the frequencies of gametes as follows (see pp. 2.52 to 2.56 for details on how to assign a negative or positive sign to the frequency of each kind of gamete):

Kind of gamete $P$				Kind of gamete $P$			
	$P$	$c(ij \cap P)$	$c(jlmn \cap P)$		$P$	$c(ij \cap P)$	$c(jlmn \cap P)$
(a)	{ 111111 }	0 +	0 +	(e)	000011 $ijkl$	2 +	2 +
(b)	011111 $i$	1 - 1	0 +		000101 $ijkm$	2 +	2 +
	101111 $j$	1 - 18	1 -		000110 $ijkn$	2 +	2 +
	110111 $k$	0 +	0 +		001001 $ijlm$	2 +	3 -
	111011 $l$	0 +	1 -		001010 $ijln$	2 +	3 -
	111101 $m$	0 +	1 -		001100 $ijmn$	2 +	3 -
	111110 $n$	0 +	1 -		010001 $iklm$	1 - 19	2 +
(c)	001111 $ij$	2 +	1 -		010010 $ikln$	1 - 23	2 +
	010111 $ik$	1 - 25	0 +		010100 $ikmn$	1 - 29	2 +
	011011 $il$	1 - 13	1 -		011000 $ilmn$	1 - 9	3 -
	011101 $im$	1 - 7	1 -		100001 $jklm$	1 - 4	3 -
	011110 $in$	1 - 3	1 -		100010 $jkln$	1 - 8	3 -
	100111 $jk$	1 - 10	1 -		100100 $jkmn$	1 - 14	3 -
	101011 $jl$	1 - 30	2 +		101000 $jlmn$	1 - 26	4 +
	101101 $jm$	1 - 24	2 +		110000 $klmn$	0 +	3 -
	101110 $jn$	1 - 20	2 +	(f)	000001 $ijklm$	2 +	3 -
	110011 $kl$	0 +	1 -		000010 $ijkln$	2 +	3 -
	110101 $km$	0 +	1 -		000100 $ijkmn$	2 +	3 -
	110110 $kn$	0 +	1 -		001000 $ijlmn$	2 +	4 +
	111001 $lm$	0 +	2 +		010000 $iklmn$	1 - 17	3 -
	111010 $ln$	0 +	2 +		100000 $jklmn$	1 - 2	4 +
(d)	111100 $mn$	0 +	2 +	(g)	{ 000000 $ijklmn$	2 +	4 +
	000111 $ijk$	2 +	1 -				
	001011 $ijl$	2 +	2 +				
	001101 $ijm$	2 +	2 +				
	001110 $ijn$	2 +	2 +				
	010011 $ikl$	1 - 21	1 -				
	010101 $ikm$	1 - 31	1 -				
	010110 $ikn$	1 - 27	1 -				
	011001 $ilm$	1 - 11	2 +				
	011010 $iln$	1 - 15	2 +				
	011100 $imn$	1 - 5	2 +				
	100011 $jkl$	1 - 6	2 +				
	100101 $jkm$	1 - 16	2 +				
	100110 $jkn$	1 - 12	2 +				
	101001 $jlm$	1 - 28	3 -				
	101010 $jln$	1 - 32	3 -				
	101100 $jmn$	1 - 22	3 -				
	110001 $klm$	0 +	2 +				
110010 $kln$	0 +	2 +					
110100 $kmn$	0 +	2 +					
111000 $lmn$	0 +	3 -					

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

The linkage value  $\lambda_{ij}$  is expressed in terms of gametic frequencies with frequencies numbered as follows (the numbers relate to those in part (b) above):

$$\begin{aligned} \lambda_{ij} = \lambda_{1\dots} &= 1 - 2\rho_{1\dots} \\ &= 1 - 2 \left( \begin{array}{l} \rho_{10000} + \rho_{10001} + \rho_{10010} + \rho_{10011} \\ + \rho_{10100} + \rho_{10101} + \rho_{10110} + \rho_{10111} \\ + \rho_{11000} + \rho_{11001} + \rho_{11010} + \rho_{11011} \\ + \rho_{11100} + \rho_{11101} + \rho_{11110} + \rho_{11111} \end{array} \right) \\ &= 1 - 2 \left( \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ p_{011111} + p_{100000} + p_{011110} + p_{100001} + p_{011100} + p_{100011} + p_{011101} + p_{100010} \\ 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ + p_{011000} + p_{100111} + p_{011001} + p_{100110} + p_{011011} + p_{100100} + p_{011010} + p_{100101} \\ 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ + p_{010000} + p_{101111} + p_{010001} + p_{101110} + p_{010011} + p_{101100} + p_{010010} + p_{101101} \\ 25 & 26 & 27 & 28 & 29 & 30 & 31 & 32 \\ + p_{010111} + p_{101000} + p_{010110} + p_{101001} + p_{010100} + p_{101011} + p_{010101} + p_{101010} \end{array} \right) \end{aligned}$$

#### Exercise 2.14.

The linkage value  $\lambda$  can be expressed as a function of the frequencies of all gametes, while the recombination value  $\rho$  can be expressed as a function of the frequencies of only half of the gametes. For example,  $\lambda_1 = p_{00} - p_{01} - p_{10} + p_{11}$  [equation (2.43A)] whereas  $\rho_1 = p_{01} + p_{10}$  [equation (2.40b)]. In addition, those functions for the linkage values are orthogonal to each other [see top of p. 2.29 and p. 2.46 below equation (2.67B)]. The linkage value is also analogous to the correlation coefficient in that the values for both of them range from  $-1$  to  $+1$ .

#### Exercise 2.15.

The deterministic model assumes that the population is so large in size that no sampling variation occurs from generation to generation. The model uses the expected values in all calculations. In contrast, the stochastic model assumes that the population is sufficiently small in size so that sampling variation does occur in allelic, gametic, and genotypic frequencies from one generation to the next generation.

#### Exercise 2.16.

a. Equation (2.95) gives the usual definition of the covariance between variables  $X$  and  $Y$ , namely,

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$$

Equation (2.95a) and (2.95b) give equivalent alternative expressions (lessor known) for the covariance between variables  $X$  and  $Y$  where the  $Y$  variable is the mean of the conditional distribution of  $Y$  for a fixed value of  $X$  [equation (2.95a)] and the  $X$  variable is the mean of the conditional distribution of  $X$  for a fixed value of  $Y$  [equation (2.95b)], namely,

$$\begin{aligned} \text{Cov}(X, Y) &= E_X [XE(Y|X)] - \mu_X \mu_Y \\ &= E_X [X\mu_{Y.X}] - \mu_X \mu_Y \quad \text{equation (2.95a)} \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(X, Y) &= E_Y [E(X|Y)Y] - \mu_X \mu_Y \\ &= E_Y [\mu_{X.Y}Y] - \mu_X \mu_Y \quad \text{equation (2.95b)} \end{aligned}$$

If both variables are discrete, the covariance between  $X$  and  $Y$  can be written more specifically as in equation (2.96), namely,

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$$

$$= \sum_i \sum_j p_{ij} X_i Y_j - \left( \sum_i p_{i.} X_i \right) \left( \sum_j p_{.j} Y_j \right)$$

Equation (2.96a) gives the equivalent alternative expression for the covariance between  $X$  and  $Y$  when the  $Y$  variable is the mean of the conditional distribution for fixed values of  $X$ .

We apply the equation (2.96a) to equation (2.89) to verify its equivalence, namely,

$$\text{Cov}(X, Y) = E[X \mu_{Y.X}] - \mu_X \mu_Y$$

$$\begin{aligned} &= \sum_i p_{i.} X_i \left( \sum_j \frac{p_{ij}}{p_{i.}} Y_j \right) - \left( \sum_i p_{i.} X_i \right) \left( \sum_j p_{.j} Y_j \right) \\ &= \frac{1}{2}(-X) \left[ \frac{1+\lambda_1}{\frac{4}{\frac{1}{2}}}(-Y) + \frac{1-\lambda_1}{\frac{4}{\frac{1}{2}}}(Y) \right] + \frac{1}{2}(X) \left[ \frac{1-\lambda_1}{\frac{4}{\frac{1}{2}}}(-Y) + \frac{1+\lambda_1}{\frac{4}{\frac{1}{2}}}(Y) \right] - (0)(0) \\ &= \frac{1}{2}(-X) \left[ \frac{-Y - \lambda_1 Y + Y - \lambda_1 Y}{2} \right] + \frac{1}{2}(X) \left[ \frac{-Y + \lambda_1 Y + Y + \lambda_1 Y}{2} \right] \\ &= \frac{1}{2}(-X) \left[ \frac{-2\lambda_1 Y}{2} \right] + \frac{1}{2}(X) \left[ \frac{2\lambda_1 Y}{2} \right] \\ &= \frac{1}{2}\lambda_1 XY + \frac{1}{2}\lambda_1 XY \\ &= \lambda_1 XY \quad \text{which agrees with equation (2.97), calculated from equation (2.96)} \end{aligned}$$

b. In part (b) we calculate the sum of cross products in an alternative and equivalent way, using the predicted value  $\hat{Y}$  instead of  $Y$  itself, namely,

$$\begin{aligned} \sum (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum (X_i - \bar{X})(\hat{Y}_i - \bar{Y}) \\ &= \sum (X_i - \bar{X})(\bar{Y} + b(X_i - \bar{X}) - \bar{Y}) \\ &= \sum (X_i - \bar{X})b(X_i - \bar{X}) \\ &= b \sum (X_i - \bar{X})^2 \\ &= \left[ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \right] \sum (X_i - \bar{X})^2 \\ &= \sum (X_i - \bar{X})(Y_i - \bar{Y}) \end{aligned}$$

### Exercise 2.17.

a. The recombination value in the larger segment under the assumption of no interference is:

$$\rho_{\bar{1}\bar{1}} = \rho_{1.} + \rho_{.1} - 2\rho_{1.}\rho_{.1} = 0.30 + 0.40 - 2(0.30)(0.40) = 0.70 - 0.24 = 0.46 \quad [\text{see equation (2.56)}]$$

b. The coefficient of coincidence is defined [see below equation (2.56b)]

$$c = \frac{\rho_{11}}{\rho_{1.}\rho_{.1}}$$

$$\rho_{11} = c\rho_{1.}\rho_{.1}$$

Therefore the recombination value in terms of the coefficient of coincidence is

$$\rho_{11}^- = \rho_{1.} + \rho_{.1} - 2\rho_{11} = \rho_{1.} + \rho_{.1} - 2c\rho_{1.}\rho_{.1} \quad [\text{see equation (2.52a) and below equation (2.56b)}]$$

Or, in terms of the interference

$$\rho_{11}^- = \rho_{1.} + \rho_{.1} - 2(1-i)\rho_{1.}\rho_{.1}$$

c. The recombination value for the coefficient of coincidence of 0.90 is

$$\rho_{11}^- = \rho_{1.} + \rho_{.1} - 2c\rho_{1.}\rho_{.1} = 0.30 + 0.40 - 2(0.9)(0.30)(0.40) = 0.70 - 0.24(0.9) = 0.70 - 0.216 = 0.484$$

As the interference increases, the recombination parameter increases in value.

**Exercise 2.18.**

We assume a diagram like equation (2.101). Considering row 1, we have

$$d_{11} = p_{11} - p_1^2 \quad d_{12} = p_{12} - p_1 p_2 \quad \cdots \quad d_{1m} = p_{1m} - p_1 p_m$$

$$\sum_{j=1}^m d_{1j} = \sum_{j=1}^m (p_{1j} - p_1 p_j) = \sum_{j=1}^m p_{1j} - \sum_{j=1}^m p_1 p_j = p_1 - p_1 \sum_{j=1}^m p_j = p_1 - p_1(1) = 0$$

Since this can be proved for every row, the double sum  $\left[ \sum_{i=1}^m \sum_{j=1}^m d_{ij} = \sum_{i=1}^m \left( \sum_{j=1}^m d_{ij} \right) = \sum_{i=1}^m (0) = 0 \right]$  is also equal to

zero.