

STA302H1F / 1001HF – Methods of Data Analysis I
Fall 2014

Lectures: Tuesdays 10am-12pm
Thursdays 10am-11am in ES1050

Tutorials: Thursdays 11am-12pm in WB 116
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 6024A

Office hours: Tuesdays and Thursdays 3-4pm. There will be additional hours around the final.

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, measurement error 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:

Assignment 1	5%	Due at the start of tutorial on Thursday, Oct. 2.
Mid-term Test	25%	Tuesday, Oct 7 from 10:10am – 12:00pm. Location: TBA.
Assignment 2	10%	Due at the start of tutorial on Thursday, Oct 30.
Assignment 3	10%	Due at the start of tutorial on Thursday, Nov 27.
Final exam	50%	During exam period.

If your exam mark is better than your term mark (including the exam), the exam weight will be 60% and the test weight will be 15%. 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 want to ask a question about the course content, a practice problem, an announcement that was made in class but you missed it because you were not present or not listening, please use the discussion forum on Portal.
- If you have an issue that must be dealt with by me, and can be handled 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.
 - If your email can be answered by reading this syllabus or the Portal discussion forum, I will not answer it. Please don't be offended.

NB: I don't check email constantly as, believe it or not, I don't have a mobile phone. I also teach over 800 students this term, and cannot handle the volume of emails that come through with that number. 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. Also, please put "STA302: " at the start of your subject, as I teach multiple courses most terms.

Important Notes

- If a test is missed for a valid reason, you must provide appropriate documentation, such as the University of Toronto Medical Certificate, University of Toronto Health Services Form, or College Registrar's Letter. You must submit this documentation to the course Instructor within one week of the test. Print on it your name, student number, course number and date, and have the doctor record the reason for the visit. No notices will be accepted without a CPSO number stamped on the form (ie. They must be a doctor in the western sense of the word).
- If documentation is not received in time, your test mark will be zero. If the test is missed for a valid reason, I reserve the right to force you to makeup the test OR to shift the weight to the final exam, at my discretion. Most likely you will be writing the makeup test.
- Any requests to have marked work re-evaluated must be made **in writing** within one week of the date the work was returned to the class. The request must contain a justification for consideration; do not simply write "please see #3".
- The course website will be used to post lecture notes, SAS examples used in lectures, practice problems, assignments and solutions, past tests/exams, other course info and important announcements. **Check it regularly**. The website also has an electronic discussion forum that you can use to communicate with other students in the course.
- If an urgent matter arises, I may contact the entire class by email. In order to receive these messages, please make sure that your ROSI account has your **utoronto.ca** email.
- In general, I am not able to answer questions about the course material by email. Before sending an email, make sure that you are not asking information that is already on the course website, or questions about the course material or assignment that are more appropriate to discuss during office hours.
- Questions about the course material can be posted on the discussion board on Blackboard. Other students may be able to answer your questions very quickly, and the TAs will check the board as well.

Computing

We will be using SAS which is available on the CQUEST system. CQUEST computer labs are available in Ramsey Wright (RW). To get an account, go to www.cquest.utoronto.ca. There you will also find information about using CQUEST. Students enrolled in STA 1001 should see me to get an account.

CQUEST is also accessible remotely by SSH. See the CQUEST website for details.

A license for SAS for personal use can be purchased from the Information Commons Licensed Software Office. Go to www.utoronto.ca/ic/softdist for more information.

I am assuming that students have never used SAS before. I will provide you with the SAS syntax for the examples in lecture, which should be sufficient for you to do your assignments. Note that there are many graphics options available to produce the sophisticated plots that are in the book, but we will focus on the basics. There are also many good books on learning SAS. Here are three suggestions if you'd like other references:

- R.P. Cody and J.K. Smith, *Applied Statistics and the SAS Programming Language*. This is my favorite book for learning SAS.
 - R.J. Freund and R.C. Littell, *SAS System for Regression*. Includes everything we need for this course and a lot more. Assumes you already know the basics of SAS.
 - SAS reference manuals are available electronically through the University of Toronto library web site. Under e-Resources select e-Books and then search for SAS.
- There are also many online tutorials and reference pages available on the internet.

Academic Integrity

You are responsible for knowing the content of the University of Toronto's Code of *Behaviour on Academic Matters* at www.artsci.utoronto.ca/osai/students.

If you have any questions about what is or is not permitted in this course, please do not hesitate to contact me. It is legitimate to discuss assignment problems with other students in the class, however, assignments must be written up completely by yourself. Do not let other students read your completed assignment solutions as this can lead to copying. Failure to comply with this is a serious academic offense.