

STA302H1F / 1001HF – Methods of Data Analysis I

Summer 2014

Lectures: Mondays and Wednesdays 6-9 p.m. in SS 2117
Tutorials: Mondays and Wednesdays 5-6 p.m. in SS 1091
During the tutorials, a teaching assistant will be available to answer questions.
Tests and assignments will be returned in this hour.

Instructor: Craig Burkett
E-mail: burkett@utstat.utoronto.ca
Web-page: <http://portal.utoronto.ca> (U of T Blackboard)
Office: SS 6025
Office hours: Mondays and Wednesdays 4 – 5 p.m. or by appointment.

Overview: This course covers some of the theory and methodology of data analysis when linear regression models are appropriate. Topics to be covered include: initial examination of data, correlation, simple and multiple regression models using least squares estimation, inference for regression parameters under assumptions of normally distributed errors, confidence and prediction intervals, diagnostics and remedial procedures when model assumptions are violated, interaction and dummy variables and model selection. This course will also be an opportunity to begin to develop skills in data analysis for which the SAS software program will be taught.

Textbook: The textbook is *Applied Linear Regression Models*, 4th edition by Kutner, Nachtsheim & Neter. We will be covering most of Chapters 1 through 8 and selected material from chapters 9 and 10, as time permits. This is a good textbook and worth the read, although it is not required for the course. I still have a copy of it on my bookshelf.

Another good textbook is *A Modern Approach to Regression with R* by Simon J. Sheather. It is currently available online (as an e-Book) through the library website. We will be covering most of Chapters 1 through 7, excluding 4. Topics in later chapters will be covered in STA 303H1. This book was used previously in the course, and the notation is noticeably different so be careful if you read it.

Datasets and other resources for Sheather are available at the textbook website <http://www.stat.tamu.edu/~sheather/book/>

Pre-requisites

Students should have a second year statistics course such as STA 248H1 / STA 261H1 / STA255H1 or ECO 227Y. I expect that you have knowledge of Appendix A (up to the end of A.7), for example. There is also a document posted on Portal (courtesy of A. Gibbs) for those who cannot see the Appendix.

Students are also expected to have the mathematics pre- and co-requisites required by students in these second-year statistics courses. You will need to know basic matrix operations. A good review of the matrix algebra that we will need can be found in the first 10 pages of this document.

<http://www.stat.ncsu.edu/people/davidian/courses/st732/notes/chap2.pdf>

Most applied courses in the Statistics Department require STA 302H1 as a pre-requisite. As a consequence, this course has a theoretical component to prepare students for more advanced work. Please do not attempt the course without the required mathematical background.

Follow-up courses

STA 303H1 (Methods of Data Analysis II) focuses on aspects of linear models that are not covered in STA 302H1 such as non-normal and correlated response variables.

Evaluation

The grading scheme is as follows:

Test 1	15%	Wednesday, May 28 from 6:15pm – 7:45pm. Location: TBA.
Assignment 1	10%	Due at the start of lecture on Wednesday, June 4.
Test 2	15%	Monday, June 16 from 6:15pm – 7:45pm. Location: TBA.
Assignment 2	10%	Due at the start of lecture on Wednesday, June 18.
Final exam	50%	During exam period.

There will be a one-hour lecture after each test. If your exam mark is better than your term mark (including the exam), the exam weight will be 70% and the test weights will be 5%. The test room will be posted on the course website prior to the test. The assignments will involve both theoretical questions and a data analysis project for which you will use SAS. You will not need to know SAS syntax on the test and exam, but you will need to interpret output from SAS.

No late assignments will be accepted without documentation of a valid reason.

STA 1001 students should speak to me regarding an optional adjustment to the marking scheme.

Practice Problems

Reading material and practice problems for each chapter will be posted on the website. They are **not** to be handed in. They will be appropriate preparation for the test and exam.

Professor Contact

There are various ways in which the TAs and I would be happy to serve you. Here are some rough guidelines:

- If you have a personal issue that you believe I can resolve in a few minutes, please speak to me before or after lecture, or during a break. You can also come to office hours if you require more time or privacy.
- If you'd like to discuss the class material in more depth, please come to office hours. You can also try me after class or at a break, but priority will be given to above.
- If you'd like to discuss the solution to homework questions, please see the TA first. They will probably be more familiar than me with specific questions. If you're not satisfied with their answers, please come to office hours.
- If you have an issue that can be dealt with in three sentences of text or less, or to report a problem with Portal or the assigned homework, or to inform me of something relevant to the course (such as a missed test), please send me an email.

NB: I don't check email constantly as, believe it or not, I don't have a mobile phone. Further, I don't really like sending/receiving emails, and would much prefer that you speak to me in person. That said, if you believe an email is appropriate, please email me using your *.utoronto.ca or *.mail.utoronto.ca address. You won't get a response if you email from other email addresses, and it probably won't even be read since my spam filter may block it. The reason for this is so that I know whom I am writing to, and so that I don't provide any personal information to someone who shouldn't be receiving it.