

STAT 331: APPLIED LINEAR MODELS

SPRING 2019

Lecture Information: MW 2:30pm – 3:50pm Arts Lecture Hall (AL) 113

Tutorial Information: M 11:30am – 12:20pm Davis Centre (DC) 1350

Note: We will not use the tutorial time slot every week. These will be used for make-up lectures and R exercises in accordance with the class schedule.

Instructor: Nathaniel T. Stevens

Office: Mathematics 3 (M3) 3143

Email: nstevens@uwaterloo.ca

Office Hours: W 12:30pm – 2:00pm

Phone: 519-888-4567 ext. 35540

COURSE DESCRIPTION

Regression modeling is one of the most useful statistical techniques available. It helps to establish the relationship between a variable of interest (the response variable) and one or more other variables thought to influence the response variable (explanatory variables). Discovering and analyzing such relationships is the basis for understanding the natural world. Consequently, regression modeling is useful in science, medicine, business, finance, marketing and countless other fields.

This course introduces students to modeling the relationship between a response variable and several explanatory variables via linear regression models. Topics include simple and multiple linear regressions; the least squares algorithm for estimation of parameters; hypothesis testing and prediction; model diagnostics and improvement; algorithms for variable selection; nonlinear regression and other methods. Students will use R to automate statistical analyses in this course.

COURSE OBJECTIVES

Any student who successfully completes this course should be able to:

- Think logically and analyze information critically in a statistical setting.
- Formulate and solve problems in an abstract framework.
- Apply statistical techniques as a solution to real-world problems involving large, complex data sets.
- Demonstrate competence with devising computational solutions to such real-world problems in R.
- Visualize, present and communicate analytical results.

COURSE HOMEPAGE

The course homepage is on LEARN (<https://learn.uwaterloo.ca>). It is my expectation that you regularly visit this webpage to download course material and receive important announcements.

COURSE COMMUNICATION

Email: All email correspondence must come from your “uwaterloo” email address.

Discussion: For discussion amongst yourselves, myself and the TAs, we will use Slack, a workplace instant-messaging app. All discussions will take place within the STAT331-Spring2019 workspace. Login using your “uwaterloo” email address and the workspace URL “stat331spring2019” at <https://slack.com/signin>.

TEXTBOOKS

Required: *Introduction to Regression Modeling* by B. Abraham and J. Ledolter, Duxbury Thomson Brooks/Cole.
*Available for 3-hour loan at the DC Library

Recommended: *An Introduction to Statistical Learning* by G. James, D. Witten, T. Hastie and R. Tibshirani, Springer.

Linear Models with R, 2nd edition by J.J. Faraway, CRC Press.

ASSESSMENT

1. Assignments (20%)

- There will be four (4) assignments, each of which is worth 5% of your final grade. These assignments will combine written solutions and computation in R.
- Assignments will be due by 11:59pm on the days listed in the table below, and they are to be submitted electronically via LEARN.
- You are expected to complete these assignments **independently** and submit your own work. Cheating is a serious offence and will be treated as such. Please refer to the Academic Integrity section below.
- If you fail to submit an assignment and have a **valid** reason with supporting documentation, the weight from that assignment will be shifted to the final exam. If the reason is not deemed valid or you do not have supporting documentation, you will receive a zero. Documentation must be provided to me within 2 business days of the missed due date.
- **Assignments up to 24 hours late will receive a penalty of 50%. Assignments more than 24 hours late will not be graded, and you will receive a zero.**

Assignment	Due Date
1	Wednesday May 22
2	Wednesday June 12
3	Wednesday July 10
4	Tuesday July 30

2. Midterm (30%)

- There will be one in-class, closed book, midterm exam on **Wednesday June 19**. This exam will last the duration of the lecture time.
- The midterm will evaluate your comprehension of the course material and may consist of a series of short answer calculations, short answer written responses, multiple choice questions, proofs and R output interpretation.
- If the midterm is missed for a **valid** reason, and you have supporting documentation, its weight will be shifted to the final exam. If the reason is not deemed valid or you do not have supporting documentation, you will receive a zero. Documentation must be provided to me within 2 business days of the midterm.
- **I will not offer a “make up” midterm for any reason.**

3. Final Exam (50%)

- There will be a 2.5-hour, closed book, cumulative final exam during the Spring 2019 Final Examination Period: **August 2 – August 16**. Please refrain from booking Summer travel before the actual exam date is scheduled.
- The format of the final exam will mimic that of the midterm. Specifically, short answer calculations, written responses, multiple choice questions, proofs and R output interpretation can all be expected.
- If the final exam is missed for a **valid** reason, and you have supporting documentation, you will receive an “incomplete” (INC) grade as long as you have at least 70% in the course going into the final exam. Otherwise, you will receive a zero. Documentation must be provided to me within 2 business days of the final exam.

*Note that for all accommodations due to illness, a [University of Waterloo Verification of Illness Form](#) (VIF) is the documentation that I require.

**If you have a dispute with your grade on an assignment or the midterm, it may be submitted to be remarked within 1 week of the assignment/ midterm being returned to you. Bear in mind that the entire assignment/ midterm is then subject to be remarked.

COURSE TOPICS

1. Simple Linear Regression

- Estimation (maximum likelihood and least squares)
- Inference (for model parameters)
- Prediction (interpolation and extrapolation)

2. Multiple Linear Regression

- Estimation (maximum likelihood and least squares)
- Inference (for model parameters)
- Prediction (interpolation and extrapolation)

3. Model Checking

- Residual diagnostics
- Addressing non-constant variance, non-normality and multicollinearity

4. Model Selection

- Goodness of fit metrics, ANOVA and partial F-tests
 - Forward, backward and hybrid selection
 - LASSO and regularization
 - Cross validation
5. Logistic Regression*
- Estimation (maximum likelihood)
 - Inference (for model parameters)
 - Prediction and classification
6. Use of R (as it pertains to the content listed above)
- *time permitting

A NOTE ON R

We will be using R throughout the course to illustrate the importance and utility of automating statistical analyses. All demonstrations and exercises will be available before lectures and tutorials. It is in your best interest to bring a laptop so that you may actively participate.

INSTITUTIONAL REQUIRED STATEMENTS

Academic integrity: In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. [See the [Office of Academic Integrity](#) for more information.]

Grievance: A student who believes that a decision affecting some aspect of his/her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read [Policy 70, Student Petitions and Grievances, Section 4](#). When in doubt, please be certain to contact the department's administrative assistant who will provide further assistance.

Discipline: A student is expected to know what constitutes academic integrity to avoid committing an academic offence, and to take responsibility for his/her actions. [See the [Office of Academic Integrity](#) for more information.] A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course instructor, academic advisor, or the undergraduate associate dean. For information on categories of offences and types of penalties, students should refer to [Policy 71, Student Discipline](#). For typical penalties, check [Guidelines for the Assessment of Penalties](#).

Appeals: A decision made or penalty imposed under [Policy 70, Student Petitions and Grievances](#) (other than a petition) or [Policy 71, Student Discipline](#) may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to [Policy 72, Student Appeals](#).

Note for students with disabilities: [AccessAbility Services](#), located in Needles Hall, Room 1401, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with AccessAbility Services at the beginning of each academic term.

Turnitin.com: Text matching software (Turnitin®) may be used to screen assignments in this course. Turnitin® is used to verify that all materials and sources in assignments are documented. Students' submissions are stored on a U.S. server, therefore students must be given an alternative (e.g., scaffolded assignment or annotated bibliography), if they are concerned about their privacy and/or security. Students will be given due notice, in the first week of the term and/or at the time assignment details are provided, about arrangements and alternatives for the use of Turnitin in this course.

It is the responsibility of the student to notify the instructor if they, in the first week of term or at the time assignment details are provided, wish to submit alternate assignment.