



POLITÉCNICA

INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Informáticos

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000896 - Statistical Data Analysis

DEGREE PROGRAMME

10BA - Master Universitario En Ciencia De Datos

ACADEMIC YEAR & SEMESTER

2022/23 - Semester 1

Index

Learning guide

1. Description.....	1
2. Faculty.....	1
3. Prior knowledge recommended to take the subject.....	2
4. Skills and learning outcomes	3
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	6
7. Activities and assessment criteria.....	8
8. Teaching resources.....	10
9. Other information.....	12

1. Description

1.1. Subject details

Name of the subject	103000896 - Statistical Data Analysis
No of credits	4.5 ECTS
Type	Compulsory
Academic year of the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10BA - Master Universitario en Ciencia de Datos
Centre	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Academic year	2022-23

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Juan Antonio Fdez Del Pozo De Salamanca	2105	juan.fdezpozo.salamanca@upm.es	Sin horario. Check office hours in September on Moodle
Antonio Jimenez Martin	2110	antonio.jimenez@upm.es	Sin horario. Check office hours in September on Moodle

Jacinto Gonzalez Pachon	2105	jacinto.gonzalez.pachon@upm.es	Sin horario. Check office hours in September on Moodle
Arminda Moreno Diaz (Subject coordinator)	2204	arminda.moreno@upm.es	Sin horario. Check office hours in September on Moodle
Maria Isabel Rodriguez Galiano	2204	mariaisabel.rodriguez@upm.es	Sin horario. Ckeck office hours in September on Moodle

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- Basic knowledge of statistics and inference will be helpful.

4. Skills and learning outcomes *

4.1. Skills to be learned

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CECD04 - Capacidad para aplicar técnicas para la generación de visualizaciones adecuadas a cada problema para el análisis y la exploración de datos, y para la correcta comunicación de los resultados del análisis.

CECD05 - Conocer y tener capacidad para aplicar técnicas de análisis estadístico avanzado a tareas de modelado y análisis de datos heterogéneos y para predicción.

CG07 - Aplicación de los últimos o más novedosos métodos para resolver problemas que, posiblemente, involucren a otras disciplinas

CG08 - Capacidad de pensamiento de forma creativa para desarrollar aproximaciones y métodos nuevos y originales

CG14 - Capacidad de trabajar y comunicarse también en contextos internacionales

CGI02 - Comprender el procedimiento, valor y límites del método científico en el campo de la Informática, siendo capaz de identificar, localizar y obtener datos requeridos en un trabajo de investigación, de diseñar y guiar investigaciones analíticas, de modelado y experimentales, así como de evaluar datos de una manera crítica y extraer conclusiones

4.2. Learning outcomes

RA2 - Conocer y aplicar técnicas de reducción de dimensionalidad y modelización de datos multivariantes.

RA1 - Conocer y aplicar las principales técnicas para explorar, describir y analizar datos multivariantes.

RA3 - Aplicar la metodología apropiada para el ajuste de series temporales.

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

The course is intended to be a non-exhaustive survey of techniques to convert multivariate data into useful information so that good decisions can be made. The perspective is twofold, theoretical and applied, covering topics such as: exploratory data analysis, statistical summaries and graphical representations, dimensionality reduction, regression techniques and time series analysis. There will be an emphasis on hands-on application of the theory and methods throughout, with extensive use of R. The course is taught in three modules which are broken down in topics below and properly referenced in the timeline.

5.2. Syllabus

1. Using data to answer statistical questions
 - 1.1. Gathering data. Experiments and observational studies.
 - 1.2. Review of univariate inference. Bootstrap and confidence intervals. Permutation tests.
 - 1.3. Methods for comparing two groups.
2. Exploratory data analysis and descriptive statistics.
 - 2.1. Aspects of multivariate data: exploration and visualization.
 - 2.2. Aspects of multivariate data: descriptive statistics.
 - 2.3. Aspects of multivariate data: dimensionality reduction.
3. Statistical modeling.
 - 3.1. Simple Linear Regression.

3.2. Multiple Linear Regression.

3.3. The General Linear Model.

4. Time Series.

4.1. Definitions, Applications and Techniques.

4.2. Stationarity and Seasonality.

4.3. Common approaches.

4.4. Box-Jenkins model identification, estimation and validation.

4.5. Forecasting.

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	Module 1: Lecture Duration: 02:00			
2	Module 1: Lecture Duration: 01:00	Computer Lab. Module 1. Duration: 01:00		
3	Module 1: Lecture Duration: 02:00			
4	Module 2: Lecture Duration: 02:00			
5	Module 2: Lecture Duration: 01:00	Computer Lab. Module 2. Duration: 01:00		
6	Module 2: Lecture Duration: 01:00	Computer Lab. Module 2. Duration: 01:00		
7	Module 3: Lecture Duration: 02:00			
8	Module 3: Lecture Duration: 02:00			Exams on Module 1 and 2. Continuous assessment Presential Duration: 02:00
9	Module 3: Lecture Duration: 02:00			
10	Module 3: Lecture Duration: 02:00			
11	Module 4: Lecture Duration: 02:00			
12	Module 4: Lecture Duration: 02:00			
13	Module 4: Lecture Duration: 01:00	Computer Lab. Module 4. Duration: 01:00		

14	Module 4: Lecture Duration: 01:00	Computer Lab. Module 4. Duration: 01:00		
15				Exams on Module 3 and 4 Continuous assessment Presential Duration: 02:00
16	Review Duration: 02:00			
17				Global exam: 4 exams, one corresponding to each module. Only for students that haven't passed the course yet. Final examination Presential Duration: 03:00

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
8	Exams on Module 1 and 2.		Face-to-face	02:00	50%	3 / 10	CB08 CG14 CECD04 CECD05 CGI02
15	Exams on Module 3 and 4		Face-to-face	02:00	50%	3 / 10	CG08 CG14 CB10 CECD05 CG07

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Global exam: 4 exams, one corresponding to each module. Only for students that haven't passed the course yet.		Face-to-face	03:00	100%	3 / 10	CGI02 CB08 CG08 CG14 CB10 CECD04 CECD05 CG07

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Extraordinary exam: all 4 exams, one corresponding to each module.		Face-to-face	03:00	100%	3 / 10	CGI02 CB08 CG08 CG14 CB10 CECD04 CECD05 CG07

7.2. Assessment criteria

Progressive evaluation:

The overall course grade will be based on the performance on 4 exams, equally weighed, corresponding to each of the modules taught. The weeks in which each exam is scheduled might change depending on the academic calendar.

Each exam should obtain a grade equal or greater than 3 (0-10 scale) for the overall course grade to be computed. The overall course grade will be the average of the grades obtained in the 4 exams and it must be equal or greater than 5 to pass the course. No makeup exams will be scheduled.

January exam:

As far as exams are concerned, the final exam in January is only for those failing to take the midterms when requested or those obtaining an overall grade less than 5. In the latter case, only those specific exams not previously passed have to be taken again.

Extraordinary exam:

In case you are opting for the extraordinary exam in July's session, you have to warn the instructor(s) in advance (at least two weeks before the exam, by email, but the sooner the better). If you have taken some of the exams on

the ordinary session and obtained a grade greater or equal than 3, these grades can be kept and no repetition of these exams is required. Otherwise, all 4 exams can be taken in this opportunity.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Moodle	Web resource	Main communication channel with students. Repository: slides, scripts, data sets, other resources.
Johnson, R.A., Whichern, D.W. (2007) Applied Multivariate Statistical Analysis. Pearson Education.	Bibliography	Friendly exposition of the most important multivariate techniques, including clustering. They also introduce other Artificial Intelligence techniques like neural networks.
Rencher, A.C. Methods of Multivariate Analysis.	Bibliography	Clear exposition of Multivariate Analysis Techniques, from a statistical point of view. Many examples.
Everitt, B.S. and Dunn G. (1997) Applied Multivariate Data Analysis. Arnold.	Bibliography	Excellent exposition of multivariate techniques. They make the Generalised Linear Model easily understandable.
Hair, J.F., Black, W.C., Babin, B.J., Anderson R.E. Multivariate Data Analysis.	Bibliography	A Global Perspective on multivariate Techniques. Very detailed examples. In almost every topic, a "Rules of Thumb" section summarizes the relevant facts.

Sharma, S (1996). Applied Multivariate Techniques. Wiley.	Bibliography	
Multivariate Analysis of Ecological Data. Greenacre, Primicerio. Fundación BBVA.	Web resource	http://www.fbbva.es/TLFU/tlfu/esp/publicaciones/libros/fichalibro/index.jsp?codigo=769 Excellent compilation and description of multivariate techniques applied to Ecological Data. Clear, educational, didactic explanations. Lots of examples
Biplots in Practice. Michael Greenacre. Fundación BBVA.	Web resource	http://www.multivariatestatistics.org/biplots.html Excellent Monograph on Biplots.
Rawlings, J.O., Pantula, S.G., Dickey, D.A. Applied Regression Analysis.	Bibliography	Almost everything about Regression Models.
Chatfield, C. (2003) The Analysis of Time Series: An Introduction. Chapman and Hall.	Bibliography	
Robert Hyndman, George Athanasopoulos. Forecasting: principles and practice. OTexts, 2018.	Web resource	Bibliography and web resource. Available at https://otexts.com/fpp2/
Robert Nau. Statistical forecasting: notes on regression and time series analysis.	Web resource	Excelent set of tips to fit ARIMA models. Available at https://people.duke.edu/~rnau/411home.htm

9. Other information

9.1. Other information about the subject

Regarding the Sustainable Developing Goals, broadly speaking, proper data analysis leads to an understanding of processes and the generation of information available to everyone. This information can help clarify and make more transparent the way in which institutions, for instance, try to meet people's needs. This information increases transparency, accountability and citizen participation in institutions and facilitates information-sharing between different actors. In this particular case, proper analysis of data might help in achieving SDG 16: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.