SISG Module Genetic Epidemiology Optional Exercise: Linkage Disequilibrium

Table 1: Haplotype and allele frequencies

Locus 2

B

b

Locus 1

A

0.04

0.50

0.54

a

0.27

0.19

0.46

0.31

0.69

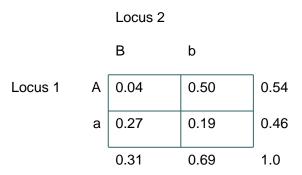
1.0

- 1. The table above provides hypothetical data, showing allele and haplotype frequencies from a population sample. Note, the designation of capital vs. lower case letters are arbitrary. At locus 2 the "b" allele is more frequent than the "B" allele in this population.
 - a. Using the notation presented in the slides, state the values for p_A, p_a, p_B, p_b, p_{AB}, p_{Ab}, p_{aB}, and p_{ab} for this population sample.
 - b. Describe in words what pA and pb represent.
 - c. What is the most frequent haplotype in this sample? The least frequent?
 - d. Based on this data, calculate D, r² and D' between Locus 1 and Locus 2.

2. Assume Locus 1 is the (untyped) disease causing SNP, and Locus 2 is the SNP you are genotyping in your case-control study. You have performed power calculations assuming that you had genotyped the disease causing allele. These power calculations indicate a sample size of 500 cases and 500 controls. Estimate the sample size you will actually need for your study.

ANSWER KEY

Table 1: Haplotype and allele frequencies



- 1. The table above provides hypothetical data, showing allele and haplotype frequencies from a population sample. Note, the designation of capital vs. lower case letters are arbitrary. At locus 2 the "b" allele is more frequent than the "B" allele in this population.
 - a. Using the notation presented in the slides, state the values for p_A, p_a, p_B, p_b, p_{AB}, p_{AB}, p_{aB}, and p_{ab} for this population sample.

ANSWER: $p_A=0.54$, $p_a=0.46$, $p_B=0.31$, $p_b=0.69$, $p_{AB}=0.04$, $p_{Ab}=0.50$, $p_{AB}=0.27$, and $p_{ab}=0.19$

- b. Describe in words what pA and pb represent.
 ANSWER: pA is the frequency of the A allele at locus 1.
 pb is the frequency of the b allele at locus 2
- c. What is the most frequent haplotype in this sample? The least frequent? **ANSWER**: The most frequent haplotype is the Ab haplotype (50% of the population. The least frequent is the AB haplotype (4% of the population).
- d. Based on this data, calculate D_{AB} , r^2 and D' between Locus 1 and Locus 2. ANSWER:

 $\begin{array}{l} D_{AB} = p_{AB} - p_{A}p_{B} = 0.04 - (0.54^{*}0.46) = -0.12 \\ r^{2} = D^{2} / (p_{A} p_{B} p_{a} p_{b}) = (-0.12^{*}2) / (0.54^{*}0.46^{*}0.31^{*}0.69) = 0.27 \\ D' = If D_{AB} < 0: D'_{AB} = D_{AB} / (min(P_{A}P_{B}, P_{a}P_{b})) = -0.12 / min(0.54^{*}0.31, 0.46^{*}0.69) \\ = -0.12 / min(0.16, 0.31) = -0.12 / 0.16 = -0.75 \end{array}$

2. Assume Locus 1 is the (untyped) disease causing SNP, and Locus 2 is the SNP you are genotyping in your case-control study. You have performed power calculations assuming that you had genotyped the disease causing allele. These power calculations indicate a sample size of 500 cases and 500 controls. Estimate the sample size you will actually need for your study.

ANSWER: N₂=N₁/r²

 $N_2=500/0.27=1851$ cases and 1851 controls.