## Overview

Max Turgeon

STAT 4690-Applied Multivariate Analysis

## Course details

- Time: MWF 10:30am-11:20am
- Office: 373 Machray Hall
- Office Hours:
- Tuesday 9:30am-11am
- Thursday 1pm-2:30pm
- Or by appointment
- Course Website:
https://maxturgeon.ca/f19-stat4690


## Textbook

- Johnson \& Wichern, Applied Multivariate Statistical Analysis. Prentice Hall (2007)
- Recommended, but not required
- A copy is available on reserve at the Science Library
- There are plenty of other textbooks on applied multivariate statistics available. See course website for some recommendations.


## Assessments

- Two assignments each worth $15 \%$ of the final grade
- One midterm (tentatively scheduled October 31 outside of class hours) worth $30 \%$ of the final grade
- There is no final exam
- There is a class project worth $40 \%$ of the final grade


## IN THE CASE OF A FIRE ALARM:

- REMAIN CALM
- IF IT IS SAFE, EVACUATE THE CLASSROOM OR LAB
- GO TO THE CLOSEST FIRE EXIT
- DO NOT USE THE ELEVATORS

IF YOU NEED ASSISTANCE TO EVACUATE THE BUILDING, INFORM YOUR PROFESSOR OR instructor NOW!!!
-IF DURING A BUILDING EVACUATION YOU NEED TO REPORT AN INCIDENT OR A PERSON LEFT BEHIND:
$>$ CONTACT ONE OF THE BUILDING FIRE WARDENS OR
> CALL SECURITY SERVICES 204-474-9341
-DO NOT REENTER THE BUILDING UNTIL
THE "ALL CLEAR" IS DECLARED BY A FIRE WARDEN, SECURITY SERVICES OR THE FIRE DEPARTMENT

## Course Objectives

- Broad overview of techniques used in multivariate analysis, with emphasis on Multivariate Linear Regression and Principal Component Analysis.

1. Make decisions on how and when to use the techniques discussed in class;
2. Apply and assess multivariate methods on real data;
3. Make sound statistical conclusions based on a multivariate analysis.

- Make you competent in the R statistical software.


## Tentative topics

- Aspects of multivariate analysis (Chapter 1)
- Matrix algebra and random vectors (Chapter 2)
- Random Samples (Chapter 3)
- Multivariate normal distribution (Chapter 4)
- Inferences about a mean vector (Chapter 5)
- Multivariate linear regression (Chapter 7)
- Principal Component Analysis (Chapter 8)
- Factor Analysis (Chapter 9)
- Canonical Correlation Analysis (Chapter 10)
- Kernel methods and Manifold Learning (if time permits)


## Multivariate Data

- Multivariate data is everywhere
- Multiple measurements collected a on given experimental unit
- Multivariate analysis is concerned with the relationship between those variables
- Note: Regression with a single outcome variable is not considered multivariate analysis.


## Multivariate Methods

- One- or two-sample inference about multivariate data (think t-test)
- MANOVA: Generalization to several populations
- Multivariate Linear Regression: Linear model for multivariate response in terms of covariates


## Multivariate Methods

- Principal Component Analysis: Reduce dimension of data by finding directions in data with maximal variance
- Factor Analysis: Understand variance in multivariate sample in terms of latent (i.e. unobserved) factors
- Canonical Correlation Analysis: Study correlations between two multivariate datasets


## Multivariate Methods (not covered in STAT 4690)

- Methods for longitudinal data (e.g. mixed-effect models or GEEs)
- Clustering: Grouping "similar" observations based on their (multivariate) measurement (STAT 4600: Statistical Learning)
- Classification and Discrimination: Grouping observations and allocating new units to previously defined classes (STAT 4600: Statistical Learning)
- The difference between the last two is whether or not we measured a class label for the observations.


## Statistical analysis

- This is an applied course, so we will be analysing data
- Although we also require a fair amount of theory
- We will mostly use $R$, and the datasets will be provided to you.
- Code for in-class examples will also be provided
- For assignments and course project, students are strongly encouraged to use Rmarkdown or knitr.
- Literate programming
- Reproducibility

