

Homework #6

Due Thursday Feb. 28 [**NOTE unusual due date!**] at the beginning of class. Assignments turned in more than 5 minutes after the beginning of class will be penalized 10 points, with an additional 10 points every 24 hours thereafter. You may discuss the homework assignment with other students, but do not share your work. **All Python programs should be run before being turned in. Even experienced programmers can seldom write a program perfectly on the first try.**

A new syndrome has been recognized among young children, and a researcher wants to determine whether it is genetic, environmental, or a mix of the two. She surveys 20 pairs of monozygotic (identical) twins and 20 pairs of dizygotic (fraternal) twins with at least one twin affected. Concordance is 19/20 monozygotic twins and 20/20 dizygotic twins. She also surveys siblings of affected children, and finds that 10% are affected.

1. (10 points) Assuming that her data collection method was sound, is this syndrome likely to be mostly genetic, mostly environmental, or a mix? Why?
2. (10 points) Describe a data-collection problem that could cast doubts on her results.

For the following three questions, determine how many haplotypes the two siblings share IBS (identical by state) and list all possibilities for sharing IBD (identical by descent). For example, you might say "These sibs share 1 haplotype IBS, and may share 1 or 0 IBD."

1. (5 points) First sibling has AB; second has CD.
2. (5 points) First sibling has AA; second has AB.
3. (5 points) First sibling has AB; second has AC.
4. (5 points) First sibling has AA; second has AA.

Briefly critique the following data collection methods. What problems might they introduce into a mapping study?

1. (10 points) Contact all members of a patients' support group for your disease and ask them to participate in your study.
2. (10 points) Survey all patients admitted with your disease from a South Seattle hospital. Use healthy UW students as a control group.

Stand-alone questions:

1. (20 points) Write a Python program which accepts two siblings (with single-letter haplotypes as shown above) and scores IBS. (You do not need to score IBD.) Test it on the sample file sibs.txt and include your test results with your homework.
2. (20 points) Write a Python program which reads observed values of P(AB),P(Ab),P(aB) and P(ab), in order, from a file. Calculate the expected values for these four numbers if linkage equilibrium were present. Compute the chi-square value comparing observed (the numbers read from the file) with expected (the numbers you calculated). You do not need to find the significance level of your chi-square. Hint:

$$\chi^2 = \sum [(observed - expected)^2 / expected]$$