Extra practice problems for final exam:

Note: these practice problems serve to cover material taught by me (GWAS and QTLs) that was NOT covered already in problem sets 5 and 6, exam III, and the previous practice problems I released. The remaining material is also relevant for the final exam, but you already have a variety of problems that cover it.

Good luck,
Aviv

Problem 1

You study the genetic basis of heart disease in the Framingham Heart Study. Within the cohort, you have identified 140 cases (with heart disease) and 200 controls (without heart disease).

1. You genotype the gene ApoE2, and obtain the following results:

	Cases	Controls
11	**60**	**60**
10	**50**	**50**
00	**30**	**90**
Totals	**140**	**200**

Estimate whether genotype status at ApoE2 is significantly associated with heart disease (chi-squared values are below). Take P<0.01 as significant.

2 You now realize that for privacy reasons, the actual genotypes have been scrambled but you can still work with the allele frequencies. As a result your table is

	Cases	Controls
1	**170**	**170**
0	**110**	**230**
Totals	**280**	**400**

Estimate whether allele 1 is significantly associated with heart disease. Take P<0.01 as significant.

3. What is the required level of nominal P-values from an individual test to achieve genome-wide significance of 0.01 if you test 4 million loci?

4. Given your answer in 3, how would your significance estimate change in 1 and 2 if the results had come as part of measuring 4 million genotypes, rather than a single locus?

5. You now study the same locus using a trio design. You have recruited 50 families with 60 heterozygous parents (i.e. in 40 trios there is only one heterozygous parent and in 10 families both parents are heterozygous). You find that 350 heterozygous parents transmitted the 1 allele and 250 transmitted the 0 allele. Estimate the significance of association of the 1 allele with heart disease.

p-value	df=1	df=2
0.1	2.705544	4.60517
0.01	6.634897	9.21034
10^{-3}	10.82757	13.81551
10^{-4}	15.13671	18.42068
10^{-5}	19.51142	23.02585
10^{-6}	23.92813	27.63102

Problem 2

You study the genetic basis of Crohn's disease. Your colleague at MGH has identified 200 cases (with Crohn's) and 350 controls (without Crohn's). You perform a genome-wide study of 500,000 loci.

1. The best-scoring locus, in chromosome 5, had the following results:

	Cases	Controls
11	**55**	**125**
10	**80**	**75**
00	**65**	**150**
Totals	**200**	**350**

Estimate whether the genotype at this locus is significantly associated with Chron's (chi-squared values are below) at a genome-wide significance level of 0.01.

2. What would have happened if you only worked with alleles, not genotypes? Would allele 1 be significantly associated with Crohn's and if so at which level? Assuming that you did NOT perform multiple tests, what could be the reason to the discrepancy between your result in 1 and 2?

p-value	df=1	df=2
0.1	2.705544	4.60517
0.01	6.634897	9.21034
10^{-3}	10.82757	13.81551
10^{-4}	15.13671	18.42068
10^{-5}	19.51142	23.02585
10^{-6}	23.92813	27.63102

NOTE: Association problems look remarkably similar. To further practice such problems, you can just replace values in the contingency tables, and repeat the calculations.

Problem 3

You study abdominal bristles in flies. You cross two inbred lines of *Drosophila* one with a mean bristle number of 30 and the other with 50, both with a standard deviation of 4.

1. You obtain F1s with a mean number of 70 bristles and a standard deviation of 8. Are these results roughly consistent with an additive model of genotypic variance? Explain.

2. You now realize that your lab mate has mixed the fly labels, and that in fact you have produced the F1s from a cross of two other lines: one with a mean bristle number of 60, the other with 80 and a variance of 16. Are these results roughly consistent with an additive model?

3. You proceed to produce F2s, which have mean bristle size of 70 and a standard deviation of 5. What are the environmental variance, genotypic variance and broad-sense heritability of bristle number?

4. What is the minimal number of genes that may be affecting bristly number in your fly population?

Problem 4

You are interested in the heritability of longevity.

1. You start by studying longevity in mice. Since lab strains appear to have little variation in lifespan, you decide to conduct a large scale breeding experiment in a wild population of mice captured in Boston. In your captured population, the mean life span is 2 years, and the standard deviation is 0.5 year. You choose as your truncation point 3 years, and obtain a population with a mean of 4 years. The

offspring of the selected animals have a mean life span of 2.6 years. What is the narrow sense heritability of lifespan in your mice?

2. You decide to repeat the selection process, choosing a new truncation of 2.7 years, obtaining a population (from the offspring) with a mean life span of 3 years. What is the expected mean life span of <u>their</u> offspring?

3. You wish to compare your estimates to ones from human studies. You write your colleague, a human geneticist, who sends you data on longevity from a study of identical twins. The correlation coefficient of longevity between identical twins is 0.3. What is the estimated broad-sense heritability of longevity according to this data?

4. While you are preparing your study for publication, another study of longevity in humans is published, this one based on measures from full-siblings. The correlation coefficient reported by that study was 0.1. What is the broad-sense heritability of longevity based on your competitor's study? How can you reconcile your finding in (3) with this new one?