MULTIVARIATE ANALYSIS (21893)

Descriptive information on the subject
Name and code of the subject: Multivariate Analysis , Code: 21893
Academic Year: 2013-14, 3rd/4th year, 1st term
Degree / studies: ECO/BMA
Language of instruction: Catalan
Teaching staff: Albert Satorra, Ferran Carrascosa

Introduction to the subject

The course deals with the techniques of multivariate statistical analysis commonly used in empirical work in economics, marketing, organization of companies, market research, and other economic and social fields. It includes principal component analysis, factorial component analysis, multivariate scales (MDS), grouping (cluster) and discriminate analysis. The course focuses on the concepts and practice of the techniques studied. Matrix calculation is used in statistical formulation computing. Part of the continuous assessment is an assignment including the application of the techniques studied to empirical data. The course uses R free software programming language intensively. Proprietary software, such as as SPSS and Stata is also used.

Competences to be obtained in the subject

The ability to analyze synthesize information from multivariate data; the ability to design a multivariate analysis; the capacity for critical assessment of a report that includes a multivariate analysis; the ability to synthesize information in the form of indices and multivariate principal components, the ability to classify units of multivariate observations; the ability to manage databases and professional statistical software; the ability to produce a report on a multivariate analysis using non-technical language.

Contents.

Topic 1. Introduction to multivariate analysis.
Terminology, organization of data, applications of multivariate analysis, distances, statistical software (R, SPSS) databases, examples of multivariate analysis.

Topic 2. Matrix Calculation
Basics of matrix algebra, vectors and eigenvalues, semi-definite positive matrix, distance, OLS approximation of a matrix (the Eckart-Young Theorem), r matrix calculation in AML.

Topic 3. Multivariate Distributions.
Random vectors, multivariate distributions, Normal multivariate linear combination of variables, conditional distributions, simple and multiple regression.

Topic 4. Principal component analysis (PCA)
Definition of a principal component, a procedure for calculating principal components, description of a PCA, plotting points, plotting variables, biplot, examples.

Topic 5. Methods of multidimensional scaling. Constructing a map from a distance matrix, metric MDS, MDS calculation procedure, non-metric MDS generalizations, stress and configurations, examples.

Topic 6. Correspondence analysis (CA)
Association in a contingency table, reducing the dimensions, configuration of points, configuration of rows, inertia, descriptive measures in CA, type of standardization, examples.

Topic 7. Cluster analysis. The problem of clustering, type of cluster analysis, hierarchical methods, distance measurements, K-means method, other classification techniques, examples.

Item 8. Factorial analysis (FA)
The FA model, estimation methods, description of the analysis, choice of the number of factors, factor rotation, examples.

Item 9. MANOVA and discriminant analysis
Testing the quality of mean vectors, dimension reduction, classification, linear and quadratic discriminant analysis, canonical discriminant functions, evaluation of a discriminant analysis, non-parametric methods, examples.

Assessment. The final grade is the result of:
(1) 15%: Participation in class and seminars
(2) 35%: Assignment
(3) 50% : Final Exam
Sections (1) and (2) are subject to continuous assessment.

On the assignment. The Assignment is a necessary part of the assessment, to be submitted before/or on the day of the final exam. This assignment should be done in groups of a maximum of three students (unless approved by the lecturer). It consists of a study of the statistical techniques covered on the course, involving the use of a context of empirical data . In this assignment, students demonstrate their ability to execute, analyze and produce a written presentation of statistical results to an audience inexperienced in statistics.

Resit exam. Students who have participated in more than half of the continuous assessment activities and who have taken the final exam of the course but failed the ordinary assessment will be able to take the resit examination. The resit exam consists of a comprehensive revalidation test of the entire syllabus of the course. Continuous assessment is not recoverable. The final grade for the course for students taking the resit is the weighted average of the grades for the continuous assessment (30%) and the resit exam (70%). In order to calculate the average, the resit grade must be at least 4, on a scale of 0 to 10.

Bibliography and teaching resources.

_ Bartholomew, D.J., F. Steele, I. Moustaki & J. I. Galbraith (2002) The analysis and interpretation of multivariate data for social scientists, Chapman and Hall
_ Jobson, J.D. Applied Multivariate Data Analysis, Berlin: Springer Verlag, 1992
_ Johnson, R.A. & Wichern, D.W. , Applied Multivariate Statistical Analysis (4rth Edition) Englewood Cliffs: Prentice Hall, 1992
_ Lattin, J. J.D. Carrol, P.E. Green, Analyzing Multivariate Data, Duxbury
_ Levy, J.P. (ed.) Analisis multivariable para las ciencias sociales, Prentice Hall, 2003
_ Manly, B. F. J. Multivariate Statistical Methods, London: Chapman and Hall, 1991
_ Mardia, K.V., Kent, J.T. , Bibby, J.M. Multivariate Analysis, Academic Press , 1988
_ Morrison, D. F. , Multivariate Statistical Methods (3rd ed.), McGraw Hill 1990
_ Pea, D. , Analisis de Datos Multivariantes, McGraw Hill, 2002
_ Norusis, M.J. SPSS Advanced statistics student guide. Chicago: SPSS Inc, 1992.

Teaching resources.
Available on the course website:
http://www.econ.upf.edu/ satorra/AnalisiMultivariant/AM2012.htm

Methodology.

A combination of lectures and seminars in small groups. There are 10 sessions of lectures of 1 hr 30 min over 10 weeks, in which concepts of the course are discussed. The seminars take place in the computer room, and use the SPSS and R programs. Social sciences and economics databases will be used during the sessions. In the seminars, we will first work on the problem that has previously has been posted on the course website, and then will resolve any questions that arise about the implementation of the statistical technique in the seminar. Seminar 4 will be entirely devoted to topics of discussion and preparation of the assignment in progress.