



POLITÉCNICA

INTERNATIONAL  
CAMPUS OF  
EXCELLENCE

COORDINATION PROCESS OF  
LEARNING ACTIVITIES  
PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## LEARNING GUIDE

### SUBJECT

**103000896 - Statistical Data Analysis**

### DEGREE PROGRAMME

10BA - Master Universitario en Ciencia de Datos

### ACADEMIC YEAR & SEMESTER

2020/21 - 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 .....	2
5. Brief description of the subject and syllabus.....	3
6. Schedule.....	5
7. Activities and assessment criteria.....	7
8. Teaching resources.....	9

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000896 - Statistical Data Analysis
<b>No of credits</b>	4.5 ECTS
<b>Type</b>	Compulsory
<b>Academic year of the programme</b>	First year
<b>Semester of tuition</b>	Semester 1
<b>Tuition period</b>	September-January
<b>Tuition languages</b>	English
<b>Degree programme</b>	10BA - Master Universitario en Ciencia de Datos
<b>Centre</b>	10 - Escuela Tecnica Superior de Ingenieros Informaticos
<b>Academic year</b>	2020-21

## 2. Faculty

---

### 2.1. Faculty members with subject teaching role

<b>Name and surname</b>	<b>Office/Room</b>	<b>Email</b>	<b>Tutoring hours *</b>
Juan Antonio Fdez Del Pozo De Salamanca	2105	juan.fdezpozo.salamanca@u pm.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 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 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	<b>Module 1: Lecture</b> Duration: 02:00		<b>Module 1: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
2	<b>Module 1: Lecture</b> Duration: 01:00	<b>Computer Lab. Module 1.</b> Duration: 01:00	<b>Module 1: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
3	<b>Module 1: Lecture</b> Duration: 02:00		<b>Module 1: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
4	<b>Module 2: Lecture</b> Duration: 02:00		<b>Module 2: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
5	<b>Module 2: Lecture</b> Duration: 01:00	<b>Computer Lab. Module 2.</b> Duration: 01:00	<b>Module 2: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
6	<b>Module 2: Lecture</b> Duration: 01:00	<b>Computer Lab. Module 2.</b> Duration: 01:00	<b>Module 2: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
7	<b>Module 3: Lecture</b> Duration: 02:00		<b>Module 3: Lecture</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00
8	<b>Review</b> Duration: 02:00		<b>Review</b> Duration: 02:00	<b>Individual Study</b>  Continuous assessment Not Presential Duration: 02:00  <b>Exam on Module 1 and 2</b>  Continuous assessment Not Presential Duration: 02:00

9	Module 3: Lecture Duration: 02:00		Module 3: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
10	Module 3: Lecture Duration: 02:00		Module 3: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
11	Module 4: Lecture Duration: 02:00		Module 4: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
12	Review Duration: 02:00		Review Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00  <b>Exam on Module 3</b> Continuous assessment Not Presential Duration: 02:00
13	Module 4: Lecture Duration: 01:00	Computer Lab. Module 4. Duration: 01:00	Module 4: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
14	Module 4: Lecture Duration: 01:00	Computer Lab. Module 4. Duration: 01:00	Module 4: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
15	Module 4: Lecture Duration: 01:00	Computer Lab. Module 4 . Duration: 01:00	Module 4: Lecture Duration: 02:00	Individual Study Continuous assessment Not Presential Duration: 02:00
16				<b>Exam on Module 4</b> Continuous assessment Not Presential Duration: 02:00
17				<b>Final Exam</b> Final examination Not 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
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	
4	Individual Study		No Presential	02:00	%	0 / 10	
5	Individual Study		No Presential	02:00	%	0 / 10	
6	Individual Study		No Presential	02:00	%	0 / 10	
7	Individual Study		No Presential	02:00	%	0 / 10	
8	Individual Study		No Presential	02:00	%	0 / 10	
8	Exam on Module 1 and 2		No Presential	02:00	50%	3 / 10	CGI02 CB08 CG14 CECD04
9	Individual Study		No Presential	02:00	%	0 / 10	
10	Individual Study		No Presential	02:00	%	0 / 10	
11	Individual Study		No Presential	02:00	%	0 / 10	
12	Individual Study		No Presential	02:00	%	0 / 10	
12	Exam on Module 3		No Presential	02:00	25%	3 / 10	CB10 CECD05 CG07 CG08
13	Individual Study		No Presential	02:00	%	0 / 10	
14	Individual Study		No Presential	02:00	%	0 / 10	
15	Individual Study		No Presential	02:00	%	0 / 10	
16	Exam on Module 4		No Presential	02:00	25%	3 / 10	CG14 CB10 CECD05

#### 7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
------	-------------	----------	------	----------	--------	---------------	------------------

17	Final Exam		No Presential	03:00	100%	3 / 10	CGI02 CB08 CG14 CB10 CECD04 CECD05 CG07 CG08
----	------------	--	---------------	-------	------	--------	---

### 7.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

## 7.2. Assessment criteria

The overall course grade will be based on the performance on 4 exams, equally weighed, corresponding to each of the modules taught.

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.

As far as 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 any of them. In that case, only those specific exams with grades less than 3 have to be taken again.

No makeup exams will be scheduled.

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.

## 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	

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