Problem Set 4 — 140.668 Statistical Genetics
Due 18 May 2001

1. In my study of disease X, I gathered a random sample of 25 cases and 25 controls (from a nice randomly mating population with no substructure) and genotyped each individual at the candidate gene MFC-1 ("my favorite candidate 1"). The allele frequencies among cases and controls were as follows. (For simplicity, we're going to ignore the fact that alleles are encased as pairs within humans.)

	Allele							
	1	2	3	4	5	6	7	8
Cases	13	9	4	4	5	8	1	6
Controls	21	8	6	10	0	2	3	0

   (a) Perform a χ^2 test on the above table. Calculate P-values via the asymptotic null distribution and (if you are able) via a permutation test.

   (b) Collapse the above 2×8 table into eight different 2×2 tables by looking at "allele i versus *not* allele i" for each of the eight alleles, and calculate the χ^2 statistic for each. Calculate a P-value for the maximum of these statistics, using the usual asymptotic null distribution with a Bonferroni correction. If you are able, also do a permutation test.

   (c) What do you learn?

2. Show that in applying the TDT to data for which one parent's genotype is missing, one may introduce bias.

   Consider data on an affected individual and his/her parents at a diallelic marker with alleles 1 and 2 at frequencies p and $1 - p$, and suppose we have Hardy-Weinberg equilibrium and random mating in the parents' generation.

   Consider a set of case-parent trios in which one parent has genotype 1/2 and the other parent's genotype is missing. Fill in the details to show that in the absence of linkage disequilibrium (LD), the TDT can still give apparent evidence for LD.