



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Modelli Lineari per Dati Categoriali

1819-1-F8203B010-F8203B011M

Learning objectives

The course introduces the linear models for categorical data according to two different settings. The first concerns the general linear model (GLM), including several special cases such as ANOVA and ANCOVA models. The second setting deals with the generalized linear models, in particular Poisson log-linear models for count data and binomial logistic models, in a GLM perspective. Analyses of empirical cases are carried out through the SAS software.

Contents

General Linear Model (GLM), one-way or more than one-way ANOVA and ANCOVA models. Generalized Linear Models (GzLM), binomial logistic model and Poisson log-linear model. Applications to real and experimental data with the SAS software.

Detailed program

- Theory of general linear model (GLM): model specification, assumptions, generalized inverse, estimable functions, testable hypotheses. Link with the constrained least-squares estimation method. Sum-to-zero and set-to-zero linear constraint approaches. Effect parameterization vs. reference category parameterization. Contrasts
- Special cases of GLM: one-way or more than one-way fixed-effects ANOVA models, ANCOVA model. SAS PROC GLM

- GLM selection: forward and stepwise methods. SAS PROC GLMSELECT
- Generalized Linear Model (GzLM): probability distribution function of response variables, link function, model specification, maximum likelihood estimation method, estimator properties, criteria for goodness-of-fit, confidence limits and statistical testing hypotheses
- Special cases of GzLM: Poisson log-linear model for count data and binomial logistic model, in a GLM perspective. SAS PROC GENMOD

Prerequisites

Knowledge of the topics covered in undergraduate courses of Multivariate Statistical Analysis is recommended

Teaching methods

Theoretical lectures in the classroom and practical exercises in the statistical-informatics laboratory with SAS software

Assessment methods

The exam consists in the preparation of a statistical data analysis with the SAS software (according to the modalities specified on the e-learning platform of the course), whose output has to be discussed during the examination, and in a written test (duration: 2 hours) concerning both theoretical and practical topics. The theoretical questions are general. The parts with a more methodological nature are the object of an optional issue.

Concerning the purely applicative part, it is required to prepare a data analysis with the SAS software using one of the methodologies covered during the course. The comment of the analysis thus produced is the object of discussion during the examination.

The methodology for preparing the analysis is assigned nominally and randomly (using a random number generator) to each student enrolled in the e-learning platform of the course.

The statistical analysis has to be prepared before the exam, following a specific track relative to the methodology assigned and published at the end of the course on the e-learning platform. During the examination, the printing of the output has then to be presented according to the modalities specified on the e-learning platform of the course.

Given the abundance of teaching material uploaded on the e-learning platform of the course, no distinction is made between exams for attending students and exams for non-attending students.

Textbooks and Reading Materials

- Teaching material uploaded on the e-learning website of the course
- Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, New York: John Wiley & Sons
- Dobson, A. (1990), *An Introduction to Generalized Linear Models*, London: Chapman & Hall
- Littell, R. C., Freund, R. J., and Spector, P. C. (2002), *SAS for Linear Models*, 4th Edition, Cary, NC: SAS Institute Inc.
- Searle, S. R. (1971), *Linear Models*, New York: John Wiley & Sons

Semester

First semester, second period

Teaching language

Italian
