

Stat 890-4, 2004-3: Introduction to statistical methods in genetic association studies

Audience: Statistics graduate students and research fellows from the CIHR Interdisciplinary Health Research Team studying Inflammatory Gene Haplotypes and Susceptibility to Cardiac, Vascular and Pulmonary disease (the “GxE” studies).

Overview: At the outset, small interdisciplinary teams consisting of statistics graduate students and research fellows will be formed. Research fellows would describe the studies they are involved in, the scientific background and the biological hypotheses being investigated. Statistics students would assist in describing the statistical underpinnings of the data analysis. The course is to be centred around issues fundamental to the analysis of the GxE studies, and will have a large collaborative component for data analysis. More specialized topics may also be pursued, time and interests permitting.

Location/time: On site at St Paul’s Hospital one day per week.

Format: Our current proposal is to have a combination lecture/lab/discussion format with a breakdown of hours per week such as:

Lecture/presentations	2 hours
Lab and group work	2 hours
Discussion	1 hour
Total	<u>5 hours</u>

Grading/Evaluation: Participation 20% (including lectures, group work and discussion), presentations 40%, report 40%. Letter grades will be assigned to the Statistics graduate students.

Course Credits: Statistics graduate students would receive 2 credits for Stat 812 (consulting) and 2 credits for Stat 890 (special topics). Research fellows who register for the course would receive a certificate from the Continuing Studies program at SFU.

Outline: Selected topics from the following

1. Introduction of the GxE studies (cohorts) and data
 - Scientific background
 - Biological hypotheses
2. Selected population genetics
 - Random mating populations and Hardy-Weinberg equilibrium (HWE)
 - Population substructure
 - Departures from HWE – the Wahlund effect
 - Marker-disease associations due to hidden population structure
 - Gametic phase (linkage) disequilibrium (LD)
 - Decay of LD over time
 - Marker-disease associations due to LD (ancestral haplotypes)
 - The common disease-common variant and other hypotheses
 - Discussion in the context of the GxE studies

3. Introduction to principles of data collection

- Observational versus experimental studies
- Observational study designs
 - Cohort studies
 - Case-control studies
 - Other biased-sampling designs
 - Matching and confounding variables
 - Population structure as a hidden confounding variable
 - Ascertainment bias
- GxE cohort descriptions and study designs

4. Summarizing/exploring data

- Introduction to GxE data
- Graphical displays and smoothing
- Contingency tables
- Measures of association between variables
 - Correlation
 - Odds-ratios
- Sources of spurious associations – confounding
- Discussion of GxE data summaries

5. Introduction to statistical inference

- Statistics versus population parameters
 - Formulation of biological hypotheses as statistical hypotheses
 - Examples with GxE studies
- Basics of statistical hypothesis testing
 - Test statistics and null distributions
 - Parametric versus nonparametric testing approaches
- Quantifying variability
 - The importance of quantifying variability in estimators
 - Confidence intervals – parametric and nonparametric

6. Quantitative outcomes – linear regression models

- Confounding and effect modification
- Interpretation of results
- Residuals

7. Dichotomous outcomes – logistic regression models

- Interpretation
- Use for case-control data

8. Other topics as time permits such as accounting for hidden population structure (genomic control and related methods), uncertain haplotype phase, etc.