



POLITÉCNICA

INTERNATIONAL  
CAMPUS OF  
EXCELLENCE

COORDINATION PROCESS OF  
LEARNING ACTIVITIES  
PR/CL/001



E.T.S. de Ingeniería  
Agronómica, Alimentaria y de  
Biosistemas

# ANX-PR/CL/001-01

## LEARNING GUIDE

### SUBJECT

**203000021 - Statistical Analysis And Data Visualization**

### DEGREE PROGRAMME

20BC - Master Universitario en Biología Computacional

### ACADEMIC YEAR & SEMESTER

2020/21 - Semester 1

## Index

---

### Learning guide

1. Description.....	1
2. Faculty.....	1
3. Skills and learning outcomes .....	2
4. Brief description of the subject and syllabus.....	3
5. Schedule.....	5
6. Activities and assessment criteria.....	7
7. Teaching resources.....	9

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	203000021 - Statistical Analysis And Data Visualization
<b>No of credits</b>	3 ECTS
<b>Type</b>	Optional
<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>	20BC - Master Universitario en Biología Computacional
<b>Centre</b>	20 - E.T.S. de Ingeniería Agronómica, Alimentaria y de Biosistemas
<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>
Raul Garcia Castro	2110	r.garcia@upm.es	Sin horario.
Antonio Jimenez Martin	2110	antonio.jimenez@upm.es	Sin horario.
Juan Antonio Fdez Del Pozo De Salamanca	2101	juan.fdezpozo.salamanca@u pm.es	Sin horario.
Jacinto Gonzalez Pachon (Subject coordinator)	2101	jacinto.gonzalez.pachon@up m.es	Sin horario.

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

## 3. Skills and learning outcomes \*

---

### 3.1. Skills to be learned

CE01 - Comprender las bases moleculares y las técnicas experimentales estándares más comunes en las investigaciones ómicas (genómica, transcriptómica, proteómica, metabolómica, interactómica, etc.).

CE04 - Utilizar diferentes bases de datos (incluidos los bigdata), conocer sus estructuras y ontologías, aplicar la estadística a su análisis, siendo capaz de utilizar herramientas de representación y visualización.

CE08 - Capacidad de integrar tecnologías y sistemas propios de la Inteligencia Artificial, con carácter generalista, y en contextos más amplios y multidisciplinares.

CE09 - Capacidad de interpretar los modelos de clasificación supervisada y no supervisada obtenidos al aplicar las técnicas de Aprendizaje Automático para un conjunto de datos.

CG02 - Familiarizarse con el trabajo y los métodos de la Biología Computacional en condiciones reales, adquiriendo la capacidad de diseñar aplicaciones/experimentos de forma independiente y describir, cuantificar, analizar y evaluar críticamente los resultados obtenidos.

CG05 - Que los estudiantes sean capaces de integrar conocimientos en el área de la Biología Computacional, de formular conclusiones, hipótesis o líneas de trabajo a partir de la información disponible, y de formarse una opinión fundamentada sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos.

CT02 - Capacidad para aplicar el método científico para la resolución de problemas de forma efectiva y creativa.

CT06 - Capacidad para liderar y trabajar en equipos multidisciplinares y multiculturales en un contexto internacional.

CT07 - Ser capaz de manejar las tecnologías de la información y comunicación en un contexto profesional.

## 3.2. Learning outcomes

RA2 - Los estudiantes deben ser capaces de identificar metodologías apropiadas e inapropiadas para una situación determinada

RA1 - Los estudiantes deben entender conceptos específicos y procedimientos estadísticos

RA3 - Los estudiantes deben ser capaces de aplicar métodos estadísticos específicos adecuados

\* 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.

## 4. Brief description of the subject and syllabus

---

### 4.1. Brief description of the subject

The course is a non-exhaustive review of techniques to convert a set of data into relevant information. Once relevant information is obtained, the next objective is to obtain the necessary knowledge to be able to make good decisions. The perspective of the course is double, both theoretical and applied. The course covers topics such as exploratory data analysis, summary statistics, data visualization, as well as models for decision-making. There will be an emphasis on the practical application of theory and methods, with extensive use of R.

### 4.2. Syllabus

1. Data Exploratory Analysis
  - 1.1. Univariate analysis
  - 1.2. Multivariate analysis
2. An introduction to Decision Theory
  - 2.1. Decision under certainty vs. decision under uncertainty
  - 2.2. Multi-criteria and multi-objective decision-making
  - 2.3. Collective decision-making
3. Data visualization
  - 3.1. Basic data visualization
  - 3.2. Models visualization
  - 3.3. Processes visualization

#### 4. Models selection

4.1. Principal component analysis

4.2. Factorial analysis

4.3. Correspondence analysis

4.4. Multidimensional scaling

## 5. Schedule

### 5.1. Subject schedule\*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1			<b>Lesson 1</b> Duration: 02:00	
2			<b>Lesson 1</b> Duration: 02:00	
3			<b>Lesson 2</b> Duration: 02:00	
4			<b>Lesson 2</b> Duration: 02:00	
5			<b>Lesson 3</b> Duration: 02:00	<b>Lecture notes about the theoretical content of lesson 2</b>  Continuous assessment Not Presential Duration: 00:00
6			<b>Lesson 3</b> Duration: 02:00	
7			<b>Lesson 3</b> Duration: 02:00	
8			<b>Lesson 3</b> Duration: 02:00	
9			<b>Lesson 3</b> Duration: 02:00	
10			<b>Lesson 3</b> Duration: 02:00	<b>Practical report about the content of lessons 1 and 3 on a specific data set</b>  Continuous assessment Not Presential Duration: 00:00
11			<b>Lesson 4</b> Duration: 02:00	
12			<b>Lesson 4</b> Duration: 02:00	<b>Practical report about the content of lesson 4 on a specific data set</b>  Continuous assessment Not Presential Duration: 00:00

13				
14				Practical report about the content of lessons 1, 3 and 4 on a specific data set  Final examination Not Presential Duration: 00:00
15				
16				
17				

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.



## 6. Activities and assessment criteria

### 6.1. Assessment activities

#### 6.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
5	Lecture notes about the theoretical content of lesson 2		No Presential	00:00	20%	5 / 10	CE08 CT02 CG05
10	Practical report about the content of lessons 1 and 3 on a specific data set		No Presential	00:00	40%	5 / 10	CT06 CE08 CE04 CT02 CE01 CE09 CG05 CG02 CT07
12	Practical report about the content of lesson 4 on a specific data set		No Presential	00:00	30%	5 / 10	CE04 CT02 CT06 CE09 CG05 CT07

#### 6.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
14	Practical report about the content of lessons 1, 3 and 4 on a specific data set		No Presential	00:00	100%	5 / 10	CE08 CE04 CT02 CE01 CE09 CG05 CG02 CT07

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

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Practical report about the content of lessons 1, 3 and 4 on a specific data set		Face-to-face	00:00	100%	5 / 10	CE04 CT02 CE08 CE01 CE09 CG05 CG02 CT07

## 6.2. Assessment criteria

Regular attendance to class is mandatory to apply the continuous evaluation. Attendance represents a 10% in the final score.

Both practical reports and lecture notes of the theoretical content will be evaluated according to their completeness, clarity, coherence and originality, as well as the ability to interconnect ideas.

The final grade of the course will be obtained from the results obtained in the activities carried out according to the percentage detailed in the table of Summative Evaluation.

### Extraordinary examination

The score in extraordinary examination will be obtained by the same concepts detailed in the table of Summative Evaluation, except the participation in class. The 10% corresponding to this activity is transferred to the practical reports, weighted with 45% and 35% respectively.

## 7. Teaching resources

### 7.1. Teaching resources for the subject

Name	Type	Notes
Chambers, J.M. (2008) Software for data analysis, Programming with R ISBN: 978-0-387-75935-7, Springer CA USA	Bibliography	Libro
French, S. (1989) Decision Theory. Ellis Horwood, Cichester	Bibliography	Libro
Hwang, C-H.; Lin, M-J. (1987). Group decision making under multiple criteria. Springer-Verlag, Berlín	Bibliography	Libro
Winton Chang (2013) R Graphics Cookbook ISBN:978-1-449-31695-2, O'Reilly Media, Inc. CA USA	Bibliography	Libro
Chun-houh Chen, Wolfgang Härdle, Antony Unwin (Editors) (2008) Handbook of Data Visualization ISBN 978-3-540-33036-3, Springer Springer-Verlag Berlin Heidelberg	Bibliography	Libro
Gentleman, R.C. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.	Bibliography	Artículos e Informes
Hadley Wickham (2016) ggplot2. Elegant Graphics for Data Analysis. Springer-Verlag, New York, 2016. ISBN 978-3-319-24275-0. 260 <a href="http://www.springer.com/gp/book/9783319242750">http://www.springer.com/gp/book/9783319242750</a>	Bibliography	Artículos e Informes

Hadley Wickham et al., R packages for data science, <a href="https://www.tidyverse.org/">https://www.tidyverse.org/</a>	Bibliography	Artículos e Informes
R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <a href="https://www.R-project.org/">https://www.R-project.org/</a> .	Bibliography	Artículos e Informes
Virtual classroom	Web resource	
Library	Others	
Assigned classroom	Others	
Moodle	Web resource	