Prior skills

The course assumes basic levels of statistics. Students should be familiar with techniques of multivariate statistics such as principal component analysis and clustering. Concepts relative to hypothesis testing and statistical significance, as well as good knowledge of analysis of variance will be appreciated. The main concepts necessary to follow the course can be found, for example, in the text "Exploratory Multivariate Analysis by Example Using R" described on FactoMiner Package website (http://factominer.free.fr/). The course assumes a good knowledge of the R programming language.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
5. CE-2. Ability to master the proper terminology in a field that is necessary to apply statistical or operations research models and methods to solve real problems.
6. CE-3. Ability to formulate, analyze and validate models applicable to practical problems. Ability to select the method and/or statistical or operations research technique more appropriate to apply this model to the situation or problem.
7. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.

Transversal:
1. SUSTAINABILITY AND SOCIAL COMMITMENT: Being aware of and understanding the complexity of the economic and social phenomena typical of a welfare society, and being able to relate social welfare to globalisation and sustainability and to use technique, technology, economics and sustainability in a balanced and compatible manner.
2. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
3. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.
4. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.
TEACHING METHODOLOGY

Learning is based on real experiments, using professional statistical tools. Combining theoretical discussion sessions with practical sessions is favored. Writing of executive reports of the practices is one of the skills that is developed.

LEARNING OBJECTIVES OF THE SUBJECT

- Understand some of the problems posed in marketing field: get to know the users, their preferences and better understand what leads them to buy.
- Understand the role of data management and data mining techniques in the decision-making process. Acquire new knowledge about statistical methods of application in marketing, but which are also applicable in a wide range of fields.
- Acquire knowledge about specific forms of data collection.
- Appreciate the contributions of statistical techniques and, at the same time, develop a critical spirit towards the results obtained.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours small group</td>
<td>15,0</td>
<td>12.00</td>
</tr>
<tr>
<td>Self study</td>
<td>80,0</td>
<td>64.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>30,0</td>
<td>24.00</td>
</tr>
</tbody>
</table>

Total learning time: 125 h

CONTENTS

**Topic 1: Structural analysis of data**

**Description:**
Analyzing large data sets (for example, surveys) requires a methodology that allows capturing the multidimensionality of this type of data, as well as allowing a synthesis easily understood by the user. Which leads to privilege a strategy that combines factorial methods and classification.
These large data sets can be structured into multiple tables for which descriptive factor analysis methods present multiple generalizations adapted to different possible combinations of complex data. For example multiple factorial techniques, mixed, dual, ...

**Full-or-part-time:** 42h
Theory classes: 15h
Self study : 27h

**Topic 2: Open questions and comments**

**Description:**
Open questions and comments are increasingly present in large data sets. They are analyzed using multidimensional methods such as correspondence analysis, multiple factor analysis, and classification methods. Correspondence analysis methods allow models to be introduced in the analysis of open responses.

**Full-or-part-time:** 12h 30m
Theory classes: 4h 30m
Self study : 8h
Topic 3: Sensory evaluation of products. Experience planning, data analysis and holistic methods

Description:
The sensory evaluation of the products is a strategic element of the development of the companies of very diverse sectors, although the preferred sector is the agri-food sector. Its objective is to characterize the products both from the sensory point of view (sight, touch, taste, smell, hearing) and from the point of view of consumer preferences. Sensory evaluations require voluminous data collections and lead to the construction of multiple tables. Statistics is the privileged tool for the conception and analysis of this type of data. Holistic methods allow the comparison of a series of products from a global point of view.

Full-or-part-time: 8h
Theory classes: 3h
Self study: 5h

Topic 4: Unsupervised clustering

Description:
Unsupervised clustering refers to the techniques that make it possible to group a set of individuals or observations according to their characteristics. Specifically, two unsupervised clustering techniques will be studied: hierarchical clustering and K-means. In addition, ways to combine both techniques and various variants will be seen. These techniques allow, for example, to conform clusters of clients or consumers of a company based on their properties and depending on the results, to establish market shares (in the case of clients) or make decisions to improve the performance of a company.

Full-or-part-time: 12h 30m
Theory classes: 4h 30m
Self study: 8h

Topic 5: Supervised clustering

Description:
Supervised clustering or discriminant analysis is applied to the set of methodologies that pursue the classification of individuals or observations. Specifically, 5 supervised clustering techniques based on Machine Learning algorithms will be studied: K-Nearest Neighbors, Naive Bayes, Conditional Trees, Random Forest and Support Vector Machine. These techniques have an eminently predictive aim and their use lies in anticipating, for example, the behavior of customers regarding the purchase of a product.

Full-or-part-time: 29h 30m
Theory classes: 10h 30m
Self study: 19h

Topic 6: Design of new products. Conjoint analysis (Conjoint analysis)

Description:
Conjoint analysis is a very powerful tool to study the valuation that customers make of the various characteristics of a product, when it does not make sense to value each characteristic separately. Conjoint analysis applies knowledge of experimental and regression designs. This methodology allows predicting the reception that a new product may have on the market, by comparison with the products already present.

Full-or-part-time: 20h 30m
Theory classes: 7h 30m
Self study: 13h
GRADING SYSTEM

The evaluation will be made from the performance of practices, and the final mark will be calculated from the mark of the corresponding reports and the mark of a final presentation of the work with a percentage of 50% for each one.

BIBLIOGRAPHY

Basic:

Complementary: