

2013 ECPR Summer School in Methods and Techniques
University of Ljubljana
Course Description Form

Course title

D3 Multiple Regression Analysis II: The Link Function

Instructor details

First name, last name : **Prof. Dr. Bernhard Kittel**

Department/Unit : Institute of Economic Sociology

Institution : University of Vienna

Full postal address for ECPR correspondence: Brünner Straße 72, 1210 Wien

Phone : 0043-1-4277-38311

E-mail : bernhard.kittel@univie.ac.at

Short Bio

Bernhard Kittel is professor of Economic Sociology at the University of Vienna. Previous engagements include the University of Oldenburg, the University of Amsterdam, the University of Bremen, and the Max Planck Institute for the Study of Societies in Cologne. His research interests cover collective decision-making, political economy, and comparative research methodology.

Short outline

The course focuses on the adequate specification of the model. Dummy variables and interaction terms will be discussed and models that are nonlinear in variables but linear in parameters will be studied. Then, non-continuous dependent variables will be introduced. The course contains an introduction to the logic of maximum likelihood estimation, logit models for binary and multinomial dependent variables and models for count data.

Long outline

The course is a direct sequel to “Multiple Regression Analysis: Estimation and Diagnostics” offered in the first week of the summer school. It is recommended to take both courses as two modules of a full two-week course on multiple regression analysis.

The course covers approaches to deal with deviations from the linearity assumption. Firstly, the inclusion of dummy variables among the regressors and multiplicative model specifications are covered, along with the associated procedures to derive correct inferences from such models. Several transformations which tackle nonlinear associations between variables but maintain a linear specification of the model are discussed. Among others, power functions, log functions, and inverse functions will be presented.

After discussion specifications of the “right hand side” (the independent variables) we turn to non-continuous dependent variable. After introducing the logic of maximum likelihood estimation, we will study logit models for analyzing binary dependent variables along with some diagnostic tools. We then extend the logit model to multinomial dependent variables and discuss poisson models for count data.

For most of the examples, individual-level data from the European Social Survey will be used. This dataset is freely available at www.europeansocialsurvey.org. Some topics will be discussed on the basis of a dataset containing aggregate political economy data. All lab sessions will use the software package R (www.r-project.org). The course is set up as follows: First a topic is presented in the lecture, then students are asked to work on a set of tasks in which they stepwise develop and test a regression model, using the ideas and formulae presented in the lecture.

Note on readings: No book is fully satisfactory in the details of presentation or the emphases laid. Therefore, I work with two textbooks which I like for different reasons, Gujarati and Fox. I expect participants to have worked through at least one of the indicated chapters in the literature when we discuss the topic in class.

It is highly recommended to bring a copy of the books as well as a laptop computer with a recent version of R installed.

The course assumes proficiency with descriptive and inferential statistics at the level of test theory and bivariate regression analysis. A brief refresher in matrix algebra will be offered during class but it is highly recommended not to rely on this refresher if you do not bring along a basic understanding of the logic of matrix algebra.

Day-to-day schedule

	Topic(s)	Details
Day 1 (3 hours)	Introduction and the Left-hand Side 1: Dummy variables and interactions	90' Lecture: Introduction; Treatment of dummy variables and interactive model specifications 90' Lab: Specifying and interpreting models with dummy variables and interaction terms
Day 2 (3 hours)	The Left-hand Side 2: Functional Forms	90' Lecture: Functions and transformations 90' Lab: Exploring nonlinearities
Day 3 (3 hours)	The Right-hand Side 1: Modelling Binary Response Variables	90' Lecture: Maximum likelihood estimation, Logit models for binary response variables 90' Lab: Specifying and testing binary logit models
Day 4 (3 hours)	The Right-hand Side 2: Multinomial and count data	90' Lecture: Multinomial models and count data 90' Lab: Specifying and testing multinomial logit models and poisson models
Day 5 (3 hours)	The Right-hand Side 3: Interpreting Generalized Linear Models	90' Lab: Model interpretation 90' Lab: Model interpretation

(Day 11 – Saturday : Exam)

Day-to-day reading list

	Readings
Day 1	Gujarati, Ch. 9 Fox, Ch. 7, 17
Day 2	Gujarati, Ch. 6 Fox, Ch. 4
Day 3	Gujarati, Ch. 15.1-15.10 Fox, Ch. 14.1
Day 4	Gujarati, Ch. 15.12-15.14 Fox, Ch. 14.2, 15.1-15.2
Day 5	Fox, Ch. 15.2-15.4

Requested prior knowledge

- It is mandatory that participants have taken the course on “Multiple Regression Analysis: Estimation and Diagnostics” in the first week of the summer school or have obtained equivalent prior knowledge through other means.
- The course makes use of the freeware statistical package R, which can be downloaded from <http://www.r-project.org>, and assumes basic skills in using the package. For students unfamiliar with R, a preparatory course will be offered prior to the course. A good introduction into using R for statistics is Field et al (2012).
- Students are expected to understand the logic of inferential statistics. Prior knowledge of multivariate statistics is useful but not necessary. Remedial reading: any good statistics textbook, for

example: Alan Agresti & Barbara Finlay (2008). Students familiar with R but in need of a refresher in basic statistics are encouraged to take part in the preparatory course on statistics.

d) Furthermore, although a brief session on the use of matrix algebra in regression analysis is scheduled at the beginning of the course, it is assumed that the logic of matrix algebra is understood at the level presented in Gujarati.

Software used

R, <http://www.r-project.org>

Literature

Agresti, A. and B. Finlay, 2008. Statistical Methods for the Social Sciences, 4th edition, Pearson.

Field, A., J. Miles and Z. Field, 2012. Discovering Statistics Using R, Sage.

Fox, J. and S. Weisberg, 2010. An R Companion to Applied Regression, London: Sage.

Fox, J., 2003. Effect Displays in R for Generalised Linear Models, Journal of Statistical Software, 2003, Vol. 8, No. 15

Fox, J., 2008. Applied Regression Analysis and Generalized Linear Models, London: Sage.

Gujarati, D.N. and D. Porter, 2009. Basic Econometrics, 5th edition, New York: McGraw-Hill.

Lecture room requirement

Ordinary lecture hall with beamer facilities, computer lab. On last day, we will need the lab for both 90 minutes sessions.