

This exam has two types of questions. First eight questions for which written answers are required, thereafter a number of questions of multiple choice type (one true and two to three false alternatives). Partial or complete solutions may be handed in, but will be considered only if the final score is slightly below 60% of the maximum score attainable. Write your full name in the top right corner of the exam and also on each page of any solutions handed in. Good luck!

1. Explain the terms DNA marker, allele, genotype and haplotype. (1p)
2. The only free parameter in a parametric linkage analysis is the recombination fraction θ . How is this parameter defined/interpreted? (1p)
3. Assume that you are trying to locate a gene responsible for an autosomal dominant disorder and that you have ascertained a large number of nuclear families. The average proportion of affected siblings in your families is close to 50%, supporting the hypothesis of autosomal dominant inheritance pattern, but one large sibship has 10 affected and no unaffected children. What is the most likely explanation to this extreme departure from 50% affected siblings in the large sibship? (1p)
4. The strength of a genetic component can for example be expressed as the average number of alleles shared IBD by an affected sib pair. Order the following three genetic models from the weakest genetic effect to the strongest: (1p)
 - A: Disease allele frequency $p = 0.001$, $f_0 = 0$, $f_1 = 1$, $f_2 = 1$;
 - B: Disease allele frequency $p = 0.001$, $f_0 = 0$, $f_1 = 0$, $f_2 = 1$;
 - C: Disease allele frequency $p = 0.001$, $f_0 = 0.2$, $f_1 = 0.5$, $f_2 = 0.8$;
5. What is meant by the *genotypic values* for a quantitative trait? (1p)
6. What is the *genetic variance* of a quantitative trait? (1p)
7. Explain the concept *gametic phase disequilibrium*? (1p)

8. Why is allelic association caused by population stratification considered spurious? (1p)
9. What is the expected number of marker alleles shared IBD by an affected sib pair under the hypothesis of no linkage? (1p)
- 0;
 - 1;
 - 1.5;
 - 2;
10. The population prevalence of a disease is defined as the proportion of individuals in a population that are affected by the disease. A rather common disease is specified by the disease allele frequency $p = 0.003$, and the penetrance parameters $f = (f_0, f_1, f_2) = (0.03, 0.60, 0, 60)$.
- (a) Calculate the population prevalence under the assumption of Hardy-Weinberg equilibrium. (1p)
- 0.0300;
 - 0.0316;
 - 0.0334;
 - 0.8700;
- (b) What is the probability that an affected individual is a disease gene carrier (i.e. that he or she carries one or two copies of the disease allele)? (1p)
- 0.1076;
 - 0.1123;
 - 0.9108;
 - 0.9323;
11. The heterozygosity of a DNA marker is defined as the probability that two alleles drawn at random from the population will be different. What is the minimum number k of equally frequent alleles (each with population frequency $1/k$) that leads to $> 70\%$ heterozygosity. Hint: the probability of heterozygosity is one minus the probability of homozygosity. (1p)

- 4;
- 5;
- 6;
- 7;

12. Consider the pedigree in Figure 1. Assume that the disease allele is very rare in the population and that the genetic model is autosomal dominant with full penetrance and zero phenocopy rate.

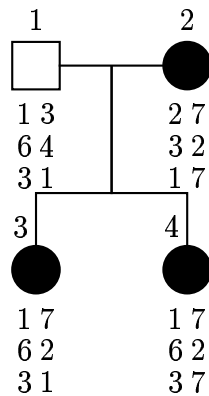


Figure 1: A nuclear family genotyped at three marker loci.

- (a) Calculate the maximum two-point parametric lod score at the first (top) marker locus: (1p)
- 0.3;
 - 0.2;
 - 0.1;
 - 0.0;
- (b) Is the support for linkage the same at all these three marker loci? (1p)
- Yes;
 - No, the support is the same for only two of the loci;
 - No, all the three lod scores are different.

13. You probably recognize the pedigree in Figure 2 since it was used both in the lectures on parametric linkage analysis and in the first computer assignment. Let us assume that the genetic model is autosomal dominant with no phenocopies and 100% penetrance. The parametric two-point lod score at $\theta = 0$ depends on the relative frequency of the 2-allele in the population, but does it also depend on the population relative frequency of the disease allele? (1p)

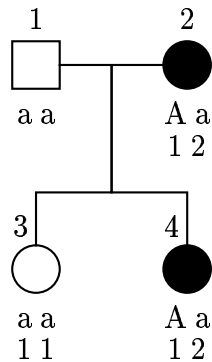


Figure 2: A nuclear family with missing marker data for the father.

- Yes, it decreases with increasing relative frequency of the disease allele;
 - Yes, it increases with increasing relative frequency of the disease allele;
 - No;
14. Five families were included in a linkage study and LOD-scores were calculated for a single marker at the recombination fractions 0, 0.1, 0.2, 0.3, 0.4, and 0.5. The results are summarized in the table below:

family	Z(0)	Z(0.1)	Z(0.2)	Z(0.3)	Z(0.4)	Z(0.5)
1	$-\infty$	-2.1	-0.1	0.4	0.1	0
2	$-\infty$	-0.6	-0.2	-0.01	-0.05	0
3	$-\infty$	-0.1	0.4	0.6	0.3	0
4	2.7	2.3	2.0	1.4	0.6	0
5	1.0	0.6	0.3	0.1	0.04	0

What is the maximum total lod score over this grid of θ -values? (1p)

- 2.49;
- 2.40;
- 2.70;
- 3.10;

15. A quantitative trait is influenced by a single QTL with 2 different alleles, A_1 and A_2 . Assume that the trait value, Y , can be modelled according to $Y = X + e$, where the mean zero environmental deviation e is uncorrelated with the genotypic value X . Suppose that the average trait value in the population is 10 for heterozygous individuals. The genetic model is specified by a homozygous effect of 3 and a dominance coefficient equal to 0.5. Suppose further that the population relative frequency of the A_2 -allele is 1%, that the population is in Hardy-Weinberg equilibrium, and that mating is random.

(a) Determine the genotypic values of A_1A_1 -homozygous and A_2A_2 -homozygous individuals, respectively. (1p)

- 7 and 13, respectively;
- 6.5 and 13.5, respectively;
- 5.5 and 11.5, respectively;
- both equal to 10;

(b) Calculate the additive allelic effects. (1p)

- $\alpha_1 = -0.0447$ and $\alpha_2 = 5.6103$;
- $\alpha_1 = -0.0447$ and $\alpha_2 = 4.4253$;
- $\alpha_1 = -0.5217$ and $\alpha_2 = 4.4253$;
- $\alpha_1 = -0.5217$ and $\alpha_2 = 5.6103$;

(c) What is the (total) genetic variance of the trait values? (1p)

- 0.3965;
- 0.7214;
- 0.6319;
- 1.4443;

- (d) What is the (narrow-sense) heritability of the trait if the variance of the environmental deviation is assumed equal to 1? (1p)
- 7.6%;
 - 28%;
 - 35%;
 - 51%;

16. Consider a great grandparent / great grandchild relative pair.

- (a) Calculate the IBD distribution for the pair, i.e. the probability that they share 0, 1 or 2 alleles identical by descent. (1p)
- 0.50, 0.50, and 0, respectively;
 - 0.25, 0.75, and 0, respectively;
 - 0.75, 0.25, and 0, respectively;
 - 0.50, 0.25, and 0.25, respectively;
- (b) Calculate the expected number of alleles shared IBD by the two relatives at any given locus. (1p)
- 0;
 - 0.25;
 - 0.75;
 - 1;
- (c) Calculate the kinship coefficient, Θ . (1p)
- 0;
 - 1/4;
 - 1/8;
 - 1/16;
- (d) Assume a quantitative trait with additive variance equal to 1. Calculate the genetic covariance of the trait values for the two relatives. (1p)
- 0;
 - 1/4;
 - 1/8;
 - 1/16;

17. In a sample of 5 family trios (each consisting of both parents and one affected child), 6 parents were found to be heterozygous at a biallelic marker locus with alleles M_1 and M_2 .
- (a) What are the smallest and largest possible value of the TDT statistic? (1p)
- 0 and 6, respectively;
 - 0 and 36, respectively;
 - 6 and 36, respectively;
 - 0 and 64, respectively;
- (b) How many of the heterozygous parents are expected to transmit an M_1 -allele under the null hypothesis of no association between marker and disease? (1p)
- 0;
 - 1;
 - 3;
 - 6;
- (c) Suppose we will reject the null hypothesis of no association if all six heterozygous parents transmitted the very same marker allele (either M_1 or M_2) to the affected child. If the true probability that a heterozygous parent transmits an M_1 -allele to the affected child is 0.8, what is the the power of our test procedure, i.e. what is the probability that the null hypothesis will be rejected? (1p)
- 0.18;
 - 0.26;
 - 0.52;
 - 0.83;
18. Consider a biallelic marker locus with alleles M_1 and M_2 . Suppose that the population prevalence of the M_1 -allele is 10%. Further assume that among M_1M_1 -homozygous individuals the proportion of affected individuals is α and that the corresponding proportions for marker heterozygous and M_2M_2 -homozygous individuals are $\alpha/2$ and $\alpha/3$, respectively. Assuming that mating is random and that the population is in Hardy-Weinberg equilibrium, calculate the probability that an

M_1M_2 -heterozygous parent has transmitted an M_1 -allele to an affected child. (1p)

- 0.52;
- 0.61;
- 0.83;
- 0.96;