

Homework 4

Introduction

In this homework it is the idea to use the simulated data we introduced before, in conjunction with the “FBAT” package, to answer relevant questions in a family-based genetic association context. All of the questions below can be answered by consulting the course notes (Chapter 5 and 6), the FBAT manual and its tutorial (cfr. FBAT Tour), all of which can be downloaded from the URL <http://www.biostat.harvard.edu/~fbat/fbat.htm>.

Specific questions on family-based genetic association analysis

Q1. How many rare SNPs are present in the simulated data, in other words, which ones have $MAF < 1\%$? Why is this information valuable?

Q2. Analyze the family data using the FBAT statistic on the continuous trait from the phenotype file, and consider the default options. What do you observe? Can you detect a significant marker?

Q3. Are the results in Q2 corrected for multiple testing? If not, how can you achieve overall significance? Think about a conservative way we saw in class...

Q4. Compute permutation based p-values for the settings of Q2 (look up “minp” in the fbat package). Vary the number of permutations. What can you tell about the precision of the “estimated” p-values?

Q5. Now re-analyze the same family data, exploring different settings in combination:

- Modes
- Models
- Options `-o` and `-e`

Compare your results with those obtained from Q2.

Write a small report, including some explanations about how you obtained the answers

Due date: 6 May 2010