STT 855/EPI 855: Introduction to Statistical Genetics

Fall 2020 (3 credits)

Time/place: T & Tr 12:40PM – 2:00 PM. Videos will be made available after the class. If you are not in the US and cannot attend the class via Zoom in a synchronous mode, please contact us to discuss possible accommodations.

Instructors contact information & office hours:

Gustavo de los Campos	Yuehua Cui
1311 IQ – building (775 Woodlot Dr.)	C432 Wells Hall
Tel: (517)884-7607	Tel: (517) 432-7098 (not forward to personal phone)
Email: gustavoc@msu.edu	Email: cuiy@msu.edu (preferred contact method)
Web: https://quantgen.github.io/	Web: http://www.stt.msu.edu/~cui
Office hours: Fridays 9:00am-10:00am (Eastern time, US)	Office hours: T Th 9:00-9:30am via Zoom (a zoom link will be shared)
Join Zoom Meeting https://msu.zoom.us/j/91599932208 Meeting ID: 915 9993 2208 Passcode: STAT2020	

Required Textbook and Course Materials: The first half will follow Introduction to Quantitative Genetics (4th Edition) 4th Edition by Douglas S. Falconer & Trudy F.C. Mackay (ISBN 0-582-24302-5). For the 2nd half, Dr. Cui will make the lecture notes available in D2L and no textbook is required.

Course platforms:

- Class meetings will be conducted through Zoom (link will be provided via email)
- This course materials will be delivered ***online*** through the course management system and you will need your MSU NetID to login to the course from the ***D2L homepage* (*http://d2l.msu.edu*).**
- In ***D2L***, you will access lecture notes, video recordings, homework, homework solutions and other reading assignments and materials.

Prerequisite: STT863 + STT442 or STT862 or equivalent.

Basic skills necessary for this course: Calculus; Basic probability distribution theory; t-test; Chi-square test; linear and logistic regression; ANOVA; Maximum likelihood estimation; Computer software (e.g., R)

Required Technologies:

- Browser/mobile support for D2L: https://documentation.brightspace.com/EN/brightspace/requirements/all/browser_support.htm
- Guide for internet speed: https://broadbandnow.com/guides/how-much-internet-speed-do-i-need.
- If you are in a time-zone that makes it difficult to attend the synchronous class, or you have problems accessing any of the course materials, please contact us promptly.

Course objectives: The course is designed for Master and PhD students in statistics and/or life science majors. Students will learn basic quantitative genetics concepts and basic and advanced statistical methods for genetic data analysis.

Course description

The course is organized in two modules:

I-Introduction to quantitative genetics *(by G. de los Campos)* will cover the following topics

1. Introduction:
    - Mendel, laws of inheritance and a brief history of Quantitative Genetics
    - The "gene" (invited lecture, by prof. David Arnosti).
2. Allele and genotype frequencies (CH1 Falconer & Mackay, F&M):
    - Allele frequencies and genotype frequencies
    - Hardy-Weinberg equilibrium (definition and testing)
    - Deviations from Hardy-Weinberg equilibrium
3. Sources genetic variation (mutation, selection and drift, CH2, F&M)
4. Values and means in a single locus model (CH6 & CH7, F&M):
    - Additive, dominance and allele substation effects
    - Introduction to GWAS (more on the topic will be covered in Module 2).
    - Variance decomposition in the single-locus model
5. Extension to multi-locus models
6. Resemblance between relatives (CH9):
    - Kinship: definition and computation from pedigree and from SNP data
    - Heritability estimation using parent-offspring- and sib-regressions.
7. Correlations between traits (if time permits)
8. Association of alleles at pairs of loci:
    - Meiosis and recombination
    - Linkage and Linkage Disequilibrium
    - Metrics (D, D') definition, estimation and testing
    - Physical versus map distance

II-Statistical methods for analyses of genetic data *(by Y. Cui)* will cover the following topics

9. Mapping quantitative traits using marginal association analyses (aka GWAS)
    - Estimation and inference in a single-locus model: OLS, chi-square, logistic regression.
    - Multiple-testing (type-I error rate and FDR control)
    - Diagnostics (qqplots, p-value inflation)
    - Population stratification and controlling for population structure
10. Haplotype analysis and gene-gene interaction analysis
11. Mapping quantitative traits using family data
    - Family-based linkage mapping
    - TDT test
12. Sequencing data and rare variants analysis
13. Genetic risk prediction

Textbook:

Required: Introduction to Quantitative Genetics by Falconer & Mackay (4th ed)

Other references:

- *The Fundamentals of Modern Statistical Genetics* by Nan M. Laird and Christoph Lange, Springer-Verlag, New York.

- *Applied Statistical Genetics with R For Population-based Association Studies* by Andrea S. Foulkes. Use R! series, Springer
- *Statistical Genetics of Quantitative Traits: Linkage, Map and QTL* by Rongling Wu, Chang-Xiang Ma and George Casella, Springer-Verlag, New York.
- *Mathematical and Statistical Methods for Genetic Analysis* by Kenneth Lange, Springer-Verlag, New York.
- *Genetics and Analysis of Quantitative Traits* by Michael Lynch, Bruce Walsh, Sinauer Assoc.

Grading

Module 1:

- 20% In-class work/assignments
- 30% Homework
- 50% midterm. The midterm will be completed during class time. There will be a regular zoom session that will allow students to ask questions as in any regular exam. If you are in a time zone that won't allow you to complete the midterm during the class time, please contact Gustavo de los Campos during the first week of the course to discuss possible arrangements.

Module 2:

- 30% homework
- 70% Individual data analysis project & report
  - Research project: The instructor will conceive of multiple problems related to the topic of the 2nd module of the course. Students can work singly or as a group. ***Each student/group will choose one of the problems as their final project***. Students with different backgrounds are encouraged to work together.
  - Alternatively, ***students are also encouraged to base their class projects on one of their own problems (with real data)***. But they should consult with the instructor first.
  - Paper review: Students can choose to review one or several papers with similar topics and write a report.
- **Format of written report (**Follow the format of a refereed journal, e.g., *Genetics* or *Biometrics*): Introduction/Motivation, Methods, Results, Discussion/Critiques, References. Page limit: ≤ 20, 12pt font double-spaced pages (including tables, figures and references).

The final grade will be the average of the grade in the first and 2nd modules, and will be based on a straight scale roughly according to the following scheme

90-100%	80-89%	70-79%	65-69%	60-64%	55-59%	50-54%	0 – 49%
4.0	3.5	3.0	2.5	2.0	1.5	1.0	0

Important Dates for Fall 2020 (tentative)

- Sept. 2 First Day of Classes
- Sept. 7 Labor Day (university closed)
- Nov. 25 All in-person instruction ends
- Nov. 26-27 Thanksgiving holiday (university closed)
- March 5-9 Spring break (No class)
- Dec. 11 Classes end

Course continuity statement:

- If you experience health problems during the course, please contact the instructor to discuss your needs and accommodations for HW, in-class assignments, midterm, or final exam.
- If any of the instructors suffer health problems for an extended period of time, we will promptly communicate with you and arrange for a backup instructor.

Course Recordings, Intellectual Property and Social Media Use:

- **Course Recordings:** Meetings of this course (Zoom sessions) will be recorded. The recordings will be available to students registered for this class. Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Recordings should not be reproduced, shared with those not in the class, or uploaded to other online environments. Doing so may result in disciplinary action.

  Related Policies:

  Institutional Data Policy:

  https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/

  Student Privacy Guidelines and Notification of Rights under FERPA

  https://reg.msu.edu/ROInfo/Notices/PrivacyGuidelines.aspx

As members of a learning community, students are expected to respect the intellectual property of course instructors. All course materials presented to students are the copyrighted property of the course instructor and are subject to the following conditions of use:

1. *Students may (may not) record lectures or another classroom activities and use the recordings only for their own course-related purposes.*
2. *Students may (may not) share the recordings with other students enrolled in the class. Sharing is limited to using the recordings only for their own course-related purposes.*
3. *Video and audio recordings made of online lectures may contain inaudible or invisible watermarks to identify shared media*: https://support.zoom.us/hc/en-us/articles/360021839031-Audio-Watermark
4. *Students **may not** post the recordings or other course materials online or distribute them to anyone not enrolled in the class without the advance written permission of the course instructor and, if applicable, any students whose voice or image is included in the recordings.*
5. *Any student violating the conditions described above may face academic disciplinary sanctions.*

Attendance Policy: While attendance is not strictly required, there will be in-class assignments. To contemplate possible difficulties attending in a synchronous mode, we will post in-class assignments in D2L and the assignments will be due 24 hour after class finished.