Stat 462/862 Computational Data Analysis: Course outline

Course website

http://www.mast.queensu.ca/~aaron/stat462
All assignments and course information will be distributed on the course website.

Instructor

Aaron Springford
Jeff room 105
aaron.springford@queensu.ca

Class times

Mondays 10:00 - 11:20 Jeff 155 lab
Wednesdays 8:30 - 9:50 Jeff 155 lab

Office hours

Mondays 11:30 – 12:30 or by appointment (Monday-Wednesday)

Course goals

Students in Statistics 462/862 will develop their skills in two important areas:

1. Applied Data Analysis
2. The use of statistical software R and SAS

This type of skill development cannot be accomplished by top-down delivery of material (i.e. traditional lecturing); rather, students should be prepared to learn by doing. This might mean that a good amount of effort is required to complete course requirements. Don't be discouraged; you will be learning two new (statistical programming) languages and this takes a certain amount of commitment. But I think you will find the end result extremely rewarding and worth the effort.

Grading

- 20% lab quizzes
- 30% final exam
- 50% data analysis project

All components of this course will receive numerical percentage marks. The final grade you receive for the course will be derived by converting your numerical course average to a letter grade according to Queen's Official Grade Conversion Scale:

Queen's Official Grade Conversion Scale **Grade**	**Numerical Course Average (Range)**
A+	90-100
A	85-89
A-	80-84
B+	77-79
B	73-76
B-	70-72

C+	67-69
C	63-66
C-	60-62
D+	57-59
D	53-56
D-	50-52
F	49 and below

Textbooks

- Data analysis and graphics using R by Maindonald and Braun 3rd Ed.
- SAS for data analysis: Intermediate statistical methods by Marasinghe and Kennedy
    - This book is available online through the Queen's library via SpringerLink. If you want, you can order a copy for $25 via SpringerLink.
- The little SAS book: A primer 5th Ed.
    - This book is available online through the Queen's library.

Additional materials

- The R Inferno by Burns available at http://www.burns-stat.com/pages/Tutor/R_inferno.pdf
- Using R for Data Analysis and Graphics: Introduction, Code and Commentary by Maindonald available at http://cran.r-project.org/doc/contrib/usingR.pdf
- R is free and available at http://cran.r-project.org/
- SAS is not free and a license that will work until April 30th is available at http://www.queensu.ca/its/software/sas/getlic.html

Lab quizzes

The lab quizzes will be held during either the Monday or Wednesday lab sessions, or both. The purpose of the quizzes is to reinforce concepts learned in class. Each quiz will include a mark for individual performance (determined by your peers) and a mark for group performance (determined by me). If you miss a lab quiz you will not have the opportunity to make it up. However, there will be twelve (12) quizzes and the lowest two (2) marks will be dropped.

Exam

There will be a final exam on the topics covered in the course during the regular exam period. The purpose of the exam is to assess whether you have developed a basic level of understanding in data analysis, R, and SAS.

Data analysis project

The data analysis project will have three parts:

1. Model parameter estimation
2. Response curve approximation
3. Prediction

Each of these parts represents a different goal of data analysis. You will be graded on your performance relative to the performance of your classmates as well as your reasoning and ability to communicate your results. Don't delay work on the data analysis project; late projects

will not be accepted except under extraordinary circumstances (at my discretion). This should not be an issue because of the long timeline for completion of the project.

Disabilities

Queen's University is committed to achieving full accessibility for persons with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities to ensure they have an equitable opportunity to participate in all of their academic activities. If you are a student with a disability and think you may need accommodations, you are strongly encouraged to contact the Disability Services Office (DSO) and register as early as possible. For more information, including important deadlines, please visit the DSO website at: http://www.queensu.ca/hcds/ds/. Students with disabilities who require accommodations are asked to make an appointment to see the instructor as soon as possible.

Academic integrity

Academic integrity is constituted by the five core fundamental values of honesty, trust, fairness, respect and responsibility (see www.academicintegrity.org). These values are central to the building, nurturing and sustaining of an academic community in which all members of the community will thrive. Adherence to the values expressed through academic integrity forms a foundation for the "freedom of inquiry and exchange of ideas" essential to the intellectual life of the University (see the Senate Report on Principles and Priorities http://www.queensu.ca/secretariat/policies/senateandtrustees/principlespriorities.html).

Students are responsible for familiarizing themselves with the regulations concerning academic integrity and for ensuring that their assignments conform to the principles of academic integrity. Information on academic integrity is available in the Arts and Science Calendar (see Academic Regulation 1 http://www.queensu.ca/artsci/academic-calendars/2011-2012-calendar/academic-regulations/regulation-1), on the Arts and Science website (see http://www.queensu.ca/artsci/academics/undergraduate/academic-integrity), and from the instructor of this course. Departures from academic integrity include plagiarism, use of unauthorized materials, facilitation, forgery and falsification, and are antithetical to the development of an academic community at Queen's. Given the seriousness of these matters, actions which contravene the regulation on academic integrity carry sanctions that can range from a warning or the loss of grades on an assignment to the failure of a course to a requirement to withdraw from the university.

I encourage discussion of the assignments between students, but each student must write up and submit their own work. The use of any reference material including books, articles, websites, other students, or professors must be attributed.

Course content

Date	Topic	Readings	Additional
Sep 8 2014	Introduction to R	DAAG Chapter 1 usingR.pdf	Lab 1
Sep 10 2014	Introduction to R	DAAG Chapter 1 usingR.pdf	
Sep 15 2014	Introduction to SAS	SDA Chapter 1-3 LSB Chapters 1-5, 8	Lab 2
Sep 17 2014	Introduction to SAS	SDA Chapter 1-3 LSB Chapters 1-5, 8	
Sep 22 2014	Styles of data analysis	DAAG Chapter 2	Lab 3
Sep 24 2014	Styles of data analysis	DAAG Chapter 2	
Sep 29 2014	Statistical models	DAAG Chapter 3	Lab 4
Oct 1 2014	Statistical models	DAAG Chapter 3	
Oct 6 2014	Inference concepts	DAAG Chapter 4	Lab 5
Oct 8 2014	Inference concepts	DAAG Chapter 4	
Oct 13 2014	Thanksgiving	NA	NA
Oct 15 2014	Regression	DAAG Chapter 5 & 6	Lab 6
Oct 20 2014	Regression	DAAG Chapter 5 & 6	
Oct 22 2014	Regression	DAAG Chapter 5 & 6	
Oct 27 2014	Extending the linear model	DAAG Chapter 7	Lab 7
Oct 29 2014	Generalized linear models	DAAG Chapter 8	
Nov 3 2014	Classification and regression trees	DAAG Chapter 11	Lab 8
Nov 5 2014	Classification and regression trees	DAAG Chapter 11	
Nov 10 2014	Multivariate methods	DAAG Chapter 12	Lab 9
Nov 12 2014	Multivariate methods	DAAG Chapter 12	
Nov 17 2014	Multi-level models	DAAG Chapter 10	Lab 10

Nov 19 2014	Multi-level models	DAAG Chapter 10	
Nov 24 2014	Topics in programming	SAS Macro Programming for Beginners	Lab 11
Nov 26 2014	Course wrapup / TBA		Final project due

Final note

Please come prepared to class by looking over the readings beforehand and by working on the labs on your own time as necessary. This is important not only for your own understanding, but also in case of an in-class quiz!