FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Análisis Multivariante
Código	E000012908
Impartido en	Grado en Administración y Dirección de Empresas (E-2) [Tercer Curso] Grado en Administración y Dirección de Empresas con Mención en Internacional (E-4) [Tercer Curso] Grado en Administración y Dirección de Empresas (E-2) - Bilingüe en inglés [Tercer Curso]
Nivel	Reglada Grado Europeo
Cuatrimestre	Semestral
Créditos	6,0 ECTS
Carácter	Optativa (Grado)
Departamento / Área	Departamento de Métodos Cuantitativos
Responsable	Antonio Rua Vieites

Datos del profesorado	
Profesor	
Nombre	Enrique Díaz-Plaza Sanz
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	ediazp@comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación

In today's data-driven world, the ability to analyze and interpret data is crucial for professionals in various fields. This course in Data Analysis aims to equip students with essential skills and techniques to extract meaningful insights from data, make informed decisions and contribute effectively to their professional roles.

This course aims to empower students to become data-savvy professionals capable of making informed and impactful contributions across various industries, fostering a data-driven culture in the workplace. Through hands-on training and practical projects, students will gain the confidence to tackle real-world data challenges and make a difference in their future careers. Thus, by the end of this course, students will have a well-rounded understanding of data analysis and its practical applications in a business setting.

The course will cover a wide range of analytical methods, focusing on practical applications and their relevance in real-world business scenarios. As highlights, the course will cover main blocks such as Practical Data Exploration, Dependency and Interdependency Analysis, Variance and Contextual Analysis.

The course will incorporate real-world case studies and projects inspired by industry challenges. Students will work with diverse datasets, apply analysis techniques, and derive actionable insights, simulating scenarios encountered in the professional world.

Also, students will develop proficiency in using statistical software like R and Python to manipulate, visualize, and analyze data. This practical skillset is highly sought-after in the industry, making graduates well-prepared for data-intensive roles.

Prerequisitos

To have completed the subjects from previous courses (especially those related to Statistics) and to have an open mind and the ability to work to learn new concepts related to data science.

Competencias - Objetivos

Competencias

CG01 Capacidad de análisis y síntesis

RA1 Analiza la información identificando sus elementos más significativos

RA2 Realiza   la abstracción y simp lificación necesa ria para modelizar estadísticamente el problema   real planteado

RA3 Integra el análisis gráfico, verbal y los datos cuantitativos y cualitativos para definir el modelo estadístico apropiado al problema

CG02 Resolución de problemas y toma de decisiones

RA1 Conoce las metodologías aplicables para resolver el problema real

RA2 Sabe aplicar dichas metodologías

RA3 Reconoce el alcance de las conclusiones y los supuestos necesarios para la validez de las mismas

CG04 Capacidad de gestionar información proveniente de fuentes diversas

RA1 Conoce fuentes y bases de datos profesionales macro y microeconómicas

RA2 Sabe tratar y juzgar críticamente las fuentes utilizadas para que sean válidas en el análisis

CG05 Conocimientos generales básicos sobre el área de estudio

RA1 Desarrolla habilidades necesarias para el estudio e investigación independiente

RA2 Encuentra por sí mismo aplicaciones y extensiones de los conceptos y metodologías estudiadas

CG06 Comunicación oral y escrita en la propia lengua

RA1 Expresa correctamente la metodología empleada y los resultados y conclusiones obtenidas del análisis efectuado

RA2 Comunica de forma efectiva, a público no especialista, los informes y análisis efectuados

CG08 Conocimientos de informática relativos al ámbito de estudio

RA1 Conoce y emplea de forma suficiente herramientas informáticas de uso común para el análisis estadístico

ESPECÍFICAS

CEOPT Conocimiento y comprensión de las principales Técnicas de Predicción y el Análisis Multivariante

RA1   Conocer los principales modelos econométricos empleados en el ámbito de las finanzas especialmente los relativos a la modelización de la volatilidad

RA2   Conocer los principales modelos econométricos empleados en el ámbito de la economía aplicada tanto en el campo microeconómico como macroeconómico

RA3 Conocer las principales técnicas clásicas de dependencias e interdependencias, sabiendo seleccionar la más adecuada según el tipo de problemas y el conjunto de datos disponible

 RA4 Adquirir una perspectiva general de las nuevas técnicas de análisis de datos aplicables al nuevo paradigma del análisis masivo de datos (Big Data)

RA5 Conocer los conceptos básicos de los procesos de predicción económica y empresarial

RA6 Saber realizar predicciones , seleccionando y aplicando la técnica más adecuada en cada caso

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos
BLOCK 1: Introduction
1.- Introduction to Data Analysis
1.1.- Overview of Data Analysis 1.2.- Data Analysis Process 1.3.- Types of Data and Data Sources 1.4.- Informational Architectures
2.- General Overview of Tools Used in Data Analysis
2.1.- Introduction to R 2.2.- Introduction to Python 2.3.- Environments 2.4.- Cloud Computing for Data Analysis
3.- Exploratory Data Analysis
3.1.- Descriptive Statistics 3.2.- Data Visualization Techniques 3.3.- Handling Missing Data 3.4.- Outlier Detection and Treatment 3.5.- Data Cleaning and Preprocessing
BLOCK 2: Dependency Analysis
4.- Correlation Analysis
4.1.- Introduction to Correlation 4.2.- Pearson's Correlation Coefficient 4.3.- Spearman's Rank Correlation Coefficient 4.4.- Kendall's Correlation Coefficient 4.5.- Correlation Matrix and Heatmaps
5.- Variance Analysis
5.1.- One-Way ANOVA (Analysis of Variance)

5.2.- Two-Way ANOVA

5.3.- ANCOVA (Analysis of Covariance)

6.- Discriminant Analysis

6.1.- Introduction to Discriminant Analysis

6.2.- Linear Discriminant Analysis (LDA)

6.3.- Quadratic Discriminant Analysis (QDA)

6.4.- Support Vector Machines (SVM)

7.- Regression Analysis for Categorical Variables: Exploring Relationships and Predictive Models

7.1.- Linear Regression

7.2.- Logistic Regression

7.3.- Multinomial Logistic Regression

7.4.- Ordinal Logistic Regression

7.5.- Poisson Regression (brief overview)

BLOCK 3: Interdependency Analysis

8.- Network Analysis

8.1.- Introduction to Network Analysis

8.2.- Graph Theory Fundamentals

8.3.- Network Metrics and Centrality Measures

8.4.- Community Detection in Networks

9.- Clustering Analysis

9.1.- Introduction to Clustering

9.2.- K-means Clustering

9.3.- Hierarchical Clustering

9.4.- Evaluation of Clustering Results

10.- Factor Analysis

10.1.- Introduction to Factor Analysis

10.2.- Factor Extraction Methods (e.g., Principal Component Analysis, Maximum Likelihood)

10.3.- Interpreting Factor Loadings and Communalities

10.4.- Applications of Factor Analysis

11.- Principal Component Analysis (PCA)

11.1.- Introduction to PCA

11.2.- Dimensionality Reduction with PCA

11.3.- Interpreting Principal Components

11.4.- Applications of PCA

12.- Association Rule Mining

12.1.- Introduction to Association Rule Mining

12.2.- Apriori Algorithm

12.3.- Evaluation of Association Rules

12.4.- Applications of Association Rule Mining

13.- Decision Trees and Random Forests

13.1.- Introduction to Decision Trees

13.2.- Splitting Criteria and Tree Pruning

13.3.- Ensemble Learning with Random Forests

13.4.- Feature Importance and Model Evaluation

14.- Time Series Analysis

14.1.- Introduction to Time Series Data

14.2.- Time Series Visualization

14.3.- Trend Analysis

14.4.- Seasonality and Decomposition

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura
Metodología Presencial: Actividades
- Mastery exposition of the general framework of each topic.   - Realization and discussion of introductory examples of practical application.   - General tutoring of practical application work.   - Introduction to and guidelines for the use of computer applications of a statistical nature, as well as obtaining and processing economic data from web sources.   - Carrying out practical tests by the students and defending them, if applicable.
Metodología No presencial: Actividades
- Preparation and realization of practical cases to replicate the structure and contents explained in class.   - Study and preparation of the final exam.

RESUMEN HORAS DE TRABAJO DEL ALUMNO

CLASSROOM HOURS	
Expository lessons	Exercises and solving cases and problems
30.00	30.00

NON-PRESENTIAL HOURS		
Individual and/or group study and organized reading	Exercises and resolution of cases and problems.	Monographic and research papers, indiv or collective.
30.00	30.00	40.00

ECTS CREDITS: 6,0 (160,00 h

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Evaluation Activities	Evaluation Criteria	Weight
Multiple-choice exam covering the entire course content, without the use of computer support.	Numerical grading from 0 to 10.	50%
Coursework and labs to be carried out based on the course content.	Numerical grading from 0 to 10.	35%
Attention and participation in class.	Numerical grading from 0 to 10.	10%
Class attendance.	Numerical grading from 0 to 10.	5%

Calificaciones

To pass the course, the student must obtain a grade equal to or higher than 5 (out of a maximum of 10) overall. Additionally, to calculate the average of all assessable parts of the subject, the student must obtain a grade equal to or higher than 4 in the final exam (if the grade in the final exam is below 4.0, the subject will be considered failed, regardless of the grades obtained in the rest of the tests).

In the course, depending on the needs and characteristics of the tests to be carried out, specialized computer support will be used, such as applications and languages like R and Python, as well as cloud environments or applications based on LLM or other technologies, such as ChatGPT, for example. The use of these tools should serve to help and facilitate the student's learning, applied to the different phases that need to be followed in the development of the assigned work. Therefore, it will not replace the student's problem-solving skills or the basic knowledge required to address the given case, but it will assist and complement them where appropriate.

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Text books

Grolemund, G., y Wickham,H.: "R for Data Science".O'Reilly Media. 2017.

Field, Andy; Miles, Jeremy; Field, Zoë: "Discovering Statistics Using R. ISBN: 9781446258460. SAGE. 2012.

Everitt, Brian; Hothorn, Torsten. "An Introduction to Applied Multivariate Analysis with R". ISBN: 9781441996497. Springer New York, NY. 2011.

Johnson Richard A., Wichern, Dean W. "Applied Multivariate Statistical Analysis". ISBN: 9780131877153. Pearson. 2007.

Dalgaard, Peter. "Introductory Statistics with R". ISBN: 9780387790534. Springer New York, NY. 2008.

McKinney, Wes. "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython". ISBN: 9781491957639. O'Reilly Media, Inc. 2017.

Aldas Manzano, J.; UrielJiménez, E. "Análisismultivariante aplicado con R". ISBN: 788428329699. Ed. Paraninfo. 2017.

Garza, J.; Morales, B.; González, B.. "Análisis estadístico multivariante, Un enfoque práctico". ISBN: 9786071508171. Ed. Mc Graw Hill. 2013.

Bibliografía Complementaria

Neter, John; Wasserman, William; Kutner, Michael H. "Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs". ISBN: 978-0256083385. CRC Press. 1990.

Mangin, Lévy; et al. "AnálisisMultivariable para las Ciencias Sociales". Editorial Pearson. 2003.

Peña, D. "Análisis de Datos Multivariantes". Editorial Mc Graw Hill. 2002.

Hair, Anderson; Tatham, Black. "Análisis Multivariante". Editorial Prentice Hall. 1999.

Gareth, James.; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert. "An introduction to statistical learning: with applications in R". Springer. 2021