Statistical Genetics
Agronomy 615
W. E. Nyquist
March 2004

EXERCISES FOR CHAPTER 3

Exercise 3.1.
a. Define random mating.
b. Discuss what random mating as defined in (a) above means in a single infinite population composed of unisexual individuals. Composed of bisexual individuals.
c. Discuss what random mating as defined in (a) above means for a cross between two populations.
d. Some workers define random mating as that situation when any individual has an equal chance of mating with any other individual of the opposite sex. Criticize the definition.

Exercise 3.2. Why is the random mating theorem so important?

Exercise 3.3. Consider one locus with two alleles only and a population with the following genotypic frequencies

$p_{AA} = 0.25$
$p_{Aa} = 0.10$
$p_{aa} = 0.65$

What will be the genotypic frequencies next generation after individuals mate at random?

Exercise 3.4. Consider a single, two allelic locus (or a single allele and all others together in a multiple allelic locus) and different populations in which the genotypic frequencies are:

	Population 1	Population 2	Population 3
$p_{AA} = p_{AA}$	0.25	0.30	0.00
$p_{Aa} = p_{A\bar{A}}$	0.10	0.00	0.60
$p_{aa} = p_{\bar{A}\bar{A}}$	0.65	0.70	0.40

Suppose that the individuals in each of the populations mate at random, what are the frequencies of the three genotypes (or categories) in the next generation?

Exercise 3.5. What percentage of a random-mating population showing 5% recessives is heterozygous?

Exercise 3.6. In two populations both of which are in Hardy-Weinberg equilibrium and consist of a single, two allelic locus with complete dominance, one has a frequency 0.02 for the dominant phenotype and the other has a frequency of 0.02 for the recessive phenotype. What would be the frequency of heterozygotes in each population?

Exercise 3.7. Consider a random-mating population with a single two-allelic locus. Under what conditions would the frequency of heterozygotes be nearly twice (equals twice in the limit) the frequency of that allele?

Exercise 3.8. Show that in a random-mating (not absolutely essential) population with two alleles one-half of the heterozygous offspring have a female parent which is heterozygous.

Exercise 3.9. Suppose that one were given a population composed only of heterozygotes. The genotypes and their frequencies are:

Genotype	Frequency
A_1A_2	$\frac{1}{6}$
A_1A_3	$\frac{1}{3}$
A_2A_3	$\frac{1}{2}$

a. What are the frequencies of the alleles in the population?

b. Give the genotypes and their frequencies for the offspring generation from actual random mating of the above parents and tallying up the kinds of offspring produced from each mating (work with fractions rather than decimals).

c. Are the offspring in Hardy-Weinberg equilibrium? Why?

Exercise 3.10. The A "allele" for the ABO blood groups actually consists of two subtypes, A_1 and A_2, either being considered "A". In Caucasians, about 3/4 of the A alleles are A_1 and 1/4 are A_2. Among individuals of genotype AO, what fraction would be expected to be A_1O? What fraction A_2O? What would be the expected proportions of A_1A_1, A_1A_2, and A_2A_2 among AA individuals? (After D. L. Hartl, 1980, Principles of Population Genetics, Chapter 2, Problem 6, p. 138.)

Exercise 3.11. In a random-mating population what is the maximum proportion of heterozygotes with two alleles? With three alleles? With m alleles? (After J.F. Crow and M. Kimura, 1970, An Introduction to Population Genetics Theory, Harper & Row, Chapter 2, Problem 6, p. 56.)

Exercise 3.12. In a random-mating population there are 8 times as many heterozygotes as homozygous recessive individuals. What is the frequency of the recessive allele? Repeat the above for 98 times. (After J.F. Crow and M. Kimura, 1970, An Introduction to Population Genetics Theory, Harper & Row, Chapter 2, Problem 1, p. 56.)

Exercise 3.13. Suppose that in an experimental or natural population for a single, two-allelic, incompletely dominant character, the following counts were obtained:

Genotype	Number
AA	13,655
Aa	678
aa	12
Total	14,345

Calculate the Hardy-Weinberg expectations. Do you think there is a significant departure from Hardy-Weinberg equilibrium? Some have suggested that a usual chi-square test, $\chi^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$, with 2 degrees of freedom be conducted. Can you criticize this test for some reason and suggest a possible change in the χ^2 test?

Exercise 3.14. If the allelic frequencies at a single autosomal locus are different in the two sexes, how many generations of random mating are required to reach Hardy-Weinberg proportions? Explain what is occurring.

Exercise 3.15. Suppose we make crosses at random between two populations

Population 1: $0.25AA + 0.50Aa + 0.25aa$

Population 2: $0.16AA + 0.48Aa + 0.36aa$

with the males coming from population 1 and the females from population 2, what is the genotypic composition of the resulting population? What is the genotypic composition of the population(s) which would result from random mating of the crossed population itself and when would equilibrium be reached?

Exercise 3.16. Suppose that two large populations, X and Y, of the same size have the following arbitrary genotypic structure with respect to one autosomal locus:

a. Give the allelic frequencies in each population (note: $p_{11}^X \neq p_{11}^Y$, etc).

X			Y		
A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
p_{11}^X	$2p_{12}^X$	p_{22}^X	p_{11}^Y	$2p_{12}^Y$	p_{22}^Y

Suppose furthermore, that these populations mate in such a way that each of the two mates is always from a different population, otherwise being selected at random.

b. Find the genotypic structure and the allelic frequencies in the cross generation. Use random-mating theorem to obtain genotypic structure.

c. Now if the individuals of the cross generation mate at random, equilibrium will be obtained in the next generation. Why? What is the genotypic structure in the population at equilibrium in terms of the allelic frequencies in the cross generation?

d. Using the formulas derived above, and the following numerical data for the two initial populations,

X			Y		
A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
0.20	0.50	0.30	0.40	0.40	0.20

find the allelic frequencies in each population, and the genotypic structure and allelic frequencies in the cross population. Find the genotypic structure in the cross population at equilibrium after one or more generations of random mating.

Exercise 3.17. In a population with a stable allelic structure at every locus, i.e., the allelic frequencies remain the same from generation to generation at all loci (no mutation, migration, or selection), does this imply a stable multilocus gametic structure or a stable genotypic structure? Explain.

Exercise 3.18. In equation (3.3)(p. 3.5), we define random mating as the condition

$$p\left(G_i^m \cap G_j^f\right) = p\left(G_i^m\right)p\left(G_j^f\right)$$

for every combination of *i* and *j*. *G* itself may represent one or more loci. On the following page (p. 3.6), it is asserted that if equation (3.3) is true, then equation (3.3) is also true for any subset of loci of *G*. For example, if the subset of *G* consists of only one locus, say locus *A*, and only one population is assumed $\left[p\left(A_i^m\right) = p\left(A_i^f\right) = p\left(A_i\right) \text{ for } i = 1, \cdots, m\right]$, then the assertion is that

$$p\left(A_i \cap A_j\right) = p\left(A_iA_j\right) = p\left(A_i\right)p\left(A_j\right)$$

$$p_{ij} = p_i p_j \qquad \text{for } i, j = 1, \cdots, m_A$$

which is to say that Hardy-Weinberg proportions exist for that locus as well as every other locus in the set represented in *G*.

a. Assuming that equation (3.3) is true with respect to the genotype *G* which consists of only two multiple allelic loci, *A* and *B*, that a single population exists so that male and female gametic arrays are equal [see equation (3.15)], and that linkage equilibrium exists in the population, prove that Hardy-Weinberg proportions exist for each of the two loci.

b. Repeat (a) under the assumption of linkage disequilibrium.

Exercise 3.19. Is it possible that linkage equilibrium may exist for some combinations of *i* and *j* at loci *A* and *B*, but linkage disequilibrium may still be present for the whole population with respect to loci *A* and *B*? Explain.

Exercise 3.20. For two alleles at each of two loci, the gametic frequencies uniting to form a population have the following properties:

$$p_A = 0.4, p_B = 0.7, \Delta_{AB} = 0.07$$

a. With random union of gametes what are the frequencies of the following genotypes in the population?

$$\frac{AB}{AB}, \frac{AB}{ab}, \frac{Ab}{aB}$$

b. If $\rho_1 = \frac{1}{4}$, what are the frequencies of the gametes from the population? Calculate them by considering the probability of every genotype and the conditional probability of every gamete given a specific genotype. Compare these gametic frequencies with those obtained by using the formula [equation (3.82)]:

$$p_{A_iB_jt} = \rho_0 p_{A_iB_jt-1} + \rho_1 p_{A_i} p_{B_j}$$

c. What is Δ_{AB} for the gamete *AB* from the population after random mating? Calculate by definition and also from the initial linkage disequilibrium in the previous or initial generation.

Exercise 3.21. Genes *A* and *B* are linked with 20% recombination between them. An initial population, denoted generation 0, is composed of *AB*/*AB*, *AB*/*ab*, and *ab*/*ab* plants in the ratio of 1:2:1. The population is allowed to pollinate at random.

a. What would be the frequencies of the four kinds of gametes produced by the initial population?
b. What would be the frequency of the *AB*/*aB* genotype in the next generation?
c. What would be the gametic frequencies when equilibrium is reached?
d. What would be the frequency of the *AB*/*aB* genotype at equilibrium?
e. How many generations would be required for the population to go halfway to equilibrium?

(After J. F. Crow and M. Kimura, 1970, An Introduction to Population Genetics Theory, Chapter 2, Problem 17, p. 58.)

Exercise 3.22. Equation (3.65) states that $\Delta_{A_1B_1} + \Delta_{A_1B_2} = 0$, and $\Delta_{A_2B_1} + \Delta_{A_2B_2} = 0$ for two-allelic loci; and similarly $\Delta_{A_1B_1} + \Delta_{A_2B_1} = 0$, and $\Delta_{A_1B_2} + \Delta_{A_2B_2} = 0$. Prove that the sum of the two-locus linkage disequilibrium values equals zero for every row and every column with multiple alleles at both loci A and B.

Exercise 3.23.

a. Define the Hardy-Weinberg condition for an autosomal locus. When is it achieved?

b. Define linkage disequilibrium in a single population. What brings about linkage disequilibrium? Is association of genes on the same chromosome essential for linkage disequilibrium? Discuss the role that chromosomal association has with linkage disequilibrium. What effect does linkage disequilibrium have upon the frequencies of the two-locus genotypic array?

c. Contrast the Hardy-Weinberg condition with linkage equilibrium.

Exercise 3.24. Suppose that in a population with two alleles at each of two loci, the allelic frequencies are $p_{A_1} = 0.3, p_{A_2} = 0.7, p_{B_1} = 0.2, p_{B_2} = 0.8$. What would be the gametic frequencies if each kind of gamete possessed the greatest amount of linkage disequilibria? What would the linkage disequilibrium values be for each kind of gamete?

Exercise 3.25. There are two alternative ways to express the frequency of the A_iB_j gamete in generation *t* in terms of conditions in generation 0. Give them. State in word what the essence of the real difference is.

Exercise 3.26. Suppose that in a population with $p_{A_i} = 0.3, p_{B_j} = 0.8,$, and $p_{A_iB_j} = 0.28$ with recombination value 0.1 between loci. What would be the frequency or probability of $p_{A_iB_j}$ after 10 generations of random mating. (You do not need to calculate the value numerically; simply show what it is equal to numerically.)

Exercise 3.27. In a random-mating population what fraction of the linkage disequilibrium between two loci is lost each generation? What fraction is lost after *t* generations?

Exercise 3.28. The loss in linkage disequilibrium in *t* generations of random mating is equal to $\left(1-\rho_0^t\right)\Delta_{A_iB_j0}$. Show that this total loss in linkage disequilibrium is the sum of a geometric series of t terms. (Review: If a geometric series of *n* terms is defined as: a, ar, ar^2, …, ar^{n-1}, then the sum of the *n* terms can be shown to be $S = \frac{a - ar^n}{1-r}$.)

Exercise 3.29. We have studied linkage disequilibrium and its rate of loss as it approaches linkage equilibrium in a large random-mating population. How would you expect inbreeding to affect the rate of loss of linkage disequilibrium in a large inbred population? Explain. (After J.F. Crow and M. Kimura, 1970, An Introduction to Population Genetics Theory, Harper and Row, Chapter 3, Problem 12, p. 113.)

Exercise 3.30.

a. Suppose that in a population with

$$p_{A_i}^m = p_{A_i}^f = 0.3, \qquad p_{B_j}^m = p_{B_j}^f = 0.8 \qquad p_{A_iB_j}^m = 0.28, \qquad p_{A_iB_j}^f = 0.26$$

(gametic frequencies are those of the gametic output from individuals in generation zero) and with recombination values $\rho_1^m = 0.0$ and $\rho_1^f = 0.1$, what are the linkage disequilibrium values for gamete A_iB_j in males and females in the initial generation and in the next generation (generation 1)? What would they be in generation 2? Derive the general expressions for linkage disequilibrium in the male and also in the female in generation *t* as a function of initial linkage disequilibria. By use of these general expressions, calculate the linkage disequilibrium values for gamete A_iB_j in males and females in generation 3. Calculate the mean linkage disequilibria for generations 1, 2, and 3 by formula, and check their correctness by averaging the male and female linkage disequilibria in each generation calculated above.

b. For the same initial conditions as in (a) above except assume that the two loci are located on separate chromosomes, $\rho_1^m = \rho_1^f = 0.5$, calculate the linkage disequilibrium values for the socalled generation 1. What condition comes about in this generation? How can one then most easily calculate the linkage disequilibrium in the socalled generation 3?

Exercise 3.31. Consider two population, two loci, each with multiple alleles, which are mated as described in Exercise 3.16 (each of the two mates is always from a different population, otherwise being selected at random). State the conditions when linkage equilibrium and linkage disequilibrium exist for the gametes produced from the cross population itself.

Exercise 3.32. In a cross between two populations delineate the conditions that can lead to linkage equilibrium and those that can retain linkage disequilibrium of the gametic output of two loci from the cross between two arbitrary populations. Note that linkage equilibrium for loci *A* and *B* is defined such that $p_{A_iB_j} = p_{A_i} p_{B_i}$ for all combinations of $i = 1, \ldots, m_A$ and $j = 1, \ldots, m_B$. If that equality fails for any two or more combinations, we have linkage disequilibrium for those two loci.

Exercise 3.33. In Example 3.2 (pp. 3.53 to 3.60), it is observed that in generations 1, 2, … , ∞, Hardy-Weinberg proportions exist at both loci, but that the two-locus genotypic frequencies are not equal to the product of the marginal, single-locus frequencies except in generation ∞. For example, in generation 2 the frequency of the double homozygote *AB*/*AB*, 9/64, is not equal to 1/4(1/4) = 1/16. In addition, it is observed that the frequency of the coupling-phase, double heterozygote is not equal to that of the repulsion-phase, double heterozygote, except in generation ∞. All of these observations presumably imply linkage disequilibrium.

On the other hand, in Example 3.4 (pp. 3.68 to 3.72) in the cross population itself, it is observed that the two-locus genotypic frequencies are equal to the product of the marginal, single-locus frequencies (e.g., 0.189 = 0.630(0.300)), which presumably implies linkage equilibrium, but yet the frequency of the coupling-phase, double heterozygote is not equal to that of the repulsion-phase, double heterozygote, which presumably implies linkage disequilibrium.

Explain why all of these observations exist, and/or why these apparent contradictions exist. Carefully sort out the essence of things.

Exercise 3.34. In the handouts we have not dealt much with linkage disequilibrium in the cross population itself. Instead we have been concerned with the linkage disequilibrium of the gametic output from the cross. In Example 3.4 (p. 3.69) two populations in linkage equilibrium and with different allelic frequencies were crossed. It is observed that the total frequency of coupling phase double heterozygotes in the cross itself was 0.075 and that of repulsion phase double heterozygotes 0.095 (pp. 3.70 and 3.71). One half of the difference between the total frequencies of the coupling and repulsion double heterozygotes is

$$\frac{1}{2}\left[\begin{pmatrix}\text{total freq}\\ \text{of coupling}\\ \text{double heter}\end{pmatrix}-\begin{pmatrix}\text{total freq}\\ \text{of repulsion}\\ \text{double heter}\end{pmatrix}\right]=\frac{1}{2}(0.075-0.095)=-0.010$$

This could be interpreted to be a measure of linkage disequilibrium in that a common measure of linkage disequilibrium in a single population is to take one-half of that difference between the total frequencies of coupling and repulsion double heterozygotes [see equation (3.68)], namely,

$$\Delta_{AB}=\tfrac{1}{2}[2p_{AB}p_{ab}-2p_{Ab}p_{aB}]=p_{AB}p_{ab}-p_{Ab}p_{aB}$$

a. What would be the linkage disequilibrium for the *AB* gamete from the *m* population measured as a deviation from the mean allelic frequencies in the cross population itself? For the *AB* gamete from the *f* population? What would be the mean linkage disequilibrium for the *AB* gamete?

b. Repeat (a) above for the three remaining kinds of gametes.

c. If you use mean gametic frequencies and calculate $p_{AB}p_{ab}-p_{Ab}p_{aB}$, what value do you obtain? Is it equal to -0.010? Why or why not? Can you suggest a reason for the relation when you use mean gametic frequencies?

Exercise 3.35. What similarities exist between two-locus and three-locus linkage disequilibrium? Why do we regard three-locus linkage disequilibrium as less important than two-locus linkage disequilibrium?

Exercise 3.36. Give an example of a set of gametic frequencies for three loci such that all three two-locus combinations are in linkage equilibrium, but the set of three is not in equilibrium. Do not assume the same allelic frequencies at the three loci. (After J. F. Crow and M. Kimura, 1970, An Introduction to Population Genetics Theory, Chapter 2, Problem 16, p. 57.)

Exercise 3.37. Suppose that for a single, sex-linked, two-allelic locus the genotypic frequencies in the two sexes in a population are:

Heterogametic sex (male): $p_A = p$, $p_a = q$

Homogametic sex (female): $p_{AA} = p^2$, $2p_{Aa} = 2pq$, $p_{aa} = q^2$

Show that the population is in equilibrium by writing out the six types of matings with appropriate frequencies, and their corresponding offspring, making a separate tabulation for each sex.

Exercise 3.38. For a certain sex-linked trait assumed to be in equilibrium the proportion of recessive individuals among females, the homogametic sex, is 20.25%. What percentage of males would you expect to show this trait? What would be the answer if the population were not in equilibrium?

Exercise 3.39. A woman whose brother is affected with a condition due to a rare, sex-linked recessive allele seeks genetic counseling (both of her parents are normal). What is the probability that the woman is a carrier of the recessive allele? Assuming she is a carrier, what is the probability that she will have an affected son? What is the overall probability that she will have an affected son? (After D. L. Hartl, 1980, Principles of Population Genetics, Chapter 1, Problem 8.)

Exercise 3.40. Algebraic expressions involving p and q, where $p + q = 1$, may be reexpressed in one or more alternative ways. Show that

a. $p^2q + pq^2 = q - q^2$

b. $q^3 + pq^2 = (1-p)^2$

c. $(1-2q)^2 = (2p-1)^2$

d. $p^2 - q^2 = p - q$

e. $1 - 4pq = (q-p)^2$

f. $\dfrac{1-2p}{pq} = \dfrac{1}{p} - \dfrac{1}{q}$

Note that $p + q = 1$, $p = 1 - q$, $q = 1 - p$, and $p^2 + 2pq + q^2 = 1$. (After C.C. Li, 1976, First Course in Population Genetics, Pacific Grove, California, Chapter 1, p. 15.)