DALHOUSIE UNIVERSITY

Faculty of Science Course Syllabus
Department of Mathematics and Statistics
STAT 4620/5620
Data Analysis
Winter 2019

Instructor(s): Dr. Joanna Mills Flemming Joanna.Flemming@Dal.Ca Chase Building 103

Lectures: MF 12:05pm-1:25pm Chase Building 319

Office Hours: W 1:05pm-2:25pm Chase Building 103

Course Description

This course will begin with an introduction to both linear models (LMs) and generalized linear models (GLMs). Additive and generalized additive models (GAMs) will then be introduced followed by their mixed model extensions. Emphasis will be placed on understanding underlying statistical theories and on method implementation. Real and relevant data sets will be used throughout the course to both demonstrate and validate the various analysis tools. The **R programming language** will be used exclusively. Each student will have the opportunity to select a dataset of interest *to them* and use this for various aspects of the assignments and the course project. Graduate students will be required to present both their project proposal and results.

Course Prerequisites

STAT 3340, STAT 3460, or the instructors consent.

Course Objectives/Learning Outcomes

The course objectives are to provide both upper level undergraduate students and graduate students with i) working knowledge of a diverse range of statistical methodologies and ii) confidence to utilize these methodologies where appropriate in order to answer scientific questions of interest.
Specific learning outcomes:

- The capacity to recognize important features of data (heterogeneity, repeated measurements, distribution of the response variable etc.).
- A thorough understanding of linear models, additive models and their extensions.
- A thorough understanding of generalized linear models, generalized additive models and their extensions.
- A full comprehension of the concepts of zero-inflation, zero-truncation and over- and under-dispersion.
- A working knowledge of hierarchical modelling frameworks along with a full grasp of random effects.
- An understanding of tree based methods.
- An appreciation for the field of spatial statistics.
- A working knowledge of the R language and environment for statistical computing and graphics including proficiency with the statistical techniques used in this course.

DALHOUSIE UNIVERSITY

Required Course Materials

Reference texts:

- *Generalized Additive Models: An Introduction with R,* Simon Wood.
- *A Beginner's Guide to R,* Alain Zuur, Elena Ieno, and Erik Meesters.
- *Mixed Effects Models and Extensions in Ecology with R,* Alain Zuur, Elena Ieno, N. Walker, A. Saveliev, G. Smith.

DALHOUSIE UNIVERSITY

Course Assessment

Component	*Weight (% of final grade)*	*Due Date(s)*
Project Proposal	0% Undergrad / 5% Grad	Feb 4-8
Project Presentation	0% Undergrad / 5% Grad	
Project	30% Undergrad / 30% Grad	April 14
Final exam	30%	April 12
Assignments (5)	40% Undergrad / 30% Grad	Jan 18, Feb 8, Mar 1, Mar 15, Mar 29

Conversion of numerical grades to Final Letter Grades follows the Dalhousie Common Grade Scale

A+ (90-100)	**B+** (77-79)	**C+** (65-69)	**D**	(50-54)
A (85-89)	**B** (73-76)	**C** (60-64)	**F**	(<50)
A- (80-84)	**B-** (70-72)	**C-** (55-59)		

Course Policies

There will be *five* assignments. These will provide students with the opportunity to review the statistical theory and methods discussed in class and apply these techniques to analyze real and relevant datasets. These assignments *must* be completed using R (https://www.r-project.org). Late assignments will not be accepted.

Cell phones and other electronic devices should be SILENCED before class begins.

Course Content

***indicates that an assignment is due*.**

Week 1	Jan 7th Introduction	Jan 11th Linear Models
Week 2	Jan 14th Linear Models with R	Jan 18th The Exponential Family*
Week 3	Jan 21st Generalized Linear Models	Jan 25th GLMs with R
Week 4	Jan 27th Generalized Additive Models	
Week 5	Feb 4th 10-MINUTE PROJECT PROPOSALS	Feb 8th 10-MIN PROJECT PROPOSALS*
Week 6	Feb 11th Zero-Inflation and Zero-Truncation	Feb 15th Generalized Estimating Equations
STUDY BREAK		
Week 7	Feb 25th Review and Case Study	Mar 1st R Practical Session*
Week 8	Mar 4th Tree Based Methods	Mar 8th Violations of Dependence
Week 9	Mar 11th Mixed Models	Mar 15th LMMs with R*
Week 10	Mar 18th LMMs with R	Mar 22nd Generalized Linear Mixed Models
Week 11	Mar 25th GLMMs and GAMMs with R	Mar 29th Introduction to Spatial Statistics*
Week 12	Apr 1st 10-MIN PROJECT PRESENTATIONS	Apr 5th 10-MIN PROJECT PRESENTATIONS
Week 13	Apr 8th FINAL EXAM	