

Advanced Quantitative Methods

Dr. Patrick Murphy
School of Mathematical Sciences
University College Dublin

patrick.murphy@ucd.ie
Library Building, Rm 530
<http://www.ucd.ie/statdept/staff/pmurphy.html>

Jos Elkink
College of Humanities Graduate School
University College Dublin

jos.elkink@ucd.ie
Library Building, Rm 512
<http://jaeweb.cantr.net/teaching>

Spring 2008

Introduction

In this course you will learn the fundamentals of multivariate regression analysis, including diagnostics analysis and extensions towards models with qualitative or limited dependent variables. At the end of the course a number of optional advanced topics will be discussed. Students will acquire a solid understanding of regression analysis, the various models that are available, and some awareness of issues that arise when dealing with actual data instead of theoretical statistics.

A foundational course in statistics, covering topics like variance and correlation, probability distributions, statistical inference, statistical tests, etcetera, is a prerequisite for participation in this course.

For the first semester, the course website is:

<http://www.ucd.ie/statdept/classpages/advancedquantmethods.htm>

For the second semester, the course website is:

<http://jaeweb.cantr.net/teaching/advanced-quantitative-methods/>

Although the course will focus on the theory and main concepts and issues in applied regression analysis, you will also learn how to use the statistical software package R,

which is freely available at <http://www.r-project.org>. You should download and install this at home, so you can get as much hands-on practice as possible.

Please note that in particular for the second semester, this course outline is preliminary. Shortly before the Christmas break a definitive version will be made available.

The course will consist of four distinct parts:

Applied linear regression

The classes in the first semester will focus on the concepts and issues in applied linear regression. How do we estimate a linear regression line? How do we interpret the results? How do we know what is statistically significant? How do we perform diagnostical analysis on the regression results? Although this part of the course will be primarily theoretical, with limited application in R, the amount of mathematics will be low.

Using R

After the Christmas break, one class will be devoted to some more nitty-gritty of using R, including importing data sets from other software, manipulating data, accessing subsets of the data, etcetera. A second class will be taught on using simulation and bootstrapping techniques in R, which will be useful for the next part of the course.

Generalized linear models

The third part of the course, and the core of the second semester, will be devoted to models with limited or qualitative dependent variables. When the dependent variable is binary (e.g. did or did not vote in an election) or has multiple categories (e.g. party voted for), linear regression leads to additional complications, for which various solutions are available. This part of the course will be both theoretical and applied, with only a limited amount of mathematics involved. Since many phenomena we are interested in in the human sciences are categorical in nature, these models are of major importance in applied social science statistics.

Advanced topics in applied regression

The last four weeks of the course will be optional for students taking the course for credit and open to outsiders. Each week a different topic, such as dealing with missing values or time series, will be discussed. These classes will be insufficient to fully apply the material, but will give you a good idea of what the main issues and solutions are, and how to find out more about applying those solutions.

Textbook

There is no one textbook for the course, although the main recommended text is Kutner, Nachtsheim and Neter (2004). The following books we recommend (in order):

Kutner, Nachtsheim and Neter (2004) provides the conceptual foundation of applied linear regression. The text is software-independent and covers all major issues in linear regression models, including simple linear regression, multiple linear regression, drawing statistical inferences, regression diagnostics, etcetera. The authors also published a longer version of the text, which is basically this book, with many chapters on how to design a research before statistical analysis and more advanced statistical issues added (Kutner et al. 2004). This book is more widely available in the library and the first part of it appears identical to the textbook.

Verzani (2005) is an introductory statistics book. The basic material of the book should already be known to participants in this course, but the book demonstrates well how to apply these concepts in R, including basic linear regression and analysis of variance. For students new to R, this book provides a very useful introduction with many examples.

Fox (2002) is a book on using R for applied regression. It is very sparse in terms of the statistical concepts, but useful to find information on how to apply these concepts in R.

Faraway (2002) is less extensive than Fox (2002), but similarly provides example code for applied regression in R, without much elaboration on conceptual issues.

If you really enjoy this course and want to go much deeper into the material, Gujarati (2003) (much more theoretical introduction to linear regression), Gelman and Hill (2007) (applied but much deeper introduction to regression, including limited dependent variables), and Long (1997) (more about limited dependent variables) are good starting points.

Grading

The grading of this course will be based on two hands-on research projects in which you are required to analyse data using regression techniques and report back on them in a paper. The first paper will contain a regular linear regression model, including diagnostics, based on a research question of your own design, using one of the data sets provided. The second paper will contain a regression with a qualitative dependent variable, using a data set of your choosing, and based on your own research question. Details about the exact assignment, somewhat depending on how the course will develop, will be made available at a later stage. The first paper will be due at 5 pm on Friday, February 8; the second at 5 pm on Friday, March 29.

Plagiarism

Although this should be obvious, plagiarism - copying someone else's text without acknowledgement or beyond "fair use" quantities - is not allowed. UCD policies concerning plagiarism can be found at http://www.ucd.ie/regist/documents/plagiarism_policy_and_procedures.pdf. A more extensive description of what is plagiarism and what is not can be found at http://www.ucd.ie/library/students/information_skills/plagiari.html.

Contact

Patrick: If you need to contact me outside class hours, you can find me in room 530 of the Library Building. Since there are no fixed office hours, it is best to make an appointment by email (patrick.murphy@ucd.ie).

Jos: If you need to contact me outside class hours, you can find me in room 512 in the Library Building. I do not have fixed office hours, so if you want to make sure I am present, you can make an appointment by email. If a personal visit is not necessary, the easiest way to reach me is by email (jos.elkink@ucd.ie).

Applied linear regression

Last 6 weeks of the first semester, covering the following topics:

- Concept of linear regression
- Hypothesis testing in linear regression
- Binary explanatory variables
- Interaction models
- Regression diagnostics
- Regression and diagnostics in R

Using R

Week 1: Types of data sets, importing data, and manipulating data

Required reading: Verzani (2005: ch. 1; §4.2) *or* Fox (2002: ch. 1-2).

Week 2: Simulation and bootstrapping

Required reading: Verzani (2005: ch. 6).

Generalized linear models

Week 3: Logit and probit - concept

Week 4: Logit and probit - interpretation and presentation

Week 5: Ordinal probit

Week 6: Multinomial logit

Week 7: Count data

Week 7: Tabular data

Advanced topics in applied regression

This part is not for credit and optional, but of course highly recommended, since the issues discussed here are very common in applied statistical analysis.

Week 8: Missing data

Week 9: Fixed and random effects

Week 10: Time-series models

Week 11: Causation and inference

References

Faraway, Julian J. 2002. *Practical regression and anova using R*.
<http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>

Fox, John. 2002. *An R and S-Plus companion to applied regression*. Sage Publications.

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multi-level/hierarchical models*. Analytical Methods for Social Research Cambridge: Cambridge University Press.

- Gujarati, Damodar N. 2003. *Basic econometrics*. 4th ed. Boston: McGraw-Hill.
- Kutner, Michael H., Christopher J. Nachtsheim and John Neter. 2004. *Applied linear regression models*. McGraw-Hill.
- Kutner, Michael H., John Neter, Christopher J. Nachtsheim and William Wasserman. 2004. *Applied linear statistical models*. McGraw-Hill.
- Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.
- Verzani, John. 2005. *Using R for introductory statistics*. Boca Raton, FL: Chapman & Hall/CRC.