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

Winter 2018

Lectures: Tuesdays 10:10 am - 12:00 pm in HS 610 and Thursdays 10:10 am - 11:00 am in HS 610.

Instructor: Alfred Benn, PhD.

E-mail: alfred.benn@gmail.com
Office: SS 1091
Telephone: (416) 553-6461

Instructor Office Hours: Thursdays 1 pm - 4 pm. Appointments for other times can be made via e-mail or in class. Additional office hours will be scheduled before the mid-term test and data analysis project due dates.

Teaching Assistants (TAs): Yuenan (Joseph) Cai, Xuan (Henry) Gu, Anthony Guerriero.

TA Office Hours: Joseph Mondays 1 pm - 2 pm (SS 623B); Henry Tuesdays 2 pm - 3 pm (SS623B); Anthony Wednesdays 1 pm - 2 pm (SS 623B).

Course announcements and a discussion board will be available through **https://portal.utoronto.ca** web pages: examples, practice problems, previous exams.

Course Content: This course involves the analysis of data when regression models are appropriate. Although the focus is on the methodology of data analysis, some underlying theory will also be developed. Core 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 and inference for non-linear regression. Additional topics may include logistic regression, generalized additive models, regression trees, random forests and neural networks.

Evaluation: Term Test 35%; Data Analysis Project 15%; Final Exam 50%.

Prerequisites: Students <u>must</u> 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) focus 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 prerequisite.

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/.

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.caorhttp://studentlife.utoronto.ca/accessibility.

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 the instructor. It is legitimate to discuss practice problems and the data project with other students in the class. However, the data analysis project report must be written independently by each student.