**Contents**

This course deals with statistical models for the analysis of quantitative and qualitative data, of the types usually encountered in social science research. The statistical methods studied are the general linear model for quantitative responses (including multiple regression, analysis of variance and analysis of covariance), binomial regression models for binary data (including logistic regression and probit models), models for count data (including Poisson regression and negative binomial models) and models for survival data (focusing on piecewise exponential models fitted via Poisson regression). All of these techniques are covered as special cases of the Generalized Linear Statistical Model, which provides a central unifying statistical framework for the entire course.

**Approach**

The course is taught at an intermediate statistical level. The emphasis is on understanding and applying statistical concepts and techniques, rather than proving theorems. However, the course assumes familiarity with basic concepts in probability theory, statistical estimation and testing theory, and statistical methodology up to multiple regression analysis, at least at the level of a serious introductory course such as WWS507c. Some familiarity with matrix algebra and calculus is necessary. Computer literacy is essential, as we make extensive use of the computer. We recommend using Stata, a general-purpose statistical package available on Windows and other platforms, but students are free to use other software packages such as R or SAS.

**Requirements**

Course requirements consist of required readings, six problem sets, and two partial exams, one near the middle and another at the end of the term. Most of the material of the course is covered in formal lectures. A set of lecture notes is available on the web, and these can be supplemented with optional readings. The problem sets deal mostly with analysis of small datasets. The two partial exams emphasize the application of techniques and the interpretation of results. Final grades are calculated as a weighted average of the grades received during the term, using weights of 40% for the problem sets and 30% for each of the two partial exams.
# List of Lectures

The following is a tentative list of the topics to be covered in each of the lectures scheduled for this term. The overall pace and/or the distribution of lectures within each topic may be altered if an adjustment seems advisable during the course of the term. The date of the final exam will be set by the WWS later in the term.

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, September 16</td>
<td>Introduction and overview of the course. Responses and predictors. Factors and covariates. The generalized linear model. Review of likelihood theory.</td>
</tr>
<tr>
<td>Monday, September 28</td>
<td>Analysis of variance models. One-way anova and regression with dummy variables. Two-way anova. The additive model. Main effects and interactions.</td>
</tr>
<tr>
<td>Wednesday, September 30</td>
<td>Analysis of covariance models. The additive model. The assumption of parallelism. Models with different intercepts and different slopes. Interpretation.</td>
</tr>
<tr>
<td>Monday, October 5</td>
<td>Regression diagnostics. Analysis of residuals. Influential observations, leverage and influence. Q-Q plots.</td>
</tr>
<tr>
<td>Wednesday, October 7</td>
<td>Regression remedies. Transforming the response. The Box-Cox family of transformations. Transforming the predictors.</td>
</tr>
<tr>
<td>Wednesday, October 14</td>
<td>Maximum likelihood estimation and testing in logistic regression models. The comparison of two groups. The odds ratio. Comparison of several groups. The one-factor model. The one-variate model.</td>
</tr>
<tr>
<td>Wednesday, October 21</td>
<td>Alternative links for binary data. Probit analysis. The c-log-log link. Regression diagnostics with binary data.</td>
</tr>
<tr>
<td>Wednesday, November 28</td>
<td>First Partial Exam</td>
</tr>
<tr>
<td>Monday, November 9</td>
<td>Models for rates of events. Exposure and the use of an offset in the linear predictor.</td>
</tr>
<tr>
<td>Wednesday, November 11</td>
<td>Extra-Poisson variation. The negative binomial model. Zero-inflated models for counts.</td>
</tr>
</tbody>
</table>
Monday, November 16

Wednesday, November 18
Sequential logits. Sequential binary choices and continuation ratio models. Equivalence with logit models.

Monday, November 23
Models for ordered categorical data. Ordered logits and probits. Latent variable formulation and interpretation of the coefficients.

Monday, November 30
Survival and event history models. The survival and hazard functions. Censoring mechanisms. The likelihood function for non-informative censoring.

Wednesday, December 2
The proportional hazards model. The baseline hazard. Relative risks. Time-varying covariates. Time-varying effects and models with interactions.

Monday, December 7

Wednesday, December 9
Discrete time models and equivalence with logistic regression. Unobserved heterogeneity. Topics in survival analysis.

Monday, December 14
The analysis of panel data. Random effects and fixed effects. Intra-class correlation.

Wednesday, December 16
Fixed and random effect models for binary and count data. Hierarchical models.

TBA
Second Partial Exam.

January 2016

Supplementary Readings
The material of this course is covered in detail in the lecture notes. The following references are pointers to more detailed supplementary discussions, classified by subject.

**Linear Models**


**Generalized Linear Models**

Other General Books


Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data, 2nd Edition. Cambridge, MA: The MIT Press. A comprehensive treatise that will be particularly useful to economists, covering the models for cross-sectional data discussed in the course as well as extensions for longitudinal data.


More Specialized Texts


