1 Setting Significance Threshold

1.1 LOD

A loci is the position of a gene on a chromosome. Multiple testing problems occur often in QTL analysis as we do number phenotypes × number loci tests. Consider one phenotype, pea lodging and the 5199 loci in Jamin's study. We could set an adjusted pea-value. An alternative method Jamin would like is calculating a genome wide adjusted LOD score at each QTL and determining significance based on that threshold.

LOD is the log_{10} liklihood ratio comparing the null that there is not a QTL to the alternative, that there is.

$H_o : y_i \sim N(\mu, \sigma^2)$ i.e. there is no genetic dependency between the phenotype and the genotype

- use MLEs for parameter estimates $\mu = \bar{y}; \sigma^2 = RSS_o/n$

$H_a : y_i|g_i \sim N(\mu_{g_i}, \sigma^2)$

- g_i = genotype of individual i at the marker (loci); each genotype (AA,AB?) group has a different mean; σ^2 = pooled RSS = RSS_1; again the MLEs.

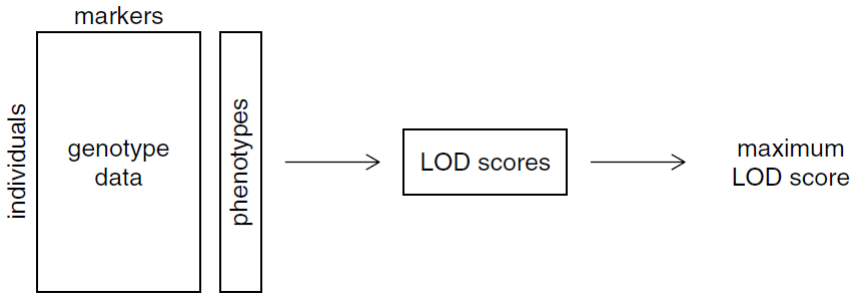
LOD = $\frac{n}{2} \times log_{10}(\frac{RSS_o}{RSS_1})$

LOD is related to the F statistic.

LOD = $\frac{n}{2} \times [F(\frac{df}{n-df-1} + 1)]$ and similar to the F statistic, large LOD values are associated with strong evidence for the alternative, that there is a relationship to the genetic loci and the phenotype.

Note that if genetic information is missing, that plant cannot be included. (Section 4.1)

1.2 Genome-wide maximum LOD



Steps:

1. Permute the phenotypic data

2. Reassign it to the genetic data

3. Calculate LOD for each loci

4. Save the maximum LOD

5. Repeat many times – Note there, is a calculation for the number of times to repeat

Use the 95%th percentile of the maximum LOD distribution as the genome wide significance threshold. It is also possible to calculate the corresponding pvalues.

Items needed:

- **Backcross**/Intercross
- Size of genome (in cM)
- Number of individuals
- Number of typed markers
- Pattern of missing genotypic data
- Phenotypic distribution

2 Considerations

- Section 4.7 – *qtl* does not (yet) have capabilities to do a joint analysis of multiple phenotypes, which Jamin has.

  A joint analysis would increase power because we aren't conducting number phenotypes × number genetic markers tests separately. It can also allow for testing of plieotrophy (vs. tight linkage) which is whether a single QTL affects multiple phenotypes.

- Did he use selective genotyping?

- The number of permutatons to run. Base this on the desired width of the pvalue CI. `http://www.rqtl.org/faq/`

look at notes from last time – what is cM

3 References

Information as well as the graphic was taken from:

Broman and Sen. 2009. *A guide to QTL mapping with R/QTL.*