

Homework 3

Style 1: literature project

This involves choosing a paper from the literature that extends or provides additional background on the material of Chapter 5 on population-based genetic association studies using unrelated individuals.

Topics may include: multiple testing, power, population stratification, genomic control, structured association, tagging SNPs.

Summarize the paper of your choice (not one that has been discussed in class), its objectives, results, and show how it fits into the broader picture.

A good reference of papers can be retrieved from the URL:

<http://www.nslj-genetics.org/ld/>

Style 2: computing project

Use existing software / applets / simulation code, or write your own simulation code on issues that are relevant for population-based association studies. These may include:

- Computing the power of a population-based genetic association study, using unrelated individuals:
 - E.g.,
 - Which factors affect the power (most)?
 - Based on your simulation study, can you make recommendations for good common practice in designing and carrying out a genetic association study?
 - Do these power calculation tools adjust for multiple testing?
- Assess the effect of introducing 5%, 10%, 15%, 20%, 25%, 30% missing genotypes on the power of your study or on study results.
 - E.g.,
 - What is the best way to analyze data with missing genotypes?
 - Impute? Ignore? ...
 - What do you observe?
 - Which percentages give rise to unreliable results?
- Assess the effects of introducing different percentages of missing phenotypes in the context of a population-based genetic association study.
 - E.g.,
 - Can you make proposals to analyze these data?
 - What is the effect on power?

- Is the situation different when multiple phenotypes are recorded or repeated measures are available?
- Can you find software / applets to simulate the power (increase? decrease?) when repeated measures are available, as opposed to a single measurement, in the presence/absence of missing data?
- Simulate scenarios of population stratification and admixture and assess its effect in the context of a genetic association study

Useful URLs :

<http://www.dorak.info/hla/stat.html>

<http://cran.r-project.org/web/views/Genetics.html>

Style 3: data analysis project

Introduction

In this homework it is the idea to use the simulated data we introduced before, in conjunction with the R “SNPassoc” package, to answer relevant questions in a population-based genetic association context. However, if you can answer the questions using another technique you have seen in other classes, this is fine as well.

Specific questions on population-based genetic association analysis

Q1. Which individuals in the data should you use for the population-based genetic association analyses we envisage?

Q2. Do a genome-wide check on the validity of the HWE condition. Which markers do not satisfy this condition? [Note: Use a function which is able to check this condition using one line only (it is not the idea that you manually perform this for each marker separately, as in a previous homework)]

Q3. Are there genotypes missing? Do you observe particular patterns in the observed/missing genotypes?

Q4. Are there monomorphic SNPs? How should they be treated further on in subsequent analyses?

Q5-6-7. In the phenotype-file, a continuous trait has been provided. Consider the following two scenario's:

- Scenario 1: restrict attention to the parents

- Scenario 2: restrict attention to a single child per family
- Scenario 3: take the entire data set, ignoring relationships

Compare the results from all scenarios. What do you observe? Is this in line with expectations?

Q8. Compute appropriate significance levels for your association tests in Q5. Try out several methods and compare / discuss your results.

Q9. Perform a subset analysis for males / females on the quantitative trait. Do you see any differences? Was this to be expected?

Q10. Perform a gene-gene interaction analysis for the quantitative trait, using the tools provided by the "SNPassoc" package. Do you observe any interesting signals?

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

Due date: 2 April 2010