



Universitat de Lleida

DEGREE CURRICULUM
**APPLICATIONS OF
MULTIVARIATE ANALYSIS TO
BIOTECHNOLOGY**

Coordination: GATIUS CORTIELLA, FERNANDO

Academic year 2023-24

Subject's general information

Subject name	APPLICATIONS OF MULTIVARIATE ANALYSIS TO BIOTECHNOLOGY			
Code	101636			
Semester	1st Q(SEMESTER) CONTINUED EVALUATION / 2nd Q(SEMESTER) CONTINUED EVALUATION			
Typology	Degree	Course	Character	Modality
	Bachelor's Degree in Biotechnology	4	OPTIONAL	Attendance-based
Course number of credits (ECTS)	6			
Type of activity, credits, and groups	Activity type	PRAULA		TEORIA
	Number of credits	3		3
	Number of groups	2		2
Coordination	GATIUS CORTIELLA, FERNANDO			
Department	ENVIRONMENT AND SOIL SCIENCES AND CHEMISTRY			
Teaching load distribution between lectures and independent student work	60 hours in class + 90 hours at home / group			
Important information on data processing	Consult this link for more information.			
Language	Catalan			
Distribution of credits	Two different blocks of electives: - GROUP 1 (maximum 12 students): 6 credits in BLOCK 1 (classes begin in SEPTEMBER) - GROUP 2 (maximum 12 students): 6 credits in BLOCK 3 (classes begin in FEBRUARY) IMPORTANT!!! At the time of registration the student must select one of the two groups because each group is limited to 12 STUDENTS MAXIMUM.			

Teaching staff	E-mail addresses	Credits taught by teacher	Office and hour of attention
GATIUS CORTIELLA, FERNANDO	fernando.gatius@udl.cat	12	

Subject's extra information

IMPORTANT: if any activity cannot be carried out normally, it will be replaced by an alternative activity or followed virtually.

Learning objectives

The student who passes the subject must: (Knowledge)

- UNDERSTAND and USE the basic terminology and the fundamental concepts addressed in the subject specified in the program.
- Know how to use the concepts to INTERPRET results obtained from the analysis of data of different nature, both in the univariate and multivariate fields.
- Obtain a scientific training based on the knowledge of the methodology involved in the data processing, as well as the universality that confers the APPLIED character of the subject to data.
- Demonstrate a preparation for the design of research based on the principles of relevance, control and generalization.

The student who passes the subject must be able to: (Capacity)

- RAISE and DEVELOP the study and treatment of different types of data according to their character.
- Express correctly the result of a study and make comparisons of numerical results with the corresponding tools and contrasts.
- ANALYZE / INTERPRET the models involved in the descriptive study of a multivariate data set.
- UNDERSTAND and APPLY the multivariate projection techniques for other studies apart from the descriptive ones, such as those involved in the calculation of relationships between different variables and the procedures for classifying samples.
- INTERPRETING, DISCUSSING and EXTRACTING CONCLUSIONS from multivariate data studies (in general).

Competences

The graduate in Biotechnology must:

CG2 Interpret scientific-technical information with a critical sense and be able to make presentations based on this information.

CG6 Know to use the software and the specific databases in the different areas of Biotechnology.

CG7 Use the scientific method to analyze data and design experimental strategies with biotechnological applications.

CG11 Acquire criteria for choosing the most appropriate analytical techniques for each specific practical case.

Subject contents

In most of the studies, we work with samples / data defined by many variables. Data treatment is essential to extract the conclusions of the studies. In this subject a series of tools are presented to be able to deal with all this information in an efficient way WITHOUT the need of advanced knowledge of mathematics or statistics, a global

vision of the analysis and treatment of data and its APPLICATION for the study and extraction of conclusions. To do this, we begin (introduction) by situating the student with the initial (very brief) review of the main concepts of the univariate data treatment (which the student already knows) that will allow him to arrive at the correct expression of the result of a study and to do basic treatments of the numerical values that define a variable. In this introduction we review basic statistical concepts that are connected with the presentation and application of projection techniques as tools to treat multivariate data. These tools are applied to the descriptive studies of data, to the relationship between variables (scope of the regression) and in the classification of objects according to pre-established groups.

It is an APPLIED data analysis subject, in which the corresponding software will be used to calculate the models involved in a study, and in which the understanding of the concepts involved in the same will be evaluated by means of the INTERPRETATION of the phenomena shown by the diagrams generated in these models. Practical classes are fundamental in which both the use of said software and the interpretation of the models / diagrams involved in the different studies.

THEORETICAL SUBJECT

INTRODUCTION: UNIVARIATE DATA PROCESSING

Errors in the quantitative analysis of results. Repeatability / reproducibility. Statistics with repeated measures. Normal Distribution. Confidence limits. Contrasts of significance: comparison of the precision and the result of 2 methods. Comparison of more than 2 values (studies with factors - ANOVA). Other studies: comparison of frequencies, detection of outliers, normality tests, ... Conclusions based on the contrasts of significance. Associated errors.

DEFINITION OF THE MULTIVARIATE SCOPE

Indirect observations and correlations. Multivariate Analysis and Multivariate Statistics. Presentation of multivariate data: the data matrix. Main objectives of multivariate analysis techniques.

METHODS OF PROJECTION. PRINCIPAL COMPONENT ANALYSIS (PCA)

Study of data defined by correlated variables. Approach and problems. The first Main Component (PC): definition. Extension to more PC's. Concepts of scores and loadings (projections / representation of samples and variables). Explained variance and residual variance in a PCA model. Residual and leverage of the samples. Study of the outliers. Interpretation of the PCA model. Associated diagrams.

RELATIONSHIP BETWEEN VARIABLES. THE MULTIVARIATE REGRESSION

Study of the regression in univariate environment. The least squares adjustment. Extension of the study to the multivariate field. Multiple linear regression (MLR). Definition of the MLR model. Limitations and problems of the MLR models. First solution to the problems: variable selection. Definitive solution to the problems: Principal Components Regression (PCR). Study of the model: scores, loadings, variances, errors. Limitations / problems of the PCR models. Solution to the problems of PCR: the Regression by Partial Least Squares (PLS). Definition of PLS factors. Type of PLS model. Study of the models: representations (scores, loadings) and associated errors.

THE VALIDATION OF MULTIVARIATE MODELS

Calibration and validation of models: requirements of the associated sets. The concept of "test set" in the validation of models. Other methods: correction by leverage and cross validation. Study of errors. Measurement of experimental error (studies with repetitions). RMSEP, RMSEC and other parameters.

METHODS OF MULTIVARIATE DATA CLASSIFICATION

Supervised and unsupervised procedures. Application of projection methods in the classification: the SIMCA method. Definition of the models that define the classes. Definition of the classification methodology. Analysis of the results (tables and diagrams) of the classification procedure.

PRACTICAL ACTIVITIES

1. Review exercises of univariate studies with spreadsheet: contrasts, analysis of variance and linear regression.
2. Data exploration. Projection methods. Principal Components Analysis (PCA). Representation of the information (scores and loadings). Residuals and leverage.
3. Regression by Principal Components (PCR). Calibration parameters. Distributions of samples and variables in the model obtained. Error of the model. Interpretation. Definition of the model: obtaining the regression coefficients.
4. Partial Least Squares Regression (PLS). Detection of outliers and treatments. Interpretation of the model found:

distributions of samples and variables. Error of the model and regression coefficients. PCR-PLS comparison. Comparison of different methods of Validation of models.

5. Classification methods. Supervised and unsupervised classification. SIMCA method. Analysis of the results of the classification related to objects, variables and models of the classes.

Evaluation

During the course, practical exercises applying the techniques explained in the theoretical sessions are solved and discussed in the classroom.

The subject is divided in three blocks: after each one there will be a practical exercise like those solved and discusses previously in the classroom.

Each of the tests will count for one third of the final qualification of the subject.

If the student does not agree with the qualification obtained in these tests, he will have the right to an extraordinary evaluation at the end of the course.

Although attendance at classes is highly recommended, those students **who can demonstrate the impossibility of attending class during the course due to work**, may request at the beginning of the course (to the center's studies director) an alternative evaluation of the entire subject that will take place on the same day of the evaluation of the third block.

Bibliography

BASIC

Multivariate data analysis in practice. ESBENSEN, K.; SCHÖNKOPF, S.; MIDTGAARD, T. CAMO AS. Trondheim. Norway. 1996.

Chemometrics. Data Analysis for the Laboratory and Chemical Plant. BRERETON, R.G. John Wiley & Sons, Chichester. 2003.

Multivariate Calibration. MARTENS, H.; NAES, T. John Wiley & Sons. Chichester. 1989.

COMPLEMENTARY

Statistics for analytical chemistry (3rd Ed). MILLER, J.C.; MILLER, J.N. Ellis Horwood Corp. Chichester. West Sussex. 1993.

A User-Friendly Guide to Multivariate Calibration and Classification. NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. NIR Publications. Chichester. UK. 2004.

Handbook of Chemometrics and Qualimetrics. MASSART, D.L.; WANDEGINSTE, B.G.M.; BUYDENS, L.M.C.; DeJONG, S.; LEWI, P.J.; SMEYERS-VERBEKE, J. Elsevier Science B.V. Amsterdam. The Netherlands. 1997 (A), 1998 (B).

Statistical Methods in Analytical Chemistry (2nd Ed.) MEIER, P.C.; ZÜND, R.E. John Wiley & Sons, Inc. Chichester. UK. 2000.