

STIN370 Selected topics in bioinformatics and applied statistics

Topic spring term 2023:

Computational and statistical methods in genetics

Thore Egeland, 2022-12-14

A description of the course in 2023 follows supplementing the generic description here:

<https://www.nmbu.no/course/STIN370>

Course responsible: Thore Egeland

Additional teachers: Solve Sæbø. Magnus Vigeland. Possibly more, to be announced.

Lectures: Wednesday 10:15 – 12:00 in MU319. First lecture (overview) Feb 1, last lecture May 10.

Guidance on projects. Compulsories: Time and place for guidance on projects will be agreed with the students. There will be two project papers, one on a given topic and one on a topic that the student chooses.

Assessment, exam: See <https://www.nmbu.no/course/STIN370>

Background

STIN 370 is taught for the first time the spring semester of 2023. The course is a revised version of STAT 370; the latter course has not been taught since 2017. As for the previous course, the topic of STIN 370 will typically change from one year to the next.

Course contents

The name of the course is admittedly quite vague. We will focus on some applications in genetics that are challenging and interesting also from a computational or statistical perspective. Furthermore, we will restrict attention to two topics that the teachers know well:

- **Population genetics:** This includes the study of genetic variation within populations, and involves the examination and modelling of changes in the frequencies of genes and alleles in populations over space and time. Specifically, we will discuss Hardy-Weinberg Equilibrium, the Wright-Fisher model, and models for mutations.
- **Relatedness and pedigrees:** Family relatedness is fundamental for a number of applications including breeding, forensic genetics and medical genetics (“genehunting”). Fundamental concepts like IBD (Identical By Descent) and inbreeding play a fundamental role. Also, pedigree likelihoods (The Elston-Stewart peeling algorithm and the Lander Green algorithm based on hidden Markov models) are fundamental.
- **Phenotype-genotype:** A common challenge within genetics is to relate phenotypic traits to high-dimensional genetic or metabolic variables (-omics data, RNA-seq etc). Typically, the aim is to identify a small subset of genetic variables through variable selection. We will focus on statistical methods suited for analysis and variable selection in high-throughput data situations in genetics, like projection methods (PCR, PLSR), regularization methods (Lasso), univariate methods with multiple testing adjustments, and gene set enrichment analysis.

Prerequisites

See <https://www.nmbu.no/course/STIN370>. Basic knowledge of R/R-studio.

Literature

Selected papers and lecture notes.