# Course Title: Advanced Quantitative Research Methods

## Hilary Term 2016

## Lecturer/s: David Barron

# Course Readings (by week and/or general)

## Introduction

This class follows on from the Statistics class taught in Michaelmas Term. We begin with some more advanced topics in linear regression, before moving on to methods suitable for the analysis of models with non-interval scale outcome variables, including binary categorical outcomes (e.g., a person is unemployed or not, a firm fails or it doesn’t), ordinal outcomes (e.g. Likert scales), integer count outcomes (e.g., number of new entries into a market in a year). We will then move to thinking about analysis for special kinds of data structure, which often reflect decisions taken at the research design stage. These data structures can have any of the types of outcome variables we have discussed, but we will focus mainly on linear regressions. The two data structures we will consider are: multilevel data (e.g., employees in firms in industries), and longitudinal data (e.g, that derive from panel surveys). Finally, we will consider models that are important in the context of the measurement of unobservable, or latent, constructs. Factor analysis is often used as a tool for exploring and perhaps verifying that a particular scale of measurement (e.g., for employee burnout) works as expected. Structural equation models allow, among other things, for measurement and substantive models to be incorporated into a single analytical model.

Each week we will have a 2 hour seminar dealing with the substantive topic, followed by a 1 hour practical session focussing on the use of the R computer package to implement the methods we will have been discussing. The R package has become one of the most widespread packages for statistical analysis. It is open source software, the main power of which comes from the large number of user-supplied packages that implement a vast array of statistical application. However, for people who have not used a package that requires the use of scripts rather than a Graphical User Interface, it might take a little while to get to grips with its use. The only way to achieve this is practice, and I will expect people to devote an appropriate amount of time to develop their facility with the package between classes.

You may want to check the OUCS course listings, as they run introductory classes in R from time to time if you feel you would benefit from some additional input to that provided in this course. There are also many online resources for learning how to use R, for example: <http://www.ats.ucla.edu/stat/r/seminars/intro.htm>. The first three chapters of the book by John Fox listed below also provides a good introduction. There is no substitute from following along this type of material yourself.

To use R you will also need a text editor. You can just use the basic editor that comes with your computer’s operating system (e.g., Notepad on Windows computers), but I recommend that you also install RStudio, which provides a very easy to use interface between an editor and R itself. This can be downloaded free of charge here: <https://www.rstudio.com/>. You want to download the appropriate version of RStudio Desktop for your computer (i.e. Windows or Mac).

## General readings

UCLA Statistical Computing site: <http://www.ats.ucla.edu/stat/r/> I strongly recommend that all students look at the R starter kit on this site before the first session. It includes instructions for installing R on your computer; you should do this before coming to the first class.

Introduction to R: <http://cran.r-project.org/doc/manuals/R-intro.pdf>

Fox, John. 2011. An R Companion to Applied Regression. Sage. Chapters 1-3 provide an introduction to the use of the R package.

## Weeks One and Two

These two sessions will cover advanced topics in linear regression, including the use and interpretation of models that include dummy variables and interaction terms; diagnosis of and solutions to: heteroscedasticity, non-linearity, collinearity, outliers, sample selection bias.

The associated lab sessions will introduce the R package and will show how these statistical methods can be carried out using this software.

### Further topics in linear regression

* Fox, J. 2011 An R Companion to Applied Regression, chapters 4 and 6 (except section 6.6)
* White, Halbert. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” Econometrica 48: 817838.
* Berk, Richard A. 1983. “An Introduction to Sample Selection Bias in Sociological Data.” American Sociological Review 48: 386-398.

## Weeks Three and Four

### Generalised linear models (including logistic and Poisson regression)

* Fox, J. 2011. An R Companion to Applied Regression, chapters 5-6
* Agresti, Alan. 1990. Categorical Data Analysis. London: PrenticeHall. Chapter 4.
* Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press. Chapter 2.

## Week Five

### Multilevel models

* Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press. Chapters 11-13.

## Week Six

### Longitudinal data analysis

* Singer, Judith D. and John B. Willett. 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press. Chapters 1-4

## Week Seven

### Factor Analysis.

* Agresti, Alan and Barbara Finlay. 1997. Statistical Methods for the Social Sciences. London: Prentice-Hall. Chapter 16.
* Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black. 1999. Multivariate Data Analysis with Readings. London: PrenticeHall. Chapter 7

## Week Eight

### Structural Equation Models

* Kline, R. Principles and Practice of Structural Equation Modelling. Guilford.
* Fox, John. Structural Equation Models.http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-SEMs.pdf
* Long, J. Scott. Covariance Structure Models. Sage.
* Long, J. Scott. Confirmatory Factor Analysis. Sage
* The lavaan project. <http://lavaan.ugent.be>. This is an R package for estimating SEMs that I find easier to use than John Fox’s sem package, but they both do the same thing.