Biostatistics 666
Problem Set 1
Due Wednesday, September 20, 2017

1. Consider a population where allele frequencies differ between the sexes. Assume that there are equal numbers of males and females and that genotypes occur in Hardy-Weinberg proportions within each sex. Focus on a single di-allelic marker in this population. The marker has allele frequency $p_M = p + \Delta$ in males and $p_F = p - \Delta$ in females, where $p=(p_F + p_M)/2$.

    a) Calculate offspring genotype frequencies after one generation of random mating.

    b) How do genotype frequencies differ from those expected under Hardy-Weinberg equilibrium?

    c) How many additional generations are required before Hardy-Weinberg equilibrium is reached?

2. In a sample of 100 individuals, 97 homozygotes for allele A, 2 homozygotes for allele B and 1 heterozygote were observed. Conditional on the number of observed A and B alleles, answer the following questions:

    a) What is the probability of this particular sample configuration?

    b) What is the probability of observing an equal or greater number of heterozygotes?

    c) What is the probability of observing a smaller number of heterozygotes?

    d) What is the chi-squared statistic for Hardy-Weinberg equilibrium?

3. Consider two loci in disequilibrium in a large population. Assume that the recombination fraction between the two loci is 0.0001. In how many generations do you expect the disequilibrium coefficient D to be halved?

4. Consider the following set of haplotype frequencies:

    $p_{AB} = 0.4$, $p_{Ab} = 0.2$; $p_{aB} = 0.2$; $p_{ab} = 0.2$

    a) Calculate the frequency of alleles A, a, B, and b.

    b) Calculate D, D’ and Δ^2 between the two markers.

    c) What is the probability that allele A will be present in a chromosome that carries allele b?

    d) What is the maximum possible value of r^2 for this marker pair?

5. The BRAVO browser (https://bravo.sph.umich.edu) lists variants and allele frequency information for many genes. Pick a gene whose name starts with the same initial as your last name. Download frequency information for missense variants in the gene and plot a histogram to summarize the data.