Homework Assignment 6
Solutions

1. Please see the code. The expected genotype counts under the null are 50, 100, and 50. Since each of these numbers is much larger than 5, we can safely use the asymptotic $\chi^2$ null distribution. The $\chi^2$ and likelihood ratio test statistics are 5.56 and 5.32, respectively. The corresponding p-values are 0.06 and 0.07. Though we do not formally reject the null hypothesis, there might be some genotyping errors present!

[3 points]

2. Please see the code. These are composite hypotheses, and we are using some functions from the lab. For locus 1, we get an A allele frequency estimate of 0.799. The likelihood ratio test statistic is 4.17, and the $\chi^2$ test statistic is 4.36. The respective p-values are 0.041 and 0.037. There is a bit of evidence that the locus is out of Hardy-Weinberg equilibrium. For locus 2, we get an A allele frequency estimate of 0.820. The likelihood ratio test statistic is 79.7 and the $\chi^2$ test statistic is 48.2. The respective p-values are tiny - the locus is very much out of Hardy-Weinberg equilibrium. The expected genotype counts under the null are (638,321,40) and (672,295,32), so we can trust the asymptotic test.

[4 points]

3. Please also see the code.

(a) The $\chi^2$ test statistic is 4.98. Using a $\chi^2$ null distribution with $(3 - 1) \times (2 - 1) = 2$ degrees of freedom, we obtain a p-value of 8.3%.
(b) We get a test statistic of 4.88, and a p-value of 8.7%.
(c) We get a p-value of 8.4%.
(d) Since the p-values are ~8%, we can conclude that there is some evidence for a difference in the survival rates for the three treatments, but it is not strong.

[4 points]

4. Please see the code. A one-sided test is more appropriate since we are interested whether the treatment helps children grow taller, and not whether there is any change (i.e. we rule out that children will shrink because of the treatment).

[3 points]

5. Please also see the code.

(a) Here is the 2x2 table:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>not A</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>268</td>
<td>48</td>
</tr>
<tr>
<td>not B</td>
<td>40</td>
<td>44</td>
</tr>
</tbody>
</table>
The question of interest is whether the prevalences of virus A and B are the same. Therefore, the null hypothesis is \( p_{01} = p_{10} \), versus the alternative \( p_{01} \neq p_{10} \) (McNemar’s test). The test statistic is \( \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} = \frac{(48 - 40)^2}{48 + 40} = \frac{64}{88} = 0.73 \). The null distribution is \( \chi^2_1 \), and we obtain a p-value of 0.39. There is no evidence against the assumption that the prevalences of virus A and B are the same.

(b) Here is the 2×2 table:

<table>
<thead>
<tr>
<th></th>
<th>Ph1</th>
<th>Ph2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>36</td>
</tr>
</tbody>
</table>

Here we are interested whether there is evidence against the assumption that the phenotype is independent of viral status, so we apply a goodness-of-fit test. For the \( \chi^2 \) test we obtain a test statistic of 5.89, and since the expected numbers in all cells are above 5, we can use the \( \chi^2_1 \) distribution as null distribution, yielding a p-value of 0.015. There is some evidence against the assumption that the phenotype is independent of viral status.