Linkage Analysis

James O'Reilly
Student Number: `r0773125`

Pedigree One

The first pedigree is given in Figure 1.

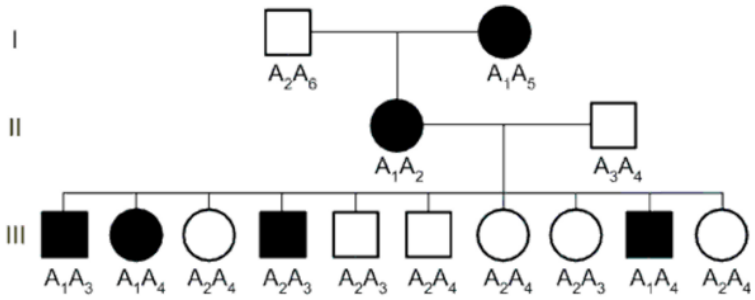

Figure 1: First pedigree

As we have the genotypes and related phenotypes for three generations, we can know the phase. Looking at the genotypes of the affected individuals, we can see that the A_1 allele is most likely the cause of the given phenotype. This means that out of the ten individuals in the third generation, recombination has only occured once. This gives a recombination fraction Θ of 0.1. The likelihood that this marker locus is linked with the disease locus is given by

$$L(\Theta) = (1 - 0.1)^9(0.1) = 0.039 \tag{1}$$

The likelihood that they are not linked is given by

$$L(\Theta = \frac{1}{2}) = (1 - 0.5)^9(0.5) = 0.00098 \tag{2}$$

The odds and log odds (LOD) are therefore given by

$$\begin{aligned} Odds &= \frac{L(\Theta)}{L(\Theta = \frac{1}{2})} = \frac{0.039}{0.00098} = 39.8 \\ \text{LOD Score Z} &= \log_{10}\left[\frac{L(\Theta)}{L(\Theta = \frac{1}{2})}\right] = 1.6 \end{aligned} \tag{3}$$

What does this mean? The odds score is greater than one and the LOD score Z is greater than zero, which indicates that it is much more likely that the A_1 marker locus is linked with the disease locus than not.

Pedigree Two

The second pedigree is given in Figure 2.

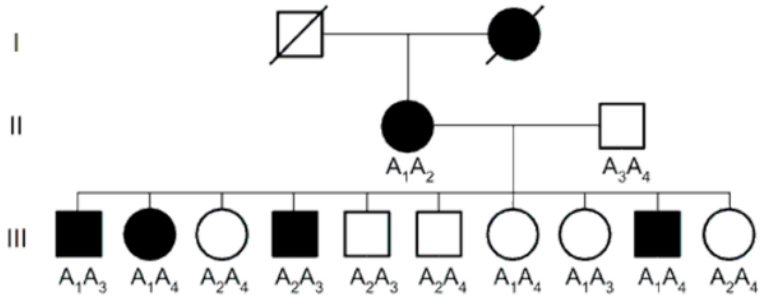

Figure 2: Second pedigree

As we have no information about the genotypes of the individuals in the first generation, we do not know the phase for certain. Therefore we assume that it either the A_1 or A_2 marker loci are linked to the disease locus, and we must estimate the likelihood for both.

If we assume that the A_1 locus is linked, this would give that two of the ten individuals in the third generation show recombination. The recombination frequency Θ_1 is therefore 0.2. If we assume that the A_2 locus is linked, this would give that nine of the ten individuals in the third generation show recombination. The recombination frequency Θ_2 is therefore 0.9. The average $L(\Theta)$ is therefore given by

$$L(\Theta) = \frac{1}{2}\left[(0.8)^8(0.2)^2 + (0.1)^1(0.9)^9\right] = 0.045 \tag{4}$$

While the likelihood that the loci are not linked is still 0.00098. Using these to calculate the LOD Score Z, we get that

$$\text{LOD Score Z} = \log_{10}\left[\frac{L(\Theta)}{L(\Theta = \frac{1}{2})}\right] = \log_{10}\left[\frac{0.045)}{0.00098}\right] = 1.66 \tag{5}$$

What does this mean? The odds score is greater than one and the LOD score Z is greater than zero, which indicates that it is likely that the either the A_1 or A_2 marker loci are linked with the disease locus than not.

I'm not actually sure if the calculation give above is correct, as I wasn't sure how to calculate $L(\Theta)$ in this case. I'm aware there is a formula given in the slides, but I wan't sure how exactly we are supposed to calculate Θ in this case. To me it felt more normal to have a Θ_1 for the recombination frequency for the A_1 loci, and Θ_2 for the A_2 loci. Otherwise, every individual is recombinant for either of of the loci, and so the recombination frequency would be 1, which doesn't make sense to me.