Course ID 009307

Generalised Linear Models

MATH 38052

Credit rating 10

ECTS credits 5

Unit coordinator: Jingsong Yuan

Semester 2

School of Mathematics

Undergraduate

Level 3

FHEQ level ' Last part of a Bachelors'

Marketing course unit overview

As an important modelling strategy Linear Models is concerned with investigating whether, and how, one or more so-called explanatory variables, such as age, sex, blood pressure, etc., influence a response variable, such as a patient's diagnosis, by taking random variations of data into account. In Linear Models, linear regression technique and Normal distribution are used to explore the possible linear relation between a continuous response and one or more explanatory variables. In this course unit we depart from linearity and normality, the very strict limitation in Linear Models. We study the extension of linearity to non-linearity and normality to a commonly encountered distribution family, called the exponential family of distributions. This extension forms Generalized Linear Models (GLM). The GLM, on the one hand, unifies linear and non-linear models in terms of statistical modelling. On the other hand, it can be used to analyze discrete data, including binary, binomial, counted and categorical data that arise very often in biomedical and industrial applications.

Course unit overview

As an important modelling strategy Linear Models is concerned with investigating whether, and how, one or more so-called explanatory variables, such as age, sex, blood pressure, etc., influence a response variable, such as a patient's diagnosis, by taking random variations of data into account. In Linear Models, linear regression technique and Normal distribution are used to explore the possible linear relation between a continuous response and one or more explanatory variables. In this course unit we depart from linearity and normality, the very strict limitation in Linear Models. We study the extension of linearity to non-linearity and normality to a commonly encountered distribution family, called the exponential family of distributions. This extension forms Generalized Linear Models (GLM). The GLM, on the one hand, unifies linear and non-linear models in terms of statistical modelling. On the other hand, it can be used to analyze discrete data, including binary, binomial, counted and categorical data that arise very often in biomedical and industrial applications.

Aims

To study an important aspect of modern statistical modelling in an integrated way, and to develop the properties and uses of GLM, focusing on those situations in which the response variable is discrete. To explore some of the wide range of real-life situations occurring in the fields of agriculture, biology, engineering, industrial experimentation, medicine and social

science that can be investigated using GLM.

Learning outcomes

On successful completion of this course unit students will have a good understanding of

- the principles and methods of statistical modelling for GLM: response and explanatory variables, maximum likelihood estimation, confidence interval and hypothesis testing, goodness of fit, etc.;
- the use of the computer statistical software R, which is available on the Mathematics PC Cluster and does not require any previous programming experience;
- the statistical analysis of both continuous and discrete data arising in practice through using the statistical software R.
- Future topics requiring this course unit
- This course unit is part of the 4th year/MSc Generalised Linear Models and Survival Analysis. It is naturally related to another 4th year unit, Longitudinal Data Analysis.

Syllabus

1.Introduction: background, review of linear models in matrix notation, model assessment, some pre-required knowledge. [2]

2.The exponential family of distributions: Definition and examples. Mean and variance, variance function and scale parameter. [2]

3.Generalized linear models (GLM): linear predictor, link function, canonical link, maximum likelihood estimation, iterative reweighted least squares and Fisher scoring algorithms, significance of parameter estimates, deviance, Pearson and deviance residuals, Pearsonâ€'s chi-square test and the likelihood ratio test, model fitting using R. [7]

4.Normal linear regression models: least squares, analysis of variance, orthogonality of parameters, factors, interactions between factors. [2]

5.Binary and Binomial data analysis: distribution and models, logistic regression models, odds ratio, one- and two-way logistic regression analysis. [5]

6.Poisson count data analysis: Poisson regression models with offset, two-dimensional contingency tables, log-linear models. [4]

Assessment methods

Other	20%
Written exam	80%

Coursework: 20%End of semester examination: two hours weighting 80%

Feedback methods

Tutorials will provide an opportunity for students' work to be discussed and provide feedback on their understanding.

Requisites

MATH20701	Probability 2	Pre-Requisite	Compulsory
MATH20802	Statistical Methods	Pre-Requisite	Compulsory

Students are not permitted to take more than one of MATH38052 or MATH48052 for credit in the same or different undergraduate year. Students are not permitted to take MATH48052 and MATH68052 for credit in an undergraduate programme and then a postgraduate programme.

Available as free choice? N

Recommended reading

- Dobson, A. J., An Introduction to Generalized Linear Models, Chapman & Hall 2002.
- Krzanowski, W., An Introduction to Statistical Modelling, Edward Arnold 1998.
- McCullagh, P. and Nelder, J. A., Generalized Linear Models, Chapman & Hall 1990.

Scheduled activity hours

Lectures	22
Tutorials	11

Independent study hours 67 hours

Version Nbr 009.0.0