UAB
Universitat Autònoma
de Barcelona

Study Abroad
Programmes
Universitat Autònoma
de Barcelona

PRE-ESTABLISHED PROGRAMME

DATA ANALYSIS

Course contact hours: 45
Recommended credits: 6 ECTS – 3 US
Language: English

Prerequisites

Basic knowledge of statistics and matrix algebra. Background in linear regression and R software is useful but not required.

Objectives and Contextualisation

The main objective of this course is to learn about the tools to analyse and visualize data to communicate results and make informed recommendations. During the course, we will learn modern statistical methods for modelling data and prediction based on them. We will learn to analyse quantitative relationships among variables. The analysis and the visualization of data will be implemented through R software. The course will cover essential tools for the analysis of data collected across all areas of science and industry, and will concentrate on the application of statistical methods with lab illustrations.

Competences

- Data analysis and visualization.
- Making sense of complex datasets.
- Linear regression modelling.
- Supervised or unsupervised statistical learning.
- Application of R software.
- Analytical skills and critical thinking.
- Communicating analytical results.

Learning Outcomes

1. Analyse complex datasets.
2. Present data through various graphs.
3. Define variables and analyse relationships among them.
4. Predict the future values of variables.
5. Implement classification and clustering analyses.
6. Use resampling methods.
7. Carry out dimension reduction and principal component analyses.
8. Implement analyses based on random forests and deep learning methods.
9. Use R software for various data analyses.

Content

1. Introduction.
   Data types and variable categories.
   Overview of statistical learning.
2. Statistical learning.
   Assessing model accuracy.
   Introduction to R software.
3. Linear Regression.
   Simple linear regression.
   Multiple linear regression.
4. Classification.
   Logistic regression.
   Linear discriminant analysis.
5. Resampling methods.
   Cross-Validation.
   The Bootstrap.
6. Linear model selection and regularization.
   Subset selection.
   Shrinkage and dimension reduction methods.
7. Tree-based methods.
   Basics of decision trees.
   Bagging, boosting, and random forests.
8. Deep learning.
   Neural networks.
   Document classification.
9. Unsupervised learning.
   Principal components analysis.
   Clustering methods.

Methodology

The course will combine both theoretical and lab sessions. Individual and group work is expected both in class and at home. It is expected from students to attend all classes, participate in class discussions, and complete all class assignments. It is also expected from students to read the materials provided and use the textbooks to reinforce what will be taught in class.

Activities

Directed:
Class sessions (theory)   20 hours
Class sessions (lab)   18 hours
Assessment   3 hours

Supervised:
Project   4 hours

Assessment

Class attendance is mandatory and subject to the corresponding policy of the Study Abroad Programmes. Active class participation is encouraged and evaluated (20%). Students are expected to work on an empirical project in a group of up to three students and present it in class (20%). Students should also take the midterm (30%) and the final (30%) exams.

Assessment Activities

Title	Weighting	Hours	Content
Participation	20%	38	1-9
Project	20%	4	3-6
Midterm exam	30%	1.5	1-5
Final exam	30%	1.5	6-9

Bibliography

Required

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2023). An Introduction to Statistical Learning with Applications in R. New York: Springer, 2nd Edition (link)

Recommended

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2nd Edition (link)

Lind, D., Marchal, W., and Wathen, S. (2023). Statistical Techniques in Business and Economics. 19th Edition, McGraw Hill (link)

Newbold, P., Carlson, W., and Thorne, B. (2020). Statistics for Business and Economics. 9th Global Edition, Pearson (link)

Software

R (link)

RStudio (link)