UNIVERSIDAD DE ALMERÍA
Vicerrectorado de Internacionalización

Internationalization Programs
Study Abroad 2020

Course title: **DATA ANALYSIS IN PRACTICE WITH R**

Teaching period: July 6 to July 17, 2020

Teaching hours: 45

Academic coordinator: Sergio Martínez Puertas/Manuel Sánchez Pérez

Knowledge area: Statistics and Operative Research / Marketing and Market Research

1. INTRODUCTION

Multivariate data analysis techniques are essential in academic and research activity. R is as one of the main open-source programming language in statistics and data science. Its power, versatility and continuous update attract more users every year. The number of R users grows by about 40%, and it is widely used in academia and science. This course combines the explanation of main statistical tools with the use of R as software to develop all statistical procedures.
This course is for everyone, from college students interested in using R for a project (e.g., from business, economics, statistics, life sciences, communications, etc.), or just beginners who want to improve their data analysis skills.

2. OBJECTIVES

1. Provide an overview of main multivariate data analysis technique.
2. Learn what you need to know to get started with R.
3. Perform exploratory analyses on a data set, visualizations and inferences.
4. Perform descriptive statistical analysis, confidence intervals, hypothesis testing and ANOVA test with several databases.
5. Apply R to estimate dependent and interdependence multivariate data techniques.
6. Implementation of panel data models with R.
7. Learn to validate and compare regression models using the base R package, and some other packages, if needed.
8. Perform the estimation of a systems of regression equations with R and estimation of a linear model with a limited dependent variable.
9. Understand techniques based on computer science as neural networks and Bayesian classifiers, and implementing them by using R packages.

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es

Internationalization Programs
Study Abroad 2020

3. CONTENT	
Modules	**Description**
MODULE A *[Introduction to R]*	• We will introduce the basics to start working with R. This module includes R objects, arithmetic in R, simple data entry and description, data frames and saving and loading R objects. • Students will perform several examples to understand the various fundamentals of R programming.
MODULE B *[Descriptive Statistics, Confidence intervals, Hypothesis testing and ANOVA]*	• We will review the main descriptive statistics, confidence intervals and usual hypothesis tests and introduce ANOVA for comparison of means. • Students will learn to perform descriptive statistical analysis with R by obtaining frequency tables, descriptive statistics and plots. • It also includes obtaining confidence intervals and hypothesis testing. • Students will perform analysis with several databases.
MODULE C *[Cluster and factor analysis]*	• We introduce the classification techniques and cluster analysis. Cluster analysis techniques. Students learn how to partition a given data into a set of groups according to certain criteria. • Exploratory factor analysis allows identifying the latent trait structure among a set of variables, obtaining a narrower number of variables that account for the initial data.
MODULE D *[Linear Regression]*	• We will explain the concept regression in a simple and intuitive way focusing in the different applications. We will show the students some applications of this technique in different areas, as for example marketing, business, environmental or medicine. Basic concepts of multivariate regression will be explained, as well as inference in the model, variable selection and comparisons between models.
MODULE E *[Regression analysis: Advances issues]*	• Extensions linear regression models • Simultaneous regression equations • Regression model with limited dependent variable. • Applications based on several databases.
MODULE F *[Structural Equation Modelling]*	• Introduction to structural equation models based on covariance structure modeling: applications, notation, key assumptions and modeling process. • Data preparation. • Measurement model: Confirmatory factor analysis, assessment and output interpretation. Practical examples. • Structural model: Procedure, assessment and output interpretation. Practical examples

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es

Internationalization Programs
Study Abroad 2020

MODULE G *[Discrete Choice Models]*	• Main features of discrete choice models (also known as qualitative response models. • Specification and use of models for the probabilities of events: probit and logit models. • Estimation of different discrete choice models with R. • Interpretation of discrete choice models: marginal effects. • Random utility models. • Applications of discrete choice models. Practical examples.
MODULE H *[Panel Data Analysis]*	• We explain the concept of panel data and the conditions to estimate a panel data model. • Different models for panel data are examined. In particular, static and dynamic models. • Endogeneity treatment. • Procedure to estimate panel data models with R. Recommendations in panel data analysis. Practical examples.
MODULE I *[Introduction to Neural Networks]*	• Neural networks are a set of algorithms designed to recognize patterns and can be used for classification or predictive analysis (regression). We will study the basic elements of a network, different types of activation functions and learning methods. • Using R packages, such us 'neuralnet', students will learn how to train, plot a neural network and to predict values using the network • Students will fit some neural networks, compare them and use the best option for prediction
MODULE J *[Bayesian classifiers]*	• We will introduce the concept of classification as the task of predicting the value of a target variable given some observed features. An example is the classification of a bank customer as defaulter or non-defaulter attending to the value of some observable client's features. We will show how the problem can be satisfactorily solved in an intuitive and interpretable way using the so-called Bayesian classifiers. • Students will learn how to construct, use, validate and compare Bayesian classifiers using some relevant R packages, mainly 'bnlearn' and 'naiveByes', and will be informed about public repositories containing databases related to a wide variety of domains that can be solved using classifiers. • Students will have to construct, validate and compare two classifiers over various datasets.
	• Closing session

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es

Internationalization Programs
Study Abroad 2020

4. METHODOLOGY

Students will learn how to construct, use, validate and compare regression models using the base R package for different proposed problems. If needed, some other related software packages may be used. The approach is applied a practical. Each module consists of a brief theoretical background about the technique, training in procedures with R, and exercises. All classes are taught on computer in a university computer-room.

5. PROFESSIONAL VISITS AND COMPLEMENTARY ACADEMIC ACTIVITIES

It is scheduled a professional visit to the Science and Technology Park of Almeria (http://pitalmeria.es/en/) in which we will be able to meet companies located in a technological park, the developments on data analysis of marketing research companies and other sectors, as well as the development of research in the agri-food field.

6. ASSESSMENT

The evaluation procedure for passing the course is based on classes attendance (30%) and the submission of a practical exercise in each module (70%).

7. LECTURERS

Ignacio Amate Fortes. PhD in Economics, and currently Associate Professor of the Department of Economics and Business at the University of Almeria. This researcher has developed an intense and continuous research activity that is reflected in various national and international publications of relevance, mainly in the fields of development economics, inequality, public economics and institutional economics. Website: https://www.researchgate.net/profile/Ignacio_Fortes

María Illescas Manzano. PhD candidate at Department of Economics and Business of University of Almeria. She received her Master degree of Finance and Accounting and her degree in Business Administration from University of Almeria. Her current research focuses on econometric methods to pricing in hospitality firms. Her works have been published in journals such as International Journal of Hospitality Management.

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es

UNIVERSIDAD DE ALMERÍA
Vicerrectorado de Internacionalización

Internationalization Programs
Study Abroad 2020

David Jiménez Castillo. PhD in Marketing and currently Associate Professor of Marketing in the Department of Economics and Business at University of Almeria, Spain. His current research interests include market information processing, digital marketing and relationship marketing. He has published in international refereed journals such as International Journal of Information Management, Information & Management, Journal of Business Research, Journal of Public Relations Research, Journal of Environmental Psychology, among others. Website: https://w3.ual.es/~djcasti/.

Helena Martínez Puertas. PhD in Mathematics from the University of Almeria and currently Associate Professor at Math Department of University of Almeria. Her research focuses on optimal allocation of a redundant component for series, parallel and k-out-of-n systems of more than two components. Recently, she has written several articles related to the parameter estimation in finite population, such as the mean, distribution function and quantiles by calibration techniques. Her works have been published in journals such as Sociological Methods & Research, Journal of Computational and Applied Mathematics and European Journal of Operational Research.

Sergio Martínez Puertas. PhD in Mathematics from the University of Almeria and currently Associate Professor at Math Department of the University of Almeria. His current research focuses on parameter estimation in finite population, such as the mean, distribution function and quantiles by calibration techniques. Recently, he has written several articles related to the analysis of large volumes of data through mixed structures of neural networks and articles related to econometric methods to pricing in hospitality firms. His works have been published in journals such as Sociological Methods & Research, International Journal of Hospitality Management, Applied Soft Computing, Journal of Computational and Applied Mathematics.

Andrés R. Masegosa Arredondo. PhD in Computer Science in 2009 at the University of Granada (Spain). Assistant Professor at the University of Almeria, Spain. He has been working or visiting different European universities, such as NTNU (Norway), Aalborg University (Denmark), TU Berlin (Germany), Copenhagen University (Denmark). His main research area is machine learning from a probabilistic approach. In addition to methodological developments, he has worked in applied fields such as bioinformatics, information retrieval and financial data analysis.

María E. Morales Giraldo. PhD in Mathematics and Associate Professor of Statistics and Operations Research. Her research is focus on Graph Theory and Bayesian Networks. Before teaching at University of Almeria, she worked as Data Analyst in the Unit of Statistics of the University of Almeria.

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es

Internationalization Programs
Study Abroad 2020

Rafael Rumí Rodríguez. PhD in Mathematics and Associate Professor of Statistics and Operations Research at the Math Department of the University of Almeria. He has been the principal investigator of two RTD projects at national level and one of regional level, and has supervised two PhD thesis in the area of machine learning and probabilistic graphical models. He has a long record of publications in relevant international journals and conferences, covering aspects of hybrid Bayesian networks, probabilistic decision graphs, classification, regression and important applications in the field of economy and environment. Since 2015 he is the Director of Teaching Affairs in the University of Almeria. Website: http://www.ual.es/personal/rrumi.

Antonio Salmerón Cerdán. PhD in Artificial Intelligence by the University of Granada and currently Professor of Statistics and Operations Research at the Math Department of the University of Almeria. He was Head of Department of Mathematics from 2015 to 2019 and also Head of the Doctoral School of the University of Almeria from 2001 to 2007. Since 2001, he has been the principal investigator of a number of RTD projects at national level, including successful industrial collaborations, and was a work-package leader in the FP7-funded AMIDST project. He has a long record of publications in relevant international journals and conferences, covering aspects of approximate inference in Bayesian networks, hybrid Bayesian networks (including modelling, inference and learning), probabilistic decision graphs, classification and regression. In 2001, he got the José Cuena award from the Spanish Association for Artificial Intelligence. Website: http://www.ual.es/personal/asalmero.

Manuel Sánchez Pérez. PhD in Economics and Business, and currently Professor of Marketing at the Economics and Business Department of the University of Almeria. His research interests are focused on marketing models, marketing strategy, tourism and food channels. He has been the principal investigator of several national research project, collaborating in others. Also, he has been supervisor of several PhD dissertations. At present, he is the Head of the Department of Economics and Business. Website: https://w3.ual.es/~msanchez/.

Organized by:

Internationalization Vicerectorate
UNIVERSIDAD DE ALMERÍA
Tel. +34 950 01 5816
E-mail: sabroad@ual.es

Ctra.Sacramento La Cañada de San Urbano 04120 Almería (España) Telf.: 34 950 015973 FAX: 34 950 015973 www.ual.es e-mail: reiual@ual.es