

STA 2101F: Methods of Applied Statistics I

Thursday, 12-3 pm Eastern

September 10 – December 3 2020

Updated September 3

From the calendar:

This course will focus on principles and methods of applied statistical science. It is designed for MSc and PhD students in Statistics, and is required for the Applied Paper of the PhD comprehensive exams. The topics covered include: planning of studies, review of linear models, analysis of random and mixed effects models, model building and model selection, theory and methods for generalized linear models, and an introduction to nonparametric regression. Additional topics will be introduced as needed in the context of case studies in data analysis.

Prerequisites: ECO374H1/ECO375H1/STA302H1 (regression); STA305H1 (design of studies)

Course Delivery:

The class will be delivered at the scheduled time (Thursdays, 12-3 pm Toronto time) using BBCollaborate. The lectures will be recorded, for viewing offline after the scheduled time. The slides for the lectures will be posted, on good weeks before the scheduled course time, and on rushed weeks just after.

The first hour will be mainly lecture-style, with breaks for discussion. The second hour will be a discussion of case studies, usually from current events, with statistical concepts reviewed as needed in that context. The third hour will be a discussion of computational methods and/or problems, questions about the course material and questions about the homework.

We will use Piazza for discussion, as it is now integrated with Quercus. You will find an entry for Piazza in the course menu. If you click it, you will be asked to sign up. Please see the instructions in the [handout](#), especially the highlighted bits.

Grading:

The grade in the course will be based on three homework sets (60%) and a take-home final test (40%). Tentative due dates for homework sets are 11.59 pm (Eastern) on: Oct 1, Oct 29, Nov 26. The take-home final released December 4, due December 11 (tentative).

Academic Integrity:

Discussion about your work with your classmates is encouraged, but the homework solutions you submit must be written, and coded, independently. You may use code provided by me without attribution, but you must acknowledge code taken from any other source using a proper bibliographic reference. To protect yourself

from potential academic integrity offences, do not share your code and written submissions. The University of Toronto's Code of Behaviour on Academic Matters is available at <http://academicintegrity.utoronto.ca>.

Computing:



I will always refer to the R computing package and I highly recommend the RStudio environment. You will need to install both of these on your laptop. I am using Version 3.6.3 of R, although Version 4 was released in April 2020. I am using Version 1.3.1073 of RStudio. You can download R from <http://cran.utstat.utoronto.ca/> and the free Desktop Version of RStudio from <https://rstudio.com/products/rstudio/#rstudio-desktop>.

I also strongly recommend using R Markdown to prepare your homework, but you can use LaTeX or Word if you must. For questions involving computing you will need to submit working code. This is easy in R Markdown, but R scripts will also be accepted.

There are many online resources for R and Rstudio. If you are new to R, you could look at [Quick-R](#). RStudio has some recommendations on their [education](#) page. For more experienced users, the [Cheatsheets](#) are invaluable.

Text:

There is no required text, but I highly recommend both *Statistical Models* by A.C. Davison (Cambridge University Press), and *Principles of Applied Statistics* by D.R. Cox and C.A. Donnelly.

The following are helpful for finding R code to fit various models: *Linear Models with R* by J.J. Faraway (Chapman & Hall); *Extending the Linear Model with R* by J.J. Faraway (Chapman & Hall); *Data Analysis and Graphics using R* by J. Maindonald and W.J. Braun (Cambridge University Press); and *An Introduction to Generalized Linear Models* by A. Dobson (Chapman & Hall).

I will give more detailed references during the course.

Course web page(s):

The course materials will be posted to the web page <http://www.utstat.utoronto.ca/reid/2101f20.html>. The Quercus page for STA2101H will lead you to this page. I will use Quercus for discussion (via Piazza), homework reminders and grades management.

Contact: Nancy Reid: reid@utstat.utoronto.ca

Office Hours: TBD

Teaching Assistant: Sangook Kim sangook.kim@mail.utoronto.ca