

Stat 877: Statistical methods for molecular biology (Spring, 2017)

Homework #1: QTL mapping

Due 16 Feb 2017

We will consider a set of simulated data from a backcross with 300 individuals, with a single quantitative phenotype. A selective genotyping strategy was used: only the top 46 and bottom 46 individuals, by phenotype, were genotyped.

1. Grab the comma-delimited data file at

<http://www.biostat.wisc.edu/~kbroman/teaching/uwstatgen/hw1.csv>

and place it in your R working directory.

Within R, you'll need to install R/qtl via `install.packages("qtl")`

Then load R/qtl via `library(qtl)`

Then import the data file via `hw <- read.cross("csv", file="hw1.csv")`

2. Use each of standard interval mapping (by the EM algorithm) and Haley-Knott regression to map QTL in this cross. Also, use a permutation test to establish significance of identified QTL, and calculate 1.5-LOD support intervals for the locations of inferred QTL.

Do two versions of the permutation test: the usual kind plus a stratified permutation test (with individuals stratified by the amount of genotyping, and with permutations performed within these two strata).

To perform the stratified permutation test, do something like this:

```
nt <- ntyped(hw)
strat <- as.numeric(nt > mean(unique(nt)))
operm <- scanone(hw, method="hk", n.perm=1000, perm.strat=strat)
```

3. How do the results change if you omit the individuals that were not genotyped?

To drop the non-genotyped, individuals, use code like

```
hw_sub <- subset(hw, ind=(ntyped(hw) > 0))
```

4. What do you conclude, regarding the behavior of standard interval mapping vs Haley-Knott regression in the presence of selective genotyping, and regarding the use of an unstratified vs stratified permutation test?