Sta437H1 S L0101 2022 Course Information

Practical techniques for the analysis of multivariate data; fundamental methods of data reduction with an introduction to underlying distribution theory; basic estimation and hypothesis testing for multivariate means and variances; regression coefficients; principal components and the partial multiple and canonical correlations; multivariate analysis of variance. The use of R software should be expected.

Instructor Zhou Zhou, Office: 700 University Avenue, Room 9072.
Email: zhou @ utstat.toronto.edu (The best way to reach me is via email).
Office Hours: Mondays 3pm-4pm on Zoom (starting from second week).

TA Tommy Guo.
Email: ygj.guo@mail.utoronto.ca
Collin Decker.
colin.decker@utoronto.ca
Miaoshiqi Liu.
miaoshiqi.liu@mail.utoronto.ca

TA office hours will be posted on Quercus.

Lectures Tuesdays 2pm-5pm from Jan. 11th to April. 5th, EXCEPT Feb. 22nd (Reading Week).

Classes will be taught online via Zoom for the first three weeks. Then the lectures will be held at BA1160.

Textbook Richard Johnson and Dean Wichern, *Multivariate Statistical Analysis*, 6th Edition, Pearson, 2007.

Readings *The Elements of Statistical Learning - Data Mining, Inference and Prediction*, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman, 2nd edition, Springer.

Evaluation Final exam: **55%** (time will be scheduled by the faculty) Cumulative.
Mid-term test: **35%** (Friday, Feb. 18th 7pm-9pm online)
There will be no make-up midterms. If you have to miss the midterm, weights will be shifted to the final exam with valid evidences for absence.
HW: **10%** 4 times. The lowest HW score will be dropped.

Syllables
- Organization of Data and Distances (Chapters 1.3 and 1.5)
- Matrix algebra and random vectors (Chapter 2),
- Sample geometry (Chapter 3),
- The multivariate normal distribution (Chapter 4),
- Multiple linear regression (Chapter 7),
- Principal component analysis, (Chapter 8) and if time permits,
- Factor analysis and Classification (Chapters 9 and 11).

Computing There will be a computing component in this course, and the statistical software R will be used throughout the course. You are also allowed to use other software if it has the same capabilities. However, please be advised that the TA and I may not be familiar with your software of choice resulting in limited assistance.