Advanced Quantitative Methods

Jos Elkink
College of Humanities Graduate School
University College Dublin

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

Fall 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. 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.

The course website can be found at:
http://jaeweb.cantr.net/teaching

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.

The course will consist of three distinct parts:

Applied linear regression

The classes in the first half 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? This part of the course will provide a very basic introduction to R (proper introduction will be provided in the second half of this course). It will also provide a clear introduction to the methodology of linear regression and model evaluation. It will include some mathematics, but only to provide a background for understanding regression. Mathematical computation or manipulation will not be required.

Using R

The second half of the course will start with a brief introduction to the nitty-gritty of using R, including importing data sets from other software, manipulating data, accessing subsets of the data, etcetera. This section will also include a session on 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 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.

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. 2005). 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 both on in-class participation and on hands-on applied work. The grade division is as follows:

10%	Participation first half
10%	Participation second half
30%	Continuous assessment (1 assignment)
50%	Course paper

For the continuous assessment assignment you will be provided with a data set and specific regression model. You will have to analyse the model, provide a diagnostics analysis and present and interpret the results (max 1500 words). For the course paper you can find your own data and research question and provide both a linear regression and a generalized regression model. This should be presented and interpreted in a proper paper (max 6000 words), including a brief theoretical background to the model. The deadlines are 17 October 2008 and 19 December 2008, respectively.

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

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

Please note: the entire schedule of classes is only preliminary - the actual pace of the course will depend on student skills and interests.

Week 1: Introduction

Recommended reading: Verzani (2005: ch 1); Faraway (2002: appendices A-C); Fox (2002: ch 1-3).

Week 2: Ordinary Least Squares estimation

Recommended reading: Kutner et al. (2005: ch 1,6); Verzani (2005: §3.4, ch 10); Faraway (2002: ch 2, "Estimation"); Fox (2002: ch 4).

Week 3: Hypothesis testing in linear regression

Recommended reading: Kutner et al. (2005: ch 2); Verzani (2005: ch 7,8,10); Faraway (2002: ch 3, "Inference").

Week 4: Binary explanatory variables and interactions

Recommended reading: Kutner et al. (2005: ch 8).

Week 5: Regression diagnostics

Recommended reading: Kutner et al. (2005: ch 3,10); Faraway (2002: ch 7, "Diagnostics"); Fox (2002: ch 6).

Week 6: More on multiple regression

Recommended reading: Kutner et al. (2005: ch 6,7).

Using R

Week 7: Types of data sets, importing data, and manipulating data

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

Week 8: Simulation and bootstrapping

Required reading: Verzani (2005: ch 6).

Generalized linear models

Week 9: Logit and probit - concept

For the logit and probit models, the main approach is taken from **?**: ch 5)King1998 and Long (1997: ch 3) is a more standard source. Both are fairly technical in nature. Kutner et al. (2005: ch 14) provides a somewhat more accessible introduction. See also Fox (2002: §5.2.1), Verzani (2005: §12.1) or Maindonald and Braun (2007: §8.1-8.3) for application in R.

Week 10: Logit and probit - interpretation and presentation

This session is to a very large extent based on King (1998: ch 5). A very useful article on this is also King, Tomz and Wittenberg (2000).

Week 11: Ordinal probit and multinomial logit

A common, but technical source is Long (1997: ch 5-6). A source for R is Fox (2002: §5.2.2) or Maindonald and Braun (2007: §8.6.1).

Week 12: optional

The topic of week 6 depends on students' interest. Options most in line with the limited dependent variables topic are:

Week 12a: Survival analysis

See Maindonald and Braun (2007: §8.7) for an introduction to survival models in R.

Week 12b: Count models

A common source is Long (1997: ch 8) and a very useful one is King (1998: ch 5). For R, have a look at Fox (2002: §5.3.1, §5.4) or Maindonald and Braun (2007: §8.4).

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 multilevel/hierarchical models.* Analytical Methods for Social Research Cambridge: Cambridge University Press.

Gujarati, Damodar N. 2003. *Basic econometrics.* 4th ed. Boston: McGraw-Hill.

King, Gary. 1998. *Unifying political methodology. The likelihood theory of statistical inference.* University of Michigan Press.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the most of statistical analyses: improving interpretation and presentation." *American Journal of Political Science* 44(2): 341–355.

Kutner, Michael H., Christopher J. Nachtsheim and John Neter. 2004. *Applied linear regression models.* McGraw-Hill.

Kutner, Michael H., Christopher J. Nachtsheim, John Neter and William Li. 2005. *Applied linear statistical models.* 5th ed. Boston: McGraw-Hill.

Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage Publications.

Maindonald, John and John Braun. 2007. *Data analysis and graphics using R. An example-based approach.* 2nd ed. Cambridge: Cambridge University Press.

Verzani, John. 2005. *Using R for introductory statistics.* Boca Raton, FL: Chapman & Hall/CRC.