ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000825 - Statistical Data Analysis

DEGREE PROGRAMME

10AX - Master Universitario Innovación Digital Ciencia de Datos Itinerario Health

ACADEMIC YEAR & SEMESTER

2020/21 - Semester 1

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

Index

Learning guide

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

1. Description

1.1. Subject details

Name of the subject	103000825 - Statistical Data Analysis
No of credits	4.5 ECTS
Type	Optional
Academic year ot the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10AX - Master Universitario Innovación Digital Ciencia de Datos Itinerario Health
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2020-21

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
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

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

* 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

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

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.

CE-HMDA02 - Capacidad para aplicar técnicas para la generación de visualizaciones adecuadas para el análisis y la exploración de datos en un contexto médico, y para la correcta comunicación de los resultados del análisis

CE-HMDA04 - Capacidad para aplicar métodos avanzados para clasificación, modelado, segmentación y predicción a partir de un conjunto de datos

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas

complejas de trabajo y estudio.

4.2. Learning outcomes

RA11 - To know and apply the main techniques to explore, describe and analyse multivariate data.

RA14 - To apply the proper methodology to analyse time series.

RA13 - To know and apply dimensionality reduction and modelization techniques to multivariate data.

* 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 four 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.

    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.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

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		**Module 1: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
2	**Module 1: Lecture** Duration: 01:00	**Computer Lab. Module 1.** Duration: 01:00	**Module 1: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
3	**Module 1: Lecture** Duration: 02:00		**Module 1: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
4	**Module 2: Lecture** Duration: 02:00		**Module 2: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
5	**Module 2: Lecture** Duration: 01:00	**Computer Lab. Module 2.** Duration: 01:00	**Module 2: Lecture** Duration: 02:00	**Work on Homework 1** Continuous assessment Not Presential Duration: 04:30 **Individual Study** Continuous assessment Not Presential Duration: 02:00
6	**Module 2: Lecture** Duration: 01:00	**Computer Lab. Module 2.** Duration: 01:00	**Module 2: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
7	**Module 3: Lecture** Duration: 02:00		**Module 3: Lecture** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00
8	**Review** Duration: 02:00		**Review** Duration: 02:00	**Individual Study** Continuous assessment Not Presential Duration: 02:00 **Exam on Module 1 and 2**

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

				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 **Exam on Module 3** Continuous assessment Not Presential Duration: 00:00
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	Module 4: Lecture Duration: 01:00	Module 4: Lecture Duration: 01:00	Individual Study Continuous assessment Not Presential Duration: 02:00
16				**Exam on Module 4** Continuous assessment Presential Duration: 02:00
17				**Final Exam** 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,

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

POLITÉCNICA | INTERNATIONAL CAMPUS OF EXCELLENCE

especially considering the COVID19 evolution.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

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	
5	Work on Homework 1		No Presential	04:30	%	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	CB07 CE-HMDA02 CG03
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	00:00	25%	3 / 10	CB07 CE-HMDA04 CB10 CG03
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		Face-to-face	02:00	25%	3 / 10	CB07 CE-HMDA04 CB10 CG03

7.1.2. Final examination

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Final Exam		No Presential	03:00	100%	3 / 10	CE-HMDA04 CB10 CG03 CB07 CE-HMDA02

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 grater or equal than 3, these grades can be kept and no repetition of these exams is required.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

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 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

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieros
Informaticos

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