

STA 302 H1F / 1001 HF – METHODS OF DATA ANALYSIS I

Fall 2011

- Lectures:** Tuesdays 10:10–11:00 in OI G162 (no lecture on Tuesday, November 8 (Fall Break))
Thursdays 10:10-12:00 in OI G162
- Tutorials:** Tuesdays 11:10–12:00 in OI G162
During most tutorials, a teaching assistant will be available to answer questions.
Tests and assignments will be returned in this hour.
This hour may occasionally be used for lecture. Announcements will be made in advance.
- Instructor:** Dr. Alison Gibbs
- E-mail:** alison.gibbs@utoronto.ca
- Office:** SS 5016A
- Telephone:** (416) 946-7589
- Office hours:** Thursdays 12:10–14:00 (priority to STA 302/1001 students)
Mondays 11:10–13:00 (priority to STA 220 students)
Appointments for other times can be made by e-mail.
More office hours will be scheduled before tests and assignment due dates.
- Tutorial teaching assistant:** Andriy Derkach
- Course web pages:** Announcements and a discussion board are available through <https://portal.utoronto.ca>
Lecture notes, SAS examples, practice problems, assignments and previous exams are at <http://www.utstat.utoronto.ca/alisong/Teaching/1112/Sta302/sta302.html>

Course content

This course covers how to analyse data when regression models are appropriate. Although the focus is on the methodology of data analysis, some underlying theory will also be developed. Topics to be covered include: initial examination of data, correlation, simple and multiple regression models using least squares, inference for regression parameters for normally distributed errors, confidence and prediction intervals, diagnostics and remedial measures when the model assumptions are violated, interactions and dummy variables, model selection, least squares estimation and inference for non-linear regression.

Pre-requisites

Students should have a second year statistics course such as STA 248H1 / STA 255H1 / STA 261H1 / ECO 227Y. Students are expected to also 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 review of matrix algebra is available on the textbook website at <http://www.stat.tamu.edu/~sheather/book/tutorials.php>.

Follow-up courses

STA 303H1 (Methods of Data Analysis II) focusses on aspects of linear models that are not covered in STA 302H1 such as non-normal and correlated response variables. Most applied courses in the Statistics Department require STA 302H1 as a pre-requisite.

Textbook

The textbook is *A Modern Approach to Regression with R* by Simon J. Sheather. We will be covering most of chapters 1 through 7. Topics in later chapters will be covered in STA 303H1. The book is available to purchase and is also available as an e-Book through the University of Toronto library website.

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

Another good reference for the course material is *Applied Linear Regression Models*, 4th edition by Kutner, Nachtsheim, and Neter. A copy of this book is on reserve at the Mathematical Sciences Library.

Course web site

The course web site

<http://www.utstat.utoronto.ca/alisong/Teaching/1112/Sta302/sta302.html> will be used to post lecture notes, practice problems, SAS examples used in lecture, assignments, and previous exams. It is recommended that you print copies of the SAS examples to bring to lecture. The lecture on Tuesday, September 20 will be an introduction to SAS. There will be notes posted on the web site that you may wish to print and bring to this lecture. When looking at previous exams, resist the temptation to look at the solutions until you have tried the questions and spent at least a day thinking about solutions to problems of which you are unsure.

Blackboard will be used for announcements. You are also encouraged to use the discussion board on Blackboard.

Accessibility Needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible:

disability.services@utoronto.ca or <http://studentlife.utoronto.ca/accessibility>.

Communication

In general, I am not able to answer questions about the course material by e-mail. Before you send an e-mail, make sure that you are not asking for information that is already on the course web site, or questions about the course material or assignments that are more appropriately discussed in tutorial or during office hours. If you do not get a response, this may be why. Questions about the course material can be posted on the discussion board on Blackboard. I will check it at least once every 2 days.

E-mail is appropriate for private communication. Use your `utoronto.ca` account to ensure that your message doesn't automatically go to my Junk folder. I will generally answer e-mail within one business day.

Announcements will be posted on Blackboard. Please check there regularly. If an urgent matter arises, I may contact the entire class by e-mail. In order to receive these messages, please make sure that your ROSI account has your `utoronto.ca` e-mail.

Evaluation

	Weight	Date	Time	Location
Term Test	24%	Thursday, October 20	10:10-11:40 a.m.	TBA
Assignment 1	7%	Tuesday, October 11	due at 12:00	
Assignment 2	7%	Thursday, November 10	due at 12:00	
Assignment 3	7%	Tuesday, November 29	due at 12:00	
Exam	55%	scheduled by Faculty		

If your exam mark is better than your test mark, the exam weight will be 67% and the test weight will be 12%.

The test will be written in a room other than the lecture room (location to be announced). Practice problems will be posted on the web for each chapter. The practice problems are to help you prepare for the tests and exam and are not to be handed in. The assignments will each be a data analysis project for which you will use SAS. You will not need to know SAS syntax on the tests and exam, but you will need to interpret output from SAS.

If the 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 within one week of the test. If documentation is not received in time, your test mark will be zero. If a test is missed for a valid reason, its weight will be shifted to the final exam.

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

Any requests to have marked work re-evaluated must be made *in writing* within two weeks of the date the work was returned to the class. The request must contain a justification for consideration.

STA 1001 students should e-mail me regarding an optional adjustment to the marking scheme.

Computing

We will be using SAS which is available on the CQUEST system. CQUEST computer labs are available in Ramsey Wright. To get an account, go to www.cquest.utoronto.ca. There you will also find information about using CQUEST. There are a few hours each week when all three computer labs are booked. There are always a few terminals running Linux available in RW 211 but you should avoid the labs during times they are booked by other courses. Students enrolled in STA 1001 should see me to get a CQUEST 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/software/index.html for more information.

I am assuming that students have never used SAS before. I will provide you with the SAS syntax for all of the examples in lecture, which should be sufficient for you to do your assignments. At the textbook website, you will find a SAS primer containing code to reproduce all of the analyses and plots in the book. 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. Under the Tutorials link at the textbook website there is a link to SAS tutorials, some of which you may find interesting. 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 favourite 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** search on SAS and then select **e-Books**.

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

<http://www.governingcouncil.utoronto.ca/policies/behaveac.htm>.

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.