

Seminar Title: Statistical Modelling and Predictive Analytics in the Era of Big Data

Presenter: Prof. Timothy E. O'Brien, Loyola University Chicago, USA

When: Thursday, January 12th 2017

Abstract: Statisticians and researchers look to determine and establish relationships between variables with an eye to making predictions – both key components of the larger field of predictive analytics. Currently, in fields such as business, marketing, medicine, and telecommunications, massive amounts of data are being collected, often involving “large n ” and/or either “ $p \approx n$ ” or “ $p > n$ ” situations. For example, in multiple linear regression, the corresponding X matrix is of dimension $n \times p$, and ordinary least squares estimation requires the existence of $(X^T X)^{-1}$ and so $n > p$, a requirement not met by typical ‘big data’. This therefore underscores some key issues when handling such big data. This talk reviews modelling techniques such as linear, logistic and nonlinear regression, and focuses on the bias-variance trade-off, classification and resampling methods, linear models selection (including ridge regression and the lasso), spline methods and local regression (generalized additive models), and tree-based methods such as bagging, random forests and boosting as tools for handling big data situations. Taking advantage of the excellent resources in R and R/Studio provided online, this talk will demonstrate the use of these freeware packages. Numerous examples are provided and seminar participants are encouraged to use the techniques introduced and illustrated here by participating in the Titanic Kaggle competition (<https://www.kaggle.com/c/titanic>) to test out and demonstrate their learned data skills.

Key Reference: the following excellent book may be freely downloaded – James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning: with Applications in R, Springer; see <http://www-bcf.usc.edu/~gareth/ISL/getbook.html>

Acknowledgement: This seminar is gratefully facilitated by a Collaborative Research Travel Grant from the Burroughs Wellcome Fund.



Preparing people to lead extraordinary lives