

# Homework 2

## Introduction

We simulated data on a number of families, including parents and/or a number of children. The data file pedigree2.ped includes relevant data in "LINKAGE" format, and hence also data on a number of genetic markers. A second data file, pheno.dat, includes data on a continuous phenotype. These data can be linked to the pedigree data using the unique identifiers for family members.

In this homework it is the idea to use the R package as much as possible to answer relevant questions in population genetics.

## Specific questions on population genetics

Q1. Which individuals in the data could be used for population analysis?

Q2. Compute LD patterns between markers in the data, using parents only. Can you observe "blocks" of markers in high LD?

Q3. Compute LD patterns between markers in the data, using children only or the entire dataset, assuming that all individuals are independent. How do these results compare to those obtained from Q2? Are the results valid?

Q4. Compute the coefficients of inbreeding for the data at hand. How can these inbreeding coefficients give you information about the quality of the data?

Q5. How is the inbreeding coefficient different from the kinship coefficient? Identify different family structures in the data (families consisting of 1, 2, ... generations – you have probably identified these by your exploration work for HW1). Can you compute kinship coefficients between members of the same family (one family per family structure)?

Q6: Compute the HWE test for all markers. Which ones are out of HWE? Do these markers cluster together?

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

**Due date:** 18 March 2011