INSTITUTO POLITÉCNICO DE BRAGANÇA

Course Unit	Data analysis	Field of study	Information Systems
Bachelor in	Management Informatics	School	School of Technology and Management
Academic Year	2021/2022	Year of study	3
Level	1-3	ECTS credits	6.0
Type	Semestral	Semester	2
Code	9186-709-3201-00-21		

Workload (hours)	162	Contact hours	T -	TP 60	PL -	TC -	S -	E -	OT -	O -

T - Lectures; TP - Lectures and problem-solving; PL - Problem-solving, project or laboratory; TC - Fieldwork; S - Seminar; E - Placement; OT - Tutorial; O - Other

Name(s) of lecturer(s): Maria Prudência Gonçalves Martins

Learning outcomes and competences

At the end of the course unit the learner is expected to be able to:
1. Organize and prepare less structured data for computer analysis, correcting errors and imputing missing data. Synthesize key characteristics and choose the most correct visualization for a data set.
2. Apply inferential techniques for analysis and modeling techniques, such as linear regression and clustering. Construct models involving data of time series.
3. Test and validate the models and results of the performed analysis. Master a high-level tool (R environment) and using it in complex data analysis and modeling problems.

Prerequisites

Before the course unit the learner is expected to be able to:
1. Apply basic concepts of quantitative methods and statistics;
2. Use knowledge of informatics and operate computer programs (software).

Course contents

Statistical inference and non parametric tests. Forecasting methods. Simple and multiple regression. Estimation of models with discrete choices.

Course contents (extended version)

1. Statistical inference and non parametric tests
   - Concepts recap: confidence intervals, hypothesis tests, p-values, types I and II statistical errors
   - Inference on quantitative data: t tests, 1-factor analysis of variance, correlation and independence
   - Inference on qualitative data: Chi-squared test for 1 sample and for contingency tables
2. Forecasting methods
   - Terminology and basic tools for analyzing time series and forecasting methods
   - Time series decomposition
   - Exponential smoothing techniques and ARIMA models
   - Introduction to state space models for computing forecasting intervals
   - Dynamic regression
3. Simple and multiple regression
   - Models' classical hypothesis
   - Ordinary Least Squares (OLS) estimators and properties
   - Extensions: Functional forms, dummy and lag variables
   - Adjustment precision indicators
   - Classical hypothesis violations: multicollinearity heteroscedasticity, autocorrelated disturbances
4. Estimation of models with discrete choices
   - Logit model: estimation and inference
   - Probit model: estimation and inference
5. Basic clustering techniques distances
   - Distances
   - Hierarchical models
   - k-means algorithm
6. Preparation of data for analysis and modeling in R environment
   - Structuring, correction of errors and imputation of missing values
7. Techniques for summarization and visualization of data

Recommended reading

1. Guimarães, R. C. & Sarsfield C. (2010). Estatística. Verlag Dashofer.
2. Hyndman, R. J. & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (http: //otexts. org/fpp/)
3. Zumel N. & Mount J. (2014). Practical Data Science with R. Manning Publications Co.
4. Gareth J. , Witten D. , Trevor H. & Robert T. (2014). An Introduction to Statistical Learning. Springer
5. Gujarati D. , Provost F. & Fawcett T. (2013). Data Science for Business. O'Reilly Media.

Teaching and learning methods

In class there will be a presentation and description of contents and analysis and resolution of small application examples accompanied by practical exercises conducted using statisticalsoftware. During the contact period the students must review the materials taught, solve application exercises and elaborate practical reports that include empirical applications of real problems.

Assessment methods

1. Alternative 1 - (Regular, Student Worker) (Final, Supplementary)
   - Practical Work - 50% (2 pratical works (30% e 20% each))
   - Final Written Exam - 50%
2. Alternative 1 - (Student Worker) (Final, Supplementary, Special)
   - Final Written Exam - 100%
3. Alternative 3 - (Regular) (Special)
   - Final Written Exam - 100%

Language of instruction

1. Portuguese
2. Portuguese, with additional English support for foreign students.

Electronic validation

Maria Prudência Gonçalves Martins	António Jorge da Silva Trindade Duarte	José Carlos Rufino Amaro	Paulo Alexandre Vara Alves
04-03-2022	10-03-2022	12-03-2022	18-03-2022