Problem Set 1, Bio 4181: Due Sept. 8, 2009

1. Voles (*Microtus ochrogaster*) were trapped in old fields in southern Indiana and were genotyped for the autosomal transferring locus. The following numbers of genotypes were recorded, where T^E and T^F represent different alleles:

Genotype	T^ET^E	T^ET^F	T^FT^F
Number	407	170	17

a. Characterize this population by its genotypic frequencies. **2 pts.**

Genotype	T^ET^E	T^ET^F	T^FT^F
Genotype Freq.	0.6852	0.2862	0.0286

b. Characterize the gene pool by the allele frequencies for T^E and T^F. **1 pt.**

Alleles	T^E	T^F
Allele Freq.	0.828	0.172

c. Using the Hardy-Weinberg law, predict the expected genotypic frequencies. **2pts.**

Genotype	T^ET^E	T^ET^F	T^FT^F
H-W. Freq.	0.6861	0.2845	0.0295

d. Test the goodness of fit of this population to the Hardy-Weinberg expectations. Use the standard 5% level of significance to accept of reject Hardy-Weinberg frequencies. **3 pts.**

	T^ET^E	T^ET^F	T^FT^F		
Exp.	407.515	168.970	17.515	Sum=chi-square	Prob. (df=1)
$(o-e)^2/e$	0.000651218	0.006282336	0.015151515	0.022085068	0.88186102

Accept hypothesis of HW.

2. Most black bears (*Ursus americanus*) are black or brown in color, but in one coastal population of British Columbia, white bears are also common. White bears occur when a bear is homozygous for a guanine at a single nucleotide site within the coding region of the autosomal melanocortin 1 receptor locus, whereas brown or black bears are homozygous or heterozygous for an adenine at this same nucleotide sites. A sample of bears from this population was scored at this nucleotide site with the following results:

Genotype	*AA*	*AG*	*GG*
Number	42	24	21

a. Characterize this population by its genotypic frequencies. **2 pts.**

Genotype	*AA*	*AG*	*GG*
Genotype Freq.	0.4828	0.2759	0.2414

b. Characterize the gene pool by the allele frequencies for *A* and *G*. **1 pt.**

Alleles	A	G
Allele Freq.	0.621	0.379

c. Using the Hardy-Weinberg law, predict the expected genotypic frequencies. **2 pts.**

Genotype	*AA*	*AG*	*GG*
H-W. Freq.	0.3853	0.4709	0.1439

d. Test the goodness of fit of this population to the Hardy-Weinberg expectations. Use the standard 5% level of significance to accept of reject Hardy-Weinberg frequencies. **3 pts.**

Genotype	*AA*	*AG*	*GG*		
Exp.	33.517	40.966	12.517	Sum=chi-square	Prob. (df=1)
(o-e)2/e	2.146871009	7.026123302	5.748646338	14.92164065	0.00011207

Reject hypothesis of HW.

e. Assuming that you can only observe the hair color of this sample of bears, what are the phenotype frequencies (black/brown versus white), and what is your estimate of the frequency of the "white" allele, assuming Hardy-Weinberg? **2 pts.**

Color	Black/Brown	White
Number	66	21
Frequency	0.75862069	0.24137931

Freq. of G allele= sqrt of q^2 = 0.491303684

f. How well does the estimated allele frequency in part e) explain the color frequencies under the assumption of random mating? Does this imply that random mating is a good model for this locus and that your estimate of the frequency of *G* from the hair color data is good? **3 pts.**
The fit is perfect, both phenotype categories have the expected freq.=observed frequency. However, there are 0 degrees of freedom, so the fit tells one nothing about the goodness of the random mating model, which is indeed rejected in part d. Also, the estimate of the freq. of *G* is not necessarily good, and indeed is discrepant with the estimator under part b).

3. Albinism in humans is an autosomal recessive trait. Let *A* and *a* be the "wild-type" and albino alleles at this locus. The frequency of albinos in one human population is 1 in 20,000. Assume random mating.

a. What is the estimated frequency of the albino allele? **2 pts.**
b. What is the frequency of carriers (heterozygotes) in the population? **2 pts.**
c. What proportion of matings would be between two carriers? **2 pts.**

freq. albino	0.00005
sqrt of q^2	0.007071068
2pq	0.014042136
4p^2q^2	0.000197182

4. Although few human populations satisfy all the assumptions of the Hardy-Weinberg model, most genetic polymorphisms fit the Hardy-Weinberg expectations for genotype frequencies. Suppose a microarray is used to score SNPs in a human population (normally thousands to hundreds of thousands of SNPs are scored, but here we will have a microarray with only 5 SNPs). Sometimes the scoring for SNP genotypes is incorrect in microarrays, with the most common error being to score a heterozygote as a homozygote. An almost universal quality

control device in genetic surveys of SNPs in humans is to score for deviations from Hardy-Weinberg. Among the five following SNPs, which (if any) are likely to be subjected to scoring errors and why? **6 pts.**

SNP	No. AA	No. Aa	No. aa	**Chi-Square**	**Prob.**
1	63	102	85	**7.8882**	**0.0050**
2	30	110	110	**0.0961**	**0.7565**
3	71	119	60	**0.5325**	**0.4656**
4	137	89	24	**2.7660**	**0.0963**
5	179	62	9	**1.4970**	**0.2211**

Answer: first test all loci for HW (1 pt. each): the results are indicated above. SNP 1 is not in HW, and it has a heterozygote deficiency. All other SNPs are in HW, so SNP 1 is likely to be subjected to a scoring error.

5. Consider two populations, each surveyed at two autosomal loci, each with two alleles (*A* and *a* at one locus, *B* and *b* at the second locus):

Gamete	*AB*	*Ab*	*aB*	*ab*	No. Of Individuals
Gamete Frequency Pop. 1	0.16	0.04	0.64	0.16	100
Gamete Frequency Pop. 2	0.12	0.28	0.18	0.42	300

a. Calculate the linkage disequilibrium (D) between these two loci in each population. **2 pts.**

In both populations D=0

b. Calculate the alternative linkage disequilibrium measure (D') between these two loci in each population. **2 pts.**

In both populations D'=0

c. Suppose the two samples are pooled and analyzed as a single population. What are D and D' between these two loci in the pooled population? **4 pts.**

The pooling is ¼ to ¾ so:

Pooled Pop.	AB	Ab	aB	ab
Freq.	0.13	0.22	0.295	0.355
D=	-0.01875			
allele freq.	p(A)=	0.35	p(B)=	0.425
	p(A)p(B)=	0.14875	p(a)p(b)=	0.37375
D'=	-0.12605042			

d. Suppose the pooled population mates at random to produce the next generation and that the two loci are on different autosomes. Calculate the gamete frequencies from this generation produced by random mating and its associated D and D' measures of disequilibrium. **5 pts.**

From Figure 2.4 in book:

Genotypes	AB/AB	AB/Ab	AB/aB	AB/ab	Ab/Ab	Ab/aB	Ab/ab	aB/aB	aB/ab	ab/ab
rm freq.	0.0169	0.0572	0.0767	0.0923	0.0484	0.1298	0.1562	0.087025	0.20945	0.126025
r=	0.5									

next genr. Gene pool.

gametes	AB	Ab	aB	ab	Sum
rm Freq.	0.139375	0.210625	0.285625	0.364375	1
D=	-0.009375				
alllele freq.	p(A)=	0.35	p(B)=	0.425	
	p(A)p(B)=	0.14875	p(a)p(b)=	0.37375	
	-				
D'=	0.06302521				