STA355H1F
Theory of Statistical Practice

Instructor: K. Knight (e-mail: `keith.knight@utoronto.ca`)

Office hours: These will be held on Mondays or Tuesdays (starting September 18) with the time and location to be determined. You should also feel free to e-mail me directly or address questions to the Piazza board.

Piazza: There will be a Piazza site (linked through Quercus) for this course where you can ask questions about course material and other issues related to the course.

Goal: The main goal of this course is to provide students with the necessary tools of mathematical statistics necessary to be a good applied statistician. The focus of the course will be on the theory behind statistical methodology (from exploratory data analysis to formal statistical inference) and there will be a substantial data analytic component.

Textbook: The recommended textbook is *Statistical Models* by A.C. Davison (Cambridge University Press); it is available online through the UofT Library System and a link is provided on the Quercus site. We will not make extensive use of this book although it will serve as a valuable reference in subsequent courses. The textbook will be supplemented with a number of handouts and some journal articles; some of these are already on Quercus and more will be added as the course progresses. Some other good references are:

> D. Nolan and T. Speed: *Stat Labs: Mathematical Statistics Through Applications.* (Springer)
>
> G. James, D. Witten, T. Hastie and R. Tibshirani: *An Introduction to Statistical Learning with Applications in R.* (Springer)

Both books are also available via the University of Toronto Library system as online books.

Computing: To recognize the role of computing in mathematical statistics as well as to emphasize the connections between applied and mathematical statistics, we will use R extensively in this course both for data analysis as well as for carrying out simple Monte Carlo (simulation) experiments. No previous experience is necessary and samples of R code will be provided. R is free software and can be downloaded (for Windows, Mac, and Linux operating systems) from `cran.utstat.utoronto.ca`. Documentation for R can also be found at `www.r-project.org` and this site also lists some books related to R. Of interest to many of you will be RStudio, which provides a very nice environment for using R; information on RStudio (including downloads) can be found at `www.rstudio.com`.

Evaluation: The course grade will be based on four homework assignments (totalling 20%), a midterm exam (35%), and a final exam (45%).

- The assignments will involve both mathematical exercises as well as some computing (using R). Two assignments will be due before the midterm and two after the midterm.

- The midterm exam is scheduled for Wednesday, November 1 during the normal lecture time (1:10–3:00pm). **If you miss the midterm exam due to illness or other circumstances beyond your control, the weight from the midterm will be carried over to the final exam.**

- The final exam will be held during the December exam period.

- Both the midterm and final exams are planned to be in-person exams although this is subject to change.

- **Students should familiarize themselves with the University's policies on academic integrity, which can be found at** `https://tinyurl.com/tsqukhx` .

Syllabus: The following topics will be covered in the course:

Short probability review. Random variables, probability distributions and expected values, convergence in distribution and in probability, related theorems (CLT, WLLN etc), Delta Method.

Statistical models. Parametric and non-parametric models, order statistics, introduction to goodness-of-fit via probability plots, spacings, density estimation, sampling variation and uncertainty in estimation.

Point and interval estimation. Substitution principle, standard errors, jackknife standard error estimates, likelihood estimation, confidence intervals, pivots (exact and approximate), introduction to Bayesian inference, credible intervals, bias/variance tradeoffs (in density estimation and non-parametric regression), robustness, methods for "big data".

Hypothesis Testing. Elements of hypothesis testing, Neyman-Pearson Lemma and its consequences, p-values (and their behaviour under the null and alternative hypotheses), goodness-of-fit testing, multiple testing, ("p-hacking" and false discovery rate), rank tests (if time permits).