

## STA302H1F/1001HF: Methods of Data Analysis I, Fall 2013

*Please be aware that the information provided below is subject to change.*

### **Class Information:**

- Lectures are given on Tues 11:00–11:50am and Thurs 10:00–11:50am in MC102.
- Tutorials are held on Tuesday 10:00–10:50am in MC102. The TA will be present to hold tutorials on R introduction and examples, practice problems and assignments, and give quizzes.
- Students are primarily expected to attend all classes and tutorials.
- Instructor: Fang Yao, [fyao@utstat.toronto.edu](mailto:fyao@utstat.toronto.edu), SS6027B, <http://www.utstat.utoronto.ca/fyao/sta302/sta302.html>
- Office hour: Tuesday 1:00-3:00pm.

**Course Description:** 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, hypothesis testing, diagnostics and remedial procedures when model assumptions are violated, interaction and dummy variables and model selection.

**Computing:** We will be using R for computing and data analysis throughout the course. R is a high level language especially designed for statistical calculations, and it is freely available at <http://www.cran.r-project.org/>. I am assuming that students have never used R before. The tutorials held by the TA will cover R introductions and examples, which will be sufficient for you to work on your assignments.

**Prerequisites:** Students should have a second year statistics course, such as STA248H1/STA261H1/STA255H1/ECO220Y1/ECO227Y1. 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 calculus and matrix operations.

**Follow-up Course:** STA303H1 (Methods of Data Analysis II) focuses on aspects of linear models that are not covered in STA302H1 such as non-normal and correlated response variables. Most applied courses in the Statistics Department require STA302H1 as a pre-requisite.

**Course Evaluation:** Assignments & Quizzes (30%), Term Test (25%) and Exam (45%).

I will post practice problems (with solutions) which the quizzes will be based on. The assignments will involve both theoretical questions and data analysis projects for which you will use R. You will not be asked to write R syntax on the quizzes, test and exam, but you will need to interpret outputs from R.

**Textbook and Resources:** *Applied Linear Regression*, 3rd edition (2005), by Sanford Weisberg, Wiley/Interscience. You may access the electronic copy of the textbook from U of T library (choose “Wiley-Blackwell Online Books - Full Collection” for a pdf version). Here is the link to the author’s website on this book: <http://www.stat.umn.edu/alr/>, which contains useful information, such as datasets used in the book. Also listed are two other good references on linear models at a similar level (not required), they are also available electronically from the U of T library.

- *A Modern Approach to Regression with R*, Sheather, S. J., Springer, 2009.
- *Linear Models with R*, Faraway, J. J., Chapman & Hall/CRC, 2005.

**Blackboard:** We will use Blackboard throughout the course. Please log into Blackboard using your UTOR ID and keep updated regarding class information and material. 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 email.

#### Important Notes:

- **There is no makeup quiz/test.** If a quiz or 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 (Fang Yao) or the Departmental Office (SS6018) **within one week of the quiz or test**. Print on it your name, student number, course number and date. If the quiz or test is missed for a valid reason, its weight will be shifted to the final exam. If the documentation is not received in time, your mark will be zero.
- You must show the necessary work to receive full credit for any problem, and any work turned in must be your own. Any request 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.
- 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 Blackboard, or questions about the course material that are more appropriate to discuss during office hours and/or tutorials. If you do not receive my response in a timely manner, this would likely be the reason.
- **Academic honesty policy:** All work submitted for credit must be your own individual effort. University of Torontos academic integrity policy states that “*Honesty and fairness are considered fundamental values shared by students, staff and faculty at the University of Toronto. The Universitys 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 offense 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>.