SAINT MARY'S UNIVERSITY SINCE 1802 | FACULTY OF SCIENCE

BIOL 6608 - Biostatistics for Graduate Students
Winter 2019
Course Outline

Instructor: Timothy R. Frasier ("Tim"), Associate Professor

Contact: Office: S327
E-mail: timothy.frasier@smu.ca
Tel: 902-491-6382

Office Hours: M 03:00–05:00pm, T 02:00–03:30pm, R 01:00–03:30pm, or **by appointment**

Class: R 09:00–11:20am MM028

Course Sites: Course content (lectures, code, and example data sets) will be available at the website https://github.com/timothyfrasier/stats-2019. Although there is a Brightspace page for this course, it will only be used minimally (i.e,. for grade postings of those students enrolled). This is because the majority of students in this course are not officially enrolled, and therefore do not have access to Brightspace

Course Description:
Analysis of biological data at the advanced level. This course will focus on Bayesian techniques: teaching students the theory and practice of common Bayesian data analysis techniques. The goal is for students to leave the course understanding, and being able to implement, Bayesian alternatives to the most common statistical analyses faced by biologists (e.g., comparing two means, regression, ANOVA), as well as build models for much more complex analyses. Ideally, students will become comfortable building their own models for whatever type of analysis is appropriate for their studies and data. This will also require a rethinking of how we test hypotheses and come to conclusions when interpreting data. Analyses will use R and Stan, and therefore students will also gain proficiency in these statistical environments.

3 credit hours

Pre-requisites: None, but must obtain permission from professor.

Co-requisites: None.

Notes: This is a somewhat advanced statistics course. Although you do not need to have a statistics background, nor have taken a statistics course, you do need to have a general familiarity with, and not be afraid of, statistics. You need to be ready and willing to work through solving difficult statistical problems.

Required Materials:
You will need a laptop computer that you can install programs on and can bring to class every day (an iPad or other tablet is not enough). There is no required textbook. However, you will need supplemental reading to fully understand Bayesian statistics (i.e., it will take more than lecture notes and class time). The problem is that this course is evolving, and there is not one textbook with which it is matched. There are, however, two books that I would highly recommend. The first is *Doing Bayesian Data Analysis 2nd Edition* by John Kruschke. This is the book one which this course was originally based, and therefore it is the closest match to the lecture content. Another excellent one is *Statistical Rethinking* by Richard McElreath. The course will pull from both of these texts, and I would actually recommend both of them.

Learning Outcomes:
By the end of this course, successful students should be able to:

- Identify common misinterpretations of results from analyses incorporating null hypothesis significance testing in the scientific literature and popular media, and explain what the errors are and why they are incorrect.
- Visualize and summarize data in a meaningful way that will lead to appropriate analyses.
- Evaluate the performance of Markov Chain Monte Carlo processes and identify if/when there are problems, and develop proper solutions.
- Develop and apply Bayesian models for the analysis of data in the formats dealt with throughout the course.
- Be fluent and comfortable extending and modifying the models developed in class for the analysis alternative data sets with different characteristics.
- Internalize and conduct "best practices" for Bayesian analyses, including the iterative processes of data visualization, model building, and model testing.

Course Content and Planned/Tentative Schedule:
The schedule below is TENTATIVE. We will try to stick to this schedule, but it is likely that we will get off-track at one point or another. Necessary changes to the schedule will be made accordingly, and you will be notified of any changes during class hours.

Date	**Topic**
Jan. 10	Introduction to R
Jan. 17	Issues with p-values and null hypothesis significance testing
Jan. 24	Intro to Bayesian statistics and Markov Chain Monte Carlo (MCMC) processes
Jan. 31	Data types and probability distributions
Feb. 7	Binomial models
Feb. 14	Simple linear regression
Feb. 21	NO CLASS - Winter Break
Feb. 28	Linear regression with one categorical predictor variable
Mar. 7	Hierarchical models
Mar. 9	Multiple regression
Mar. 14	Binomial predicted variables
Mar. 21	Categorical predicted variables
Mar. 28	Ordinal predicted variables

Methods of Course Delivery:
This course will primarily follow a "workshop" format, where the instructor will talk through and demonstrate statistical techniques, and students are expected to perform the steps and analyses on their computers simultaneously along with the instructor. Therefore, it is essential that students have laptop computers available for every class. There will also be a fair amount of in-class problem solving expected of the students. A typical day will include: (1) introduction of a new type of Bayesian model by the instructor; (2) development and application of the model on a simple data set, led by the instructor; (3) students then work in pairs to modify and apply the model to different data sets with more difficult characteristics, and evaluate its performance; (4) students present their model and results to the class, explaining the reasoning behind their choices during model development and alteration. Step 3 will often extend into "homework" for the students, where we will begin the following class with the results from the previous exercise.

Marking Scheme:
Final letter grades are assigned according to the Undergraduate Ratings, Grades and Grade Points table listed on p. 32 in the Academic Regulation section of the Academic Calendar.

Component	Weight (%)	Expectations
In-class participation	20%	- Have a good attitude throughout course - Contribute regularly, constructively, and thoughtfully to discussions - Work well with others during pair or group-based problem solving exercises (meaning that you work constructively and diligently to solve the problem, and that you also have a good attitude and are easy to work with throughout the process - Are well-prepared for each class (have the appropriate materials/ supplies, have done the assigned readings, activities, etc. - Are on time for each class
Weekly exercises	50%	- Have completed the assigned exercises/activities each week - Have developed well thought-out solutions to the problems (most of the time it is fine if your model/solution is not correct, as long as it was well thought-out. It is easy for a small detail to make a model completely fail. However, it should be clear that you put in a good and thoughtful effort into finding a solution. - Have done all assigned readings and have really thought about them and how they relate to our class and your previous knowledge (and not just that you let your eyes pass over the words) - Communicate your results/interpretation in a clear, organized, and professional manner. - Have made a substantial effort to fill any knowledge gaps required for completing the assignments on your own, rather than immediately asking the instructor
Research projects	30%	- Research topic chosen was an appropriate level of difficulty and required an appropriate amount of work (e.g., was not too easy) - Developed, appropriately tested, and appropriately implemented a suitable and well thought-out solution to the problem - Showed a strong effort at self-directed learning to develop a solution - Communicated results/interpretation in a clear, organized, and professional manner

Description of Course Components:

In-class Participation	Each class will contain several aspects where active in-class participation is required. These include (but are not limited to): (a) conducting analyses along with the instructor during in-class demonstrations and asking questions when needed; (b) responding frequently, and in a thoughtful and professional manner, when questions are asked of the students; (c) participating constructively and professionally during group-work exercises; and (d) being on time and having completed any pre-class preparations needed to be ready for the day. Details regarding expectations can be found in the table above.
Weekly Exercises	Most weeks students will have "homework", consisting of reading assigned papers or developing/modifying models and analyzing data. Doing this work is the primary means through which students will struggle with Bayesian analyses on their own, and therefore really learn how to apply it to their own research projects. Students are expected to tackle these assignments with enthusiasm, professionalism, and to be thorough in their analyses. They are also expected to use self-directed learning to fill any gaps in their understanding separating what was covered in class from what they need to do to complete the assignment. Students are also expected to present their results to the class in a clear, organized, and professional manner. Details regarding expectations can be found in the table above.
Research Projects	In addition to the weekly exercises, students will complete individual research projects, where they apply Bayesian analysis techniques to their own data for their theses (hopefully), or to simulated data representative of the types of data that students expect to get as part of their research. This project should be complicated enough that it warrants the use of several different types of Bayesian models, but is also contained enough to be achievable within the semester. All research projects should be cleared with the instructor prior to beginning in earnest. At the end of the course students will give a formal presentation of their research projects, approach, and interpretation. Details regarding expectations can be found in the table above.

Student Responsibilities, Academic Integrity, & Code of Conduct:

1. To ensure that all students and guest speakers have an interactive audience for their presentations **attendance and participation are mandatory**. You must let me know in advance of any known absences.

2. Treat your colleagues and instructors with respect and give others your attention when they are speaking.

3. Smart phones may be used to take notes, but all the sounds must be turned off and you should not receive/send calls/texts, or check email during class.

4. **Academic integrity: As in all courses, plagiarism and cheating will not be tolerated.** You must hand in your own work, written in your own words. Plagiarism will be dealt with according to policies outlined in the Academic Calendar. It is your responsibility to familiarize yourself with Saint Mary's policies on Academic Integrity by consulting the "*Academic Integrity and Student Responsibility*" (p. 19–27) and "*Academic Regulations*" (p. 28–42) sections of the Academic Calendar, in order to be well informed on the consequences of dishonest behaviour.

5. Technology in the classroom: Please do not record lectures without my direct approval.

Missed Classes:
SMU faculty no longer accept "sick notes" for missed days of class or exams. Instead, if students miss a day of class, particularly when there was something due that day (e.g., a research presentation or mid-term exam), they need to read, print out, fill out, and sign a copy of the Declaration of Extenuating Circumstances. This should then be submitted to the professor, and they will keep a copy, and also give a copy to the Science Advising Centre for your records.

Accessibility:
The Fred Smithers Centre establishes individualized support services to help students with physical, medical, and learning disabilities. Accommodations work best for all concerned if the student comes forward to the Smithers Centre early. Students are encouraged to seek more information by visiting the Centre, or its website.