Do these problems, and email the results (if you can, in a PDF or else in plain text in an email message, as I have trouble with some Microsoft Word formats). Like all of the weekly homeworks this one totals 25 points.

1. (12 points) (This is Exercise 4 from Chapter 1 of the textbook) Suppose there are two populations that have genotype frequencies:

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop. 1</td>
<td>0.64</td>
<td>0.32</td>
<td>0.04</td>
</tr>
<tr>
<td>Pop. 2</td>
<td>0.09</td>
<td>0.42</td>
<td>0.49</td>
</tr>
</tbody>
</table>

If a researcher draws a sample, thinking it is coming from a single population, but it is actually composed of individuals two-thirds of whom came from population 1, and one-third from population 2,

(i) If these individuals are simply collected together, but have no time to interbreed, what will the genotype frequencies in the sample expected to be?

(ii) What will the gene frequencies be expected to be in that sample?

(iii) If we mistakenly assume that the sample is from a random-mating population, and use that sample gene frequency, what proportion of heterozygotes will we expect to see?

(Make sure that when doing averages of genotype frequencies or gene frequencies, you weight the averages 2/3 : 1/3 rather than just doing a simple average.)

2. (13 points) Suppose we draw a “panel” of 20 diploid individuals from a population and proceed to discover SNPs (single nucleotide polymorphisms) by looking for sites that vary and have more than one copy of the rare allele. We do that to exclude sites that might simply have a sequencing error. Suppose that about one site in every 1000 has a true SNP, and the rate of error in sequencing is 0.002 per copy sequenced. Further, suppose that the SNPs all happen to have their minor allele gene frequency be $p$.

(i) What fraction of the putative SNPs we find will be true ones?

(ii) What would that fraction be if we included as putative SNPs also sites where the minor allele only occurred once?

(iii) For $p = 0.01$ and for $p = 0.01$ what is the fraction computed in each of these two cases?

Hint for these: If an allele occurs with frequency $q$ in a sample of 20 diploids, the chance that it shows up $k$ times is

\[
\binom{20}{k} q^k (1 - q)^{20 - k}
\]

Note that you will need to compute the probability of 0 or 1 occurrences of an allele (or an error) and then subtract from 1, in order to get the probability that more than there are more than one occurrence. The combinatorial factor for these terms will just be $1$ or $40$. 