ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000896 - Statistical Data Analysis

DEGREE PROGRAMME

10BA - Master en Ciencia de Datos

ACADEMIC YEAR & SEMESTER

2019/20 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	103000896 - Statistical Data Analysis
No of credits	4.5 ECTS
Type	Compulsory
Academic year ot the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10BA - Master en Ciencia de Datos
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2019-20

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
Antonio Jimenez Martin	2110	antonio.jimenez@upm.es	Sin horario. Check office hours in September

Jacinto Gonzalez Pachon	2105	jacinto.gonzalez.pachon@upm.es	Sin horario. Check office hours in September
Arminda Moreno Diaz (Subject coordinator)	2204	arminda.moreno@upm.es	Sin horario. Check office hours in September

* 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

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.

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

* 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 techinques to convert multivariate data into useful information so that good decisions can be made. The perspective is twofold, theorical 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. Exploratory data analysis and descriptive statistics.
    1.1. Aspects of multivariate data:exploration and visualization.
    1.2. Aspects of multivariate data: manipulation.
    1.3. Aspects of multivariate data: descriptive statistics.
    1.4. Aspects of multivariate data: dimensionality reduction.
2. Statistical modeling.
    2.1. Simple Linear Regression.
    2.2. Multiple Linear Regression.
    2.3. The General Linear Model.
3. Time Series.
    3.1. Definitions, Applications and Techniques.
    3.2. Stationarity and Seasonality.
    3.3. Common approaches.
    3.4. Box-Jenkins model identification, estimation and validation.
    3.5. Forecasting.

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1	**Module 1: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00
2	**Module 1: Lecture** Duration: 01:00	**Computer Lab. Module 1.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00
3	**Module 1: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 1** Continuous assessment Duration: 02:00
4	**Module 1: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 1** Continuous assessment Duration: 02:30
5	**Module 1: Lecture** Duration: 01:00	**Computer Lab. Module 1.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 1** Continuous assessment Duration: 04:30
6	**Module 1: Lecture** Duration: 01:00	**Computer Lab. Module 1.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 1** Continuous assessment Duration: 04:30

7	**Module 2: Lecture** Duration: 02:00			**Work on Homework 2** Continuous assessment Duration: 04:00 **Individual Study** Continuous assessment Duration: 02:00
8	**Module 2: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 2** Continuous assessment Duration: 04:00 **Exam on Module 1** Continuous assessment Duration: 02:00
9	**Module 2: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Upload Homework 1** Continuous assessment Duration: 00:00 **Work on Homework 2** Continuous assessment Duration: 04:00
10	**Module 2: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 2** Continuous assessment Duration: 03:30
11	**Module 3: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Exam on Module 2** Continuous assessment Duration: 02:00

12	**Module 3: Lecture** Duration: 02:00			**Individual Study** Continuous assessment Duration: 02:00 **Upload Homework 2** Continuous assessment Duration: 00:00
13	**Module 3: Lecture** Duration: 01:00	**Computer Lab. Module 3.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 3** Continuous assessment Duration: 04:30
14	**Module 3: Lecture** Duration: 01:00	**Computer Lab. Module 3.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 3** Continuous assessment Duration: 04:30
15	**Module 3: Lecture** Duration: 01:00	**Computer Lab. Module 3.** Duration: 01:00		**Individual Study** Continuous assessment Duration: 02:00 **Work on Homework 3** Continuous assessment Duration: 04:30
16	**There will be 6 more hours of teaching classes throughout the semester. These hours will be properly announced to the students.** Duration: 04:00			**Upload/Presentation Homework 3.** Continuous assessment Duration: 03:00
17				**Final Exam** Final examination Duration: 03:00 **Homework uploaded when requested (weeks 9, 12 and 16)** Final examination Duration: 00:00

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

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 subject schedule is based on a previous theorical planning of the subject plan and might go to through experience some unexpected changes along throughout the academic year.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
1	Individual Study		No Presential	02:00	%	0 / 10	
2	Individual Study		No Presential	02:00	%	0 / 10	
3	Individual Study		No Presential	02:00	%	0 / 10	
3	Work on Homework 1		No Presential	02:00	%	0 / 10	
4	Individual Study		No Presential	02:00	%	0 / 10	
4	Work on Homework 1		No Presential	02:30	%	0 / 10	
5	Individual Study		No Presential	02:00	%	0 / 10	
5	Work on Homework 1		No Presential	04:30	%	0 / 10	
6	Individual Study		No Presential	02:00	%	0 / 10	
6	Work on Homework 1		No Presential	04:30	%	0 / 10	
7	Individual Study		No Presential	02:00	%	0 / 10	
7	Work on Homework 2		No Presential	04:00	%	0 / 10	
8	Individual Study		No Presential	02:00	%	0 / 10	
8	Work on Homework 2		No Presential	04:00	%	0 / 10	
8	Exam on Module 1		Face-to-face	02:00	20%	3 / 10	CGI02 CECD04
9	Individual Study		No Presential	02:00	%	0 / 10	
9	Work on Homework 2		No Presential	04:00	%	0 / 10	
9	Upload Homework 1		No Presential	00:00	15%	3 / 10	CB10 CG14 CGI02 CB08 CECD04
10	Individual Study		No Presential	02:00	%	0 / 10	
10	Work on Homework 2		No Presential	03:30	%	0 / 10	
11	Individual Study		No Presential	02:00	%	0 / 10	
11	Exam on Module 2		Face-to-face	02:00	20%	3 / 10	CG08 CG07
12	Individual Study		No Presential	02:00	%	0 / 10	

12	Upload Homework 2		No Presential	00:00	15%	3 / 10	CG14 CB10 CG08 CG07
13	Individual Study		No Presential	02:00	%	0 / 10	
13	Work on Homework 3		No Presential	04:30	%	0 / 10	
14	Individual Study		No Presential	02:00	%	0 / 10	
14	Work on Homework 3		No Presential	04:30	%	0 / 10	
15	Individual Study		No Presential	02:00	%	0 / 10	
15	Work on Homework 3		No Presential	04:30	%	0 / 10	
16	Upload/Presentation Homework 3.		Face-to-face	03:00	30%	3 / 10	CB10 CG14 CECD05

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Final Exam		Face-to-face	03:00	40%	3 / 10	CG08 CG07 CGI02 CECD04
17	Homework uploaded when requested (weeks 9, 12 and 16)		No Presential	00:00	60%	3 / 10	CB10 CG08 CG14 CG07 CECD05 CGI02 CB08 CECD04

7.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

7.2. Assessment criteria

The final course grade will be based on the homework assignments grades and the exams grades broken down as follows:

Homework assignment 1 15%
Exam 1 20%
Homework assignment 2 15%
Exam 2 20%
Homework assignment 3 30%

The assignments will consist on applying the methods and techniques studied in class to different data sets. Each assignment and exam should obtain a grade equal or greater than 3 (0-10 scale) for the final grade to be computed. This final course grade must be equal or greater than 5 to pass the course. Eventually, students may be asked to present orally the conclusions of their work.

As far as the assignments are concerned it is mandatory to hand them in when requested. No late assignments will be accepted. If you miss the deadline, the grade for this assignment will be set to 0 and the overall grade couldn't be computed as it wouldn't meet the minimum requirements.

As far as the exams are concerned, the final exam is for those failing to take the midterms when requested or those obtaining a grade less than 3.

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). In this extraordinary exam you are allowed to hand in the three assignments and take the corresponding exam.

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.
Computer Lab	Equipment	Computer Room for hands-on sessions.
Johnson, R.A., Whichern, D.W. (2007) Applied Multivariate Statistical Analysis. Pearson Education.	Bibliography	Friendly exposition of the most important multivariate techiques, 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