

Write your full name in the top right corner of each page handed in, and use separate pages for solutions of different problems. Good luck!

1. (a) Define the concept marker homozygosity. (1p)

(b) Is marker homozygosity good or bad in a linkage study. Give a short motivation to your answer. (1p)

2. (a) Define the concepts penetrance and phenocopy? (1p)

(b) Explain the two concepts using a simple example. (1p)

(c) Four cases of neurofibromatosis can be seen in the pedigree in Figure 0 (last page of the exam). Suggest a penetrance model that fits the disease pattern observed in this family? Hint: This problem has no unique solution. (1p)

3. A study aimed at finding a gene governing extreme male adult length was initiated by an anatomist who knew nothing about quantitative genetics. He decided upon an affected-sib-pair (ASP) design and defined the binary trait "extreme length" as adult length ≥ 195 cm. Fifty ASPs were identified. One of the pairs is formed by the two brothers numbered 10 and 11 in Figure 0.

(a) Assume that these two brothers have genotypes (1,2) and (3,4), respectively, at a marker locus M and that the parents were not genotyped for this marker. Count and list the possible combinations of parental genotypes at the marker locus given the genotypes of the two sons. Hint: Mother (X,Y) and Father (Z,W) or ...? (1p)

- (b) The genotypes of the brothers at another marker locus N , which is a SNP marker, turned out to be (1,2) and (1,2). In this situation it is not possible to determine the IBD count unambiguously. Determine the IBD probabilities instead. Assume that the two alleles are equally frequent in the population. (1p)
- (c) Is the IBS count at marker locus N observable, and if so, what is it? (1p)
- (d) One marker with high level of heterozygosity turned out to be so informative that all the 50 IBD counts could be determined unambiguously. Twenty ASP:s share two alleles IBD and the remaining 30 pairs share one allele IBD at the marker locus. Calculate the NPL score, and compare it to the recommended genome-wide-scan thresholds 3.1 for suggestive linkage and 4.1 for significant linkage. (1p)
4. The NPL score is standardized. What is meant by that and what is the general purpose of standardization of test statistics? (2p)
5. A population on an island was formed by immigration of 70% individuals with the genotype A_1B_1/A_1B_1 and 30% A_2B_2/A_2B_2 .
- (a) What is the initial level of linkage disequilibrium (LD)? (1p)
- (b) What is the expected level of LD after four generation of random mating if the recombination frequency between A and B is 0.01? (1p)
- (c) How many generations are required to reduce the LD by 50%?(1p)
6. In Europe Phenylketonuria is a genetic disease affecting approximately 1 in 10000 newborns. The allele causing the disease is completely recessive. Assume random mating.

- (a) What are the allele frequencies of the disease causing allele and the dominant functional allele? (1p)
- (b) What proportion of all individuals are heterozygotes (=carriers)? (1p)
- (c) A healthy woman with healthy parents but a brother who is affected by Phenylketonuria marries a healthy man. Draw a pedigree describing the situation above. (1p)
- (d) What is the probability that the first child of the woman will be affected? (1p)
7. Describe concisely three scenarios which might cause linkage disequilibrium (LD). Indicate if the LD is transient or stable. Answers can be given in Swedish or English. (3p)
8. Two regression analyses were made. In analysis 1 stature (=body length) of fathers was used as the x-variable and stature of their sons was used as the y-variable. In analysis 2 uncles and nephews were used instead.
- (a) Which analysis will yield the steepest slope? Why? (1p)
- (b) What is the expected ratio of the slopes? (1p)
- (c) What would you expect if the the same analyses were made for body weight? (1p)
9. Two female monozygotic twins meet two male monozygotic twins. They marry pairwise and each couple has a baby. What is the genetic relationship between these babies in terms of the coefficient of relatedness and the coefficient of identity? (2p)

10. The pygmy locus in mouse influences body size. If the two alleles occurring are called A_1 and A_2 the genotypic values for weight is A_1A_1 14 g, A_1A_2 12 g and A_2A_2 6 g.

(a) What are the a and the d values for this locus? (2p)

(b) Make a diagram which illustrates how the additive and dominance variance can be defined through a regression analysis. Indicate in the diagram which values are used to calculate V_A , V_D and V_G , respectively. (2p)

11. In an exercise you have seen that the likelihood function for the cross $AB/ab \times Ab/ab$ have the following form;

$$L(\theta) = K[(2 - \theta)/4]^a[(1 + \theta)/4]^b(\theta/4)^c[(1 - \theta)/4]^d$$

where a is the number of AB -, b is the number of Ab -, c is the number of aB - and d is the number of ab individuals, respectively. Assume the following observation, $a = 45$, $b = 30$, $c = 5$ and $d = 20$.

(a) Calculate the expected numbers for a , b , c and d if $\theta = 0.5$. (2p)

(b) Test the hypothesis $H_0: \theta = 0.5$ vs. $H_1: \theta < 0.5$. (The number of degrees of freedom is "the number of classes" - 1. The classes in this case would be the phenotypic classes.) (1p)

12. The members of the pedigree shown in Figure 1 has been genotyped at two marker loci. Assume a dominant model with full penetrance and no phenocopies.

(a) Estimate the recombination fraction for the first (upper) locus. Hint: Fill in the disease genotypes, which are known under this genetic model, and use the top generation to deduce the phase. Then you solve the problem by counting the number of recombinants and non-recombinants. (1p)

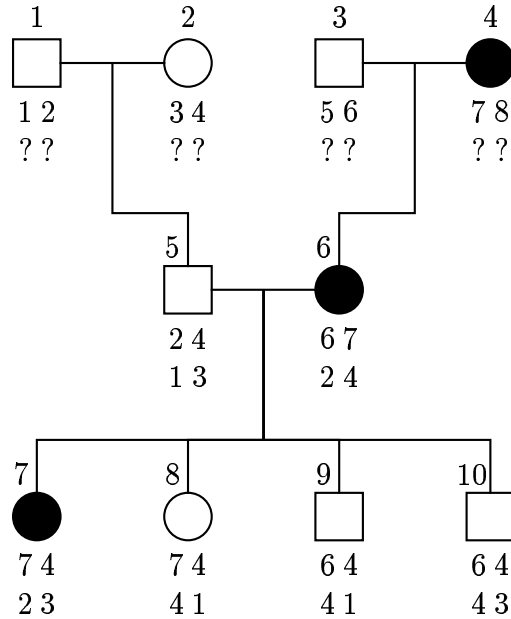


Figure 1: A three-generation pedigree with incomplete genotype data.

- (b) What is the lod score at the estimated recombination fraction at the first locus? (1p)
- (c) What is the lod score at $\theta = 0$ at the first locus? (1p)
- (d) What is the lod score at $\theta = 0.5$ at the first locus? (1p)
- (e) Estimate the recombination fraction for the second (bottom) locus? (1p)
- (f) What is the lod score at the estimated recombination fraction at the second locus? (1p)

13. Let us now return to the Holberg pedigree. Using DNA, extracted from a tissue sample from one of the affected girls (number 14), it was possi-

ble to exclude both the two hitherto identified neurofibromatosis genes NF1 and NF2. Exclusion of these two genes had also been possible in other families with several cases of this horrible disease. Could there be an additional NF-gene? Using these NF1/NF2-negative families, a set of trios consisting of an affected individual plus his/her parents were collected (one trio for each extended family) and all individuals in all the trios were genotyped for a dense set of SNP-markers covering the whole genome.

- (a) Which test would you suggest as a test for association between a specific marker allele, at one of the SNP loci, and the disease allele? (1p)

- (b) Assume that the SNP marker M has alleles M_1 and M_2 . Assume furthermore that 20 parents (in 20 different trios) were found to be heterozygous at this marker locus and that the remaining parents were homozygous M_2M_2 . Seventeen of the children in the trios with one heterozygous parent were heterozygous at the marker locus. Is the association significant? (2p)