

Intermediate Social Statistics

Course providers: Ray Duch (raymond.duch@nuffield.ox.ac.uk)
and Tom Snijders (Tom.Snijders@nuffield.ox.ac.uk)

Course Outline in PDF format

Lecture Handouts

[Matrices and Linear Regression](#)
[Maximum likelihood and model selection](#)
[Binary logit and probit](#)
[Multinomial logit and Ordered Logit/Probit](#)
[Factor analysis and Principal Component Analysis](#)
[Scale Construction](#)
[Event Count and Duration Models](#)
[Models with Non-random Selection](#)

Take home exam ([Click here](#))

Course structure

There will be a two hour lecture on Wednesday 2-4 pm in Lecture Theatre in the Manor Road Building every week during Hilary term. These lectures are accompanied by four classes where the practical aspects of the models discussed in the lecture will be taken up. Classes will be run by Mark Pickup and David Armstrong on Thursdays of weeks 3, 4, 6, 7, and 8 in the Manor Road Building IT room from 4-6pm. If you are not already familiar with Stata you are strongly encouraged to attend the special introductory session at 4-6 pm on Thursday of week 3. Also, there will be a two day work-shop from 2-5 pm on Wednesday and Thursday of 0th week of Hilary titled *An Introduction to Mathematics for the Social Sciences* led by Indridi Indridason which will introduce some of the fundamental mathematical concepts used in statistical analysis. Attendance is completely voluntary for this workshop, but recommended for those interested in quantitative data analysis.

Topics covered each week

Week	Topic
1	<i>Matrices and Linear regression (Snijders)</i> This week introduces linear regression in matrix notation, and the necessary knowledge of matrix-algebra as background. The aim is to increase understanding of linear regression and to provide a basis for more advanced models. Literature: Freeman, parts of chapters 2-4.
2	<i>Maximum likelihood estimation and model selection (Duch)</i> This lecture introduces the idea of maximum likelihood as a method of estimating parameters in a statistical model and thereby providing further theoretical background to the models introduced in later weeks. MLE also provides a method of model selection. Freeman, chapter 6

	Long, chapter 2
3	<p><i>Binary logit and probit models with applications (Duch)</i></p> <p>This week introduces regression models for binary dependent variables such as yes/no or voted/abstained.</p> <p>Long, J. Scott. 1997 Regression Models for Categorical and Limited Dependent Variables. Chapter 3 and 4.</p>
4	<p><i>Multinomial and ordered logit/ probit. (Duch)</i></p> <p>These models are for ordered categorical variables, e.g. survey questions where there are ordered categories such as strongly agree, agree, neither agree nor disagree, disagree, strongly disagree.</p> <p>Long, J. Scott. 1997 Regression Models for Categorical and Limited Dependent Variables. Chapter 5 and 6</p>
5	<p><i>Factor analysis and principal component analysis (Snijders)</i></p> <p>These are methods for summarizing the association between several interval or ratio level variables. They are often used when it is thought that some of the variables are essentially measuring the same phenomenon, or that there are a few underlying ('latent') phenomena affecting the values of the observed variables.</p> <p>Literature: Bartholomew et al., chapters 5-6.</p>
6	<p><i>Scale Construction (Snijders)</i></p> <p>Methods will be discussed for constructing scales from binary items, and from items consisting of ordered categories. The main method discussed will be Mokken scaling.</p> <p>Literature: to be announced.</p>
7	<p><i>Event count and Duration models (Duch)</i></p> <p>Models for count data are used for relatively infrequent events, with say an average between 0 and 7 (e.g. the number of presidential vetoes in a year).</p> <p>Long, J. Scott. 1997 Regression Models for Categorical and Limited Dependent Variables. Chapter 8; and TBA</p>
8	<p><i>Models with Non-random Selection (Duch)</i></p> <p>Long, J. Scott. 1997 Regression Models for Categorical and Limited Dependent Variables. Chapter 7; and TBA</p>

Assessment

For those who are taking this course for credit, 1) half of your grade will be based on weekly assignments and 2) half your grade will be based on an exam to be set by the relevant lecturer(s) no later than Friday of the eighth week of Hilary Term. Your homework assignments should be handed into your class instructors at the date indicated on each of the weekly assignments. Answers to the exam should be handed in to the Politics Graduate Admin Office by 12 noon on Friday of 6th week of Trinity Term. Students should

not write their names on their answers, but identify their scripts with their candidate number only. Candidates will be expected to demonstrate skill in the application of, and ability to evaluate critically, the models and methods discussed in both the lectures and classes.

Principal Texts:

J Scott Long. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage.

J Scott Long and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press.

David J. Bartholomew, Fiona Steele, Irini Moustaki, and Jane I. Galbraith. *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Chapman & Hall/CRC, 2002.

David A. Freedman. *Statistical Models; Theory and Practice*. Cambridge University Press, 2005.

David J. Bartholomew, Fiona Steele, Irini Moustaki, and Jane I. Galbraith. *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Chapman & Hall/CRC, 2002.

Other Readings

John Aldrich. 1984. *LPM, Logit, and Probit*. Sage.

Vani Kant Borooah. 2001. *Logit and Probit: Ordered and Multinomial Models*. Sage.

Dayton, Mitchell C. 1999. *Latent Class Scaling Analysis*. Sage.

John Fox. 1991. *Regression Diagnostics*. Sage.

John Fox. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Sage.

Thomas H. Wonnacott and Ronald J. Wonnacott. 1990. *Introductory Statistics*. Wiley.