
Ano Letivo 2022-23

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 - ODS (Indicar até 3 objetivos) 4;8

Línguas de Aprendizagem Inglês

Modalidade de ensino

Presencial (e/ou à distância)

Docente Responsável

Efigénio da Luz Rebelo

DOCENTE	TIPO DE AULA	TURMAS	TOTAL HORAS DE CONTACTO (*)
Efigénio da Luz Rebelo	S	S1	9S
Patrícia Susana Lopes Guerrilha dos Santos Pinto	S	;S1	6S
Jorge Miguel Lopo Gonçalves Andraz	S	;S1	6S
Célia Maria Quitério Ramos	S	;S1	6S
Guilherme José Fresca Mirador de Andrade Castela	S	;S1	12S
Luís Miguel Soares Nobre de Noronha e Pereira	S	;S1	6S

* 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; compreender os conceitos da análise categórica, e perceber o interesse da análise conjunta; dominar os métodos de três vias para a análise de dados e perceber os procedimentos *STATIS* e *COSTATIS*; saber validar uma solução na análise de clusters e dominar as aprendizagens de *machine learning*.

Conteúdos programáticos

1. Modelos com Variável Dependente Binária, Análise de regressão, Modelo Linear Probabilístico versus Modelos *Logit* e *Probit*
2. Modelos Multinomiais e de Variável Dependente Limitada, Modelos de resposta sem ordem e modelos de resposta ordenada, Amostras truncadas, Dados censurados
3. Sucessões Cronológicas, Estacionaridade, Modelos de tipo *ARMA*, Modelos de tipo *GARCH*
4. Modelos de Equações Estruturais com Variáveis Latentes, *CB-SEM* vs *PLS-SEM*, Modelos, variáveis, equações e parâmetros, Aplicação do método e Desenvolvimentos em *SEM*, *SmartPLS* e *AMOS*
5. Análise Categórica e Procedimento *STATIS*, Estruturas de dados, Análise Categórica de Componentes Principais, Análise de co-inércia, *STATIS*, *COSTATIS*, Análise Conjunta
6. Análise de Clusters, Medidas de semelhança, Algoritmos de formação de clusters, Número de clusters, Perfil e validação
7. Aprendizagem automática, Conceitos de aprendizagem automática, Aprendizagem supervisionada, Aprendizagem não supervisionada, Mineração de texto.

Metodologias de ensino (avaliação incluída)

A componente letiva da unidade curricular estrutura-se em 7 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, em pelo menos um dos módulos, nesse(s) módulo(s) o estudante terá o acompanhamento do(s) docente(s).

Bibliografia principal

Abdi, H., Williams, L. J., Valentin, D. e Bennani-Dosse, M. (2012), STATIS and DISTATIS: Optimum Multitable Principal Component Analysis and Three-way Metric Multidimensional Scaling. *Wiley Interdisciplinary Reviews: Computational Statistics* . 4 (2), 124-167.

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

Cameron, A. and Triverdi, P. (2005), *Microeconometrics: Methods and Applications* , Cambridge, UK.

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

Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th Edition). Wiley, USA.

Gujarati, D.N. (2020), *Basic Econometrics* , McGrawHill

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.

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

Academic Year 2022-23

Course unit Advanced Data Analysis

Courses TOURISM

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 Efigénio da Luz Rebelo

Teaching staff	Type	Classes	Hours (*)
Efigénio da Luz Rebelo	S	S1	9S
Patrícia Susana Lopes Guerrilha dos Santos Pinto	S	;S1	6S
Jorge Miguel Lopo Gonçalves Andraz	S	;S1	6S
Célia Maria Quitério Ramos	S	;S1	6S
Guilherme José Fresca Mirador de Andrade Castela	S	;S1	12S
Luís Miguel Soares Nobre de Noronha e Pereira	S	;S1	6S

* 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; understand the main concepts and techniques of categorical analysis and realize the interest on conjoint analysis; understand the three-way methods to analyze data, and understand the STATIS and COSTATIS procedures; validate a particular solution using cluster analysis; dominate both the supervised and unsupervised machine learning techniques.

Syllabus

Module I - Binary Dependent Variable Models

1. Classical regression analysis
2. Linear Probability Model
3. Logit & Probit models

Module II - Multinomial Models & Limited Dependent Variable Models

1. Unordered response models
2. Ordered response models
3. Truncated samples
4. Censored data

Module III - Time Series

1. Stationarity
2. ARMA models
3. GARCH models

Module IV - Structural Equations Models with Latent Variables

1. CB-SEM vs PLS-SEM
2. Models, variables, equations and parameters
3. Steps to apply the method
4. Advances in SEM
5. SmartPLS and Amos

Module V - Categorical Analysis and STATIS Procedure

1. Data structures
2. Main Component Categorical Analysis
3. Co-inertia Analysis
4. STATIS
5. COSTATIS
6. Conjoint Analysis

Module VI - Cluster Analysis

1. Similarity measures
2. Clustering algorithms
3. Number of clusters
4. Profile and validation

Módulo VII ? Machine Learning

1. Machine Learning concepts
2. Supervised Learning
3. Unsupervised Learning
4. Text Mining
5. Case study

Teaching methodologies (including evaluation)

The lecture component is structured in 7 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 at least one of the modules's contents. To prepare the essay, the students will be assisted by the lecturer(s).

Main Bibliography

Abdi, H., Williams, L. J., Valentin, D. e Bennani-Dosse, M. (2012), STATIS and DISTATIS: Optimum Multitable Principal Component Analysis and Three-way Metric Multidimensional Scaling. *Wiley Interdisciplinary Reviews: Computational Statistics* . 4 (2), 124-167.

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

Cameron, A. and Triverdi, P. (2005), *Microeconometrics: Methods and Applications* , Cambridge, UK.

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

Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th Edition). Wiley, USA.

Gujarati, D.N. (2020), *Basic Econometrics* , McGrawHill

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.

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