

---

**Ano Letivo** 2023-24

---

**Unidade Curricular** ANÁLISE AVANÇADA DE DADOS

---

**Cursos** TURISMO (3.º Ciclo)

---

**Unidade Orgânica** Faculdade de Economia

---

**Código da Unidade Curricular** 30001091

---

**Área Científica** MÉTODOS QUANTITATIVOS

---

**Sigla**

---

**Código CNAEF (3 dígitos)** 460

---

**Contributo para os Objetivos de  
Desenvolvimento Sustentável - 4;8  
ODS (Indicar até 3 objetivos)**

---

**Línguas de Aprendizagem** Inglês

---

**Modalidade de ensino**

Presencial (e/ou à distância)

---

**Docente Responsável**

Patrícia Susana Lopes Guerrilha dos Santos Pinto

---

DOCENTE	TIPO DE AULA	TURMAS	TOTAL HORAS DE CONTACTO (*)
---------	--------------	--------	-----------------------------

\* Para turmas lecionadas conjuntamente, apenas é contabilizada a carga horária de uma delas.

---

ANO	PERÍODO DE FUNCIONAMENTO*	HORAS DE CONTACTO	HORAS TOTAIS DE TRABALHO	ECTS
1º	T2	36S; 12OT	560	20

\* A-Anual;S-Semestral;Q-Quadrimestral;T-Trimestral

---

**Precedências**

Sem precedências

---

**Conhecimentos Prévios recomendados**

Inferência estatística

---

**Objetivos de aprendizagem (conhecimentos, aptidões e competências)**

Complementar o conhecimento de análise de regressão, de forma a selecionar e aplicar técnicas apropriadas para a análise de determinadas questões de investigação; reconhecer as limitações do modelo linear probabilístico e as potencialidades da regressão logística; conduzir testes de hipóteses em modelos binomiais, multinomiais e de variável dependente limitada e interpretar os resultados; distinguir diferentes modelos de sucessões cronológicas de forma a selecionar e saber aplicar o modelo adequado; validar modelos de equações estruturais com variáveis latentes, enquanto modelos de estrutura de covariância ou segundo a abordagem *pls*, utilizando um software adequado; saber validar uma solução na análise de clusters e dominar as aprendizagens de *machine learning*.

---

**Conteúdos programáticos**

I - Modelos de Variável Dependente Binária

Modelo de Probabilidade Linear

Modelos Logit e Probit

II - Modelos Multinomiais e Modelos de Variável Dependente Limitada

Modelos de resposta não ordenada

Modelos de resposta ordenada

Amostras truncadas

dados censurados

III - Séries Temporais

Estacionaridade

modelos ARMA

modelos GARCH

IV - Modelos de Equações Estruturais com Variáveis Latentes

CB-SEM vs PLS-SEM

Modelos, variáveis, equações e parâmetros

Passos para aplicar o método

Avanços em SEM

SmartPLS e Amos

V - Análise de Cluster

Medidas de similaridade

Algoritmos de agrupamento

Número de clusters

Perfil e validação

VI ? *Machine Learning*

Conceito de *Machine Learning*

Aprendizagem Supervisionada

Aprendizagem não supervisionada

Mineração de texto

Estudo de caso

---

### **Metodologias de ensino (avaliação incluída)**

A componente letiva da unidade curricular estrutura-se em 6 módulos, a que correspondem um total de 36 horas teórico-práticas de ensino e 12 horas de orientação tutorial.

De avaliação obrigatória por trabalho individual num dos módulos. Nesse módulo o estudante seguirá as linhas de orientação facultadas pelo respetivo docente.

---

### **Bibliografia principal**

Brooks, C. (2019), *Introductory Econometrics for Finance*, 4th Edition, Cambridge, UK.

Cameron, A. and Triverdi, P. ( 2022 ), *Microeconometrics Using Stata*, Stata Press.

Collier, J.E. (2020), *Applied Structural Equation Modeling Using Amos*. NY: Routledge.

Dolnicar, S., Grun, B., & Leisch, F. (2018). *Marketing Segmentation Analysis: Understanding It, Doing It, and Making It Useful*. Springer.

Egger, R. (Ed.). (2022). *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*. Springer Nature.

Hair J.F., Hult G.T., Ringle C. (2017), *A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM)* (2nd Edition), Los Angeles, CA: SAGE Publications.

Orme, B. (2018). *Getting started with conjoint analysis: Strategies for product design and pricing research*. Research Publishers LLC.

---

**Academic Year** 2023-24

---

**Course unit** Advanced Data Analysis

---

**Courses** TOURISM (3rd cycle)

---

**Faculty / School** THE FACULTY OF ECONOMICS

---

**Main Scientific Area**

---

**Acronym**

---

**CNAEF code (3 digits)** 460

---

**Contribution to Sustainable Development Goals - SGD (Designate up to 3 objectives)** 4;8

---

**Language of instruction** English

---

**Teaching/Learning modality** Face-to-face (and/or by remote means)

**Coordinating teacher**                      Patrícia Susana Lopes Guerrilha dos Santos Pinto

Teaching staff	Type	Classes	Hours (*)
----------------	------	---------	-----------

\* For classes taught jointly, it is only accounted the workload of one.

Contact hours	T	TP	PL	TC	S	E	OT	O	Total
	0	0	0	0	36	0	12	0	560

T - Theoretical; TP - Theoretical and practical ; PL - Practical and laboratorial; TC - Field Work; S - Seminar; E - Training; OT - Tutorial; O - Other

**Pre-requisites**

no pre-requisites

**Prior knowledge and skills**

Statistical inference

**The students intended learning outcomes (knowledge, skills and competences)**

Complement the knowledge on regression analysis, in order to select and apply the appropriate techniques to modelling and analyzing relevant research questions; recognize the limitations of the Linear Probability Model when compared to the Binary Logistic Regression; conduct hypothesis testing in the context of binomial, multinomial, and limited dependent variable models; distinguish between different types of time series models so that the appropriate model can be chosen and properly applied; estimate and validate structural equation models with latent variables, either as covariance structured models or as a result of a pls approach; validate a particular solution using cluster analysis; dominate both the supervised and unsupervised machine learning techniques.

## Syllabus

### I - Binary Dependent Variable Models

Linear Probability Model  
Logit & Probit models

### II - Multinomial Models & Limited Dependent Variable Models

Unordered response models  
Ordered response models  
Truncated samples  
Censored data

### III - Time Series

Stationarity  
ARMA models  
GARCH models

### IV - Structural Equations Models with Latent Variables

CB-SEM vs PLS-SEM  
Models, variables, equations and parameters  
Steps to apply the method  
Advances in SEM  
SmartPLS and Amos

### V - Cluster Analysis

Similarity measures  
Clustering algorithms  
Number of clusters  
Profile and validation

### VI ? Machine Learning

Machine Learning concepts  
Supervised Learning  
Unsupervised Learning  
Text Mining  
Case study

### Teaching methodologies (including evaluation)

The lecture component is structured in 6 modules, corresponding to a total of 36 hours of mixed lecture classes and 12 hours of tutorial orientation.

The students will submit, to be marked, an individual essay covering one of the modules's contents. To prepare the essay, the students will follow the gridelines provided by the lecturer.

---

### Main Bibliography

Brooks, C. (2019), *Introductory Econometrics for Finance*, 4th Edition, Cambridge, UK.

Cameron, A. and Triverdi, P. ( 2022 ), *Microeconometrics Using Stata*, Stata Press.

Collier, J.E. (2020), *Applied Structural Equation Modeling Using Amos*. NY: Routledge.

Dolnicar, S., Grun, B., & Leisch, F. (2018). *Marketing Segmentation Analysis: Understanding It, Doing It, and Making It Useful*. Springer.

Egger, R. (Ed.). (2022). *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*. Springer Nature.

Hair J.F., Hult G.T., Ringle C. (2017), *A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM)* (2nd Edition), Los Angeles, CA: SAGE Publications.

Orme, B. (2018). *Getting started with conjoint analysis: Strategies for product design and pricing research*. Research Publishers LLC.