STA302H1F / 1001HF – Methods of Data Analysis I

Summer 2012

Lectures: Mondays and Wednesdays 6-9 p.m. in PB B250

Instructor: Dr. Hadas Moshonov

E-mail: hadas@utstat.utoronto.ca

Web-page: Available through http://portal.utoronto.ca

Office: SS 6015

Phone: 416-978-4308

Office hours: TBA.

Course Description

This course covers some of the theory and methodology of data analysis when 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.

Textbook

The textbook is *A Modern Approach to Regression with R* by Simon J. Sheather. We will be covering most of chapters 1 through 6. 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/

Pre-requisites

Students should have a second year statistics course such as STA 248H1 / STA 261H1 / STA 255H1 or 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) focuses 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.

Evaluation

The grading scheme is the following:

Term Test	30%	To be held on Wednesday, May 30, 6:10 – 7:40 p.m. (room TBA).
Assignment	15%	Due on Monday, June 11 in the beginning of the lecture.
Final exam	55%	To be held on Wednesday, June 27, 6 – 9 p.m. (room TBA).

If your exam mark is better than your test mark, the exam weight will be 70% and the test weight will be 15%. The test room will be posted on the course website prior to the test.

The assignment will involve both theoretical questions and 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 outputs from SAS.

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

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

Practice Problems

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

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. Students enrolled in STA 1001 should see me to get an account. CQUEST is also accessible remotely by ssh. See 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 assignment. 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 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.

Important Notes

- There is no makeup test. 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 to the course Instructor (Hadas Moshonov) or the Departmental Office (SW6018) within one week of the test. Print on it your name, student number, course number and date. If documentation is not received in time, your test mark will be zero. If the test is missed for a valid reason, its weight will be shifted to the final exam.

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

- The course web site 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 bulletin board that you can use to communicate with other students in the course.

- 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.

- In general, I am not able to answer questions about the course material by e-mail. 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. I will check it at least every 2 days.

- The lecturer and TAs are there to help. Ask questions and alert us to any problems right away!

Academic Honesty Policy

All work submitted for credit must be your own individual effort.
University of Toronto's academic integrity policy states that "Honesty and fairness are considered fundamental values shared by students, staff and faculty at the University of Toronto. The University's policies and procedures that deal with cases of cheating and plagiarism are designed to protect the integrity of the institution. As a result, the University treats cases of cheating and plagiarism very seriously. Any student accused of committing an academic offence will find that the accusation is dealt with formally and that the penalties can be severe if it is determined that they did, in fact, cheat." For more information visit http://www.utoronto.ca/academicintegrity.

Tentative Work Schedule

Week Date	Topic	Reading
1 May 14	Introduction to regression models. Simple linear regression models and least square estimates.	Chapter 1. Chapter 2, section 2.1.
2 May 16	Assumptions for SLR models. Properties of least square estimates. Gauss-Markov Theorem.	Chapter 2, section 2.2.
3 May 21	**Victoria day – no class**	
4 May 23	Introduction to SAS. Review of distribution theory for *t*-tests and confidence intervals.	SAS lecture notes.
5 May 28	Inference about the slope and intercept of the regression line. Confidence interval for the population regression line. Prediction interval. Decomposition of sums of squares – Analysis of Variance.	Chapter 2, sections 2.3-2.5, 2.7.
6 May 30	**Term test + short lecture**	
7 June 4	Dummy variable regression. Diagnostics for simple linear regression. Residuals and residual plots.	Chapter 2, section 2.6. Chapter 3, sections 3.1, 3.2
8 June 6	Dealing with violation of assumptions. Transformations. Random vectors.	
9 June 11	Linear regression in matrix form. Multiple regression models. **Assignment is due.**	Chapter 3, sections 3.3
10 June 13	More on multiple regression models. Regression models with interaction.	Chapter 5.
11 June 18	Diagnostics for Multiple regression models. Multicolinearity.	Chapter 6, parts of sections 6.1 and 6.4
12 June 20	Diagnostics for Multiple regression models. Review and final summary.	