



Universidad  
de Alcalá

# GUÍA DOCENTE

## PROGRAMACIÓN Y ANÁLISIS DE DATOS CON R

**Máster en Ciencias Actuariales y  
Financieras**

**Universidad de Alcalá**

**Curso Académico 2018/2019**

**Segundo Curso – Segundo Cuatrimestre**

## GUÍA DOCENTE

Nombre de la asignatura:	<b>Programación y Análisis de Datos con R</b>
Código:	<b>202052</b>
Titulación en la que se imparte:	<b>Máster en Ciencias Actuariales y Financieras</b>
Departamento y Área de Conocimiento:	<b>Economía / Métodos Cuantitativos para la Economía y la Empresa</b>
Carácter:	<b>Obligatorio</b>
Créditos ECTS:	<b>6 créditos</b>
Curso y cuatrimestre:	<b>Segundo Curso – Segundo Cuatrimestre</b>
Profesorado:	Fco. Javier Callealta Barroso Pablo Alonso González
Horario de Tutoría:	<b>Se comunicará al comienzo del curso</b>
Idioma en el que se imparte:	Español

### 1a. PRESENTACIÓN

Superados los niveles de conocimiento teóricos fundamentales y aprendidas las habilidades básicas que le permitan observar, analizar y modelizar objetivamente la realidad en situaciones de riesgo, esta asignatura busca el estudio de la interrelación entre variables que definen un fenómeno a estudiar. Para ello, se requiere de la utilización de herramientas, tanto teóricas como de ejecución, que permitan llevar a cabo el estudio en cuestión. Las herramientas teóricas a utilizar son las conocidas como técnicas multivariantes, las cuales, dada su complejidad inherente, requieren de la utilización de sistemas de cálculo y programación. Este tipo de técnicas de análisis son de amplio uso en entorno caracterizados por la utilización de base de datos de gran tamaño y con un elevado número de variables.

Por ello, la selección de contenidos se adapta a la confección de dos bloques temáticos, cuyos objetivos básicos se expresan a continuación:

Introducción al programa R. En este bloque se pretende que el alumno se familiarice con el entorno característico de este programa, que es el estándar en el ámbito científico. Se espera que logre cierta soltura en el manejo de datos y en el cálculo básico de estadísticas. En la última parte de este bloque se introducen los conceptos básicos de programación. Se pretende que el alumno domine, ya no sólo la utilización básica de los procedimientos pre-programados en las librerías del programa estadístico R, sino que aprenda a realizar sus propios programas de cálculo, preparación y análisis de datos y simulación

Análisis Multivariante de Datos. En este bloque, el alumno deberá familiarizarse con los principales conceptos relacionados con un conjunto básico de técnicas de Análisis Multivariante de Datos desde una perspectiva teórico-práctica, aprendiendo a detectar las similitudes, diferencias y aplicabilidad de los diversos métodos y conceptos revisados, y llegando a utilizar dichas técnicas desde una perspectiva aplicada en el entorno del programa estadístico R.

Dado el carácter eminentemente práctico de la asignatura y la necesidad de realizar cálculos informáticos, se sugiere que el alumno siga las sesiones con un ordenador portátil al que se le deberá cargar el programa R.

## 1b. PRESENTATION

Once students have surpassed the levels of fundamental theoretical knowledge and have learned the basic skills that allow to observe, analyze and model the reality objectively at risk, this course seeks to study the interrelationship between variables that define this phenomenon. Doing this requires the use of tools, both theoretical and practical, allowing to carry out the study in question.

On the one hand, theoretical tools will be multivariate techniques, which given its inherent complexity, require the use of computing systems and programming. On the other hand, this type of analysis techniques is widely used in environments characterized by the use of database large with a large number of variables. Therefore, the contents in this course has been split in two thematic blocks, whose basic objectives are given below:

Introduction to "R". In this block is intended to familiarize the student with the characteristic environment of this program, which is an standard in Sciences. It is expected to achieve some ease in handling data and the calculation of basic statistics. In the latter part of this block, basic programming concepts are introduced. It is intended that the student have a good command, not only of the basic use of pre-programmed statistical programs in bookstores "R", but of making their own calculation programs, as well as data analysis programs and simulation procedures.

Multivariate Data Analysis. In this block, students will become familiar with concepts related to main Multivariate Data Analysis techniques, from both theoretical and practical perspectives, and will learn to consider its differences and properties in order to use them properly from an applied perspective with the "R" statistical program.

Given the practical nature of the subject and the need for computer calculations, it is suggested that students follow the sessions with a laptop computer with the "R" program installed.

## 2. COMPETENCIAS

Las competencias que deben acrecentar los estudiantes en esta asignatura son, entre otras, las siguientes:

### Competencias genéricas:

1. Capacidad de observación y síntesis: de aspectos de la realidad
2. Capacidad de abstracción: de la realidad a modelos teóricos.
3. Capacidad de análisis: de problemas prácticos reales.
4. Capacidad crítica: frente a nuevos conocimientos
5. Capacidad de trabajo: individualmente y en equipo.

### Competencias específicas:

1. Habilidad para aplicar correctamente técnicas de programación y simulación con R, en el ámbito del seguro y las finanzas.
2. Capacidad de reconocer los principales conceptos relacionados con el análisis estadístico multivariante de datos desde una perspectiva teórica (similitudes y diferencias entre los diversos conceptos y métodos) y práctica (utilización básica de programas de ordenador e interpretación de resultados).

## 3. CONTENIDOS

**Observación:** la planificación temporal que se propone en este apartado es sólo aproximada. Por tanto, es susceptible de sufrir variaciones a lo largo del curso, en función del desarrollo de la asignatura.

Bloques de contenido (se pueden especificar los temas si se considera necesario)	Total de clases, créditos u horas
<b>Parte I: Introducción al R</b>	
Tema 1: El entorno del programa estadístico “R” Descarga e Instalación. Tipos de datos en R. Importación y exportación de datos. Recodificación de variables. La interfaz R-Commander. Operadores lógicos.	▪ 1 semana
Tema 2: Cálculo matricial en R Operaciones con vectores y matrices. Los array	▪ 1 semana
Tema 3: Estadística descriptiva y creación de gráficos. Análisis unidimensional: creación de gráficos y cálculo de estadísticos. Tablas y gráficos en los casos de variables uni, bi o multidimensionales	▪ 2 semanas

<p><b>Tema 4: Modelo de regresión lineal</b> Regresiones con una y varias variables explicativas. Generación de residuos y predicciones.</p>	<ul style="list-style-type: none"> <li>▪ 1 semana</li> </ul>
<p><b>Tema 5: Fundamentos de programación en R</b> Los “Scripts”. Bucles: for, while, repeat. Instrucciones if e ifelse. Generación de funciones por el usuario</p>	<ul style="list-style-type: none"> <li>▪ 1 semana</li> </ul>
<p><b>Tema 6: Bootstrapping y Montecarlo</b> Generación de muestras por Bootstrapping. Aplicaciones estadísticas. Regresión por Bootstrap. Montecarlo simple y estructurado.</p>	<ul style="list-style-type: none"> <li>▪ 1 semanas</li> </ul>
<p><b>Parte II: Análisis multivariante de datos</b></p>	
<p><b>Tema 7: Introducción al Análisis de Datos con R y Análisis Preliminar de Datos</b> Introducción a R-Commander. Análisis Exploratorio de Datos. Datos atípicos. Comportamientos asimétricos: detección, efectos y tratamiento. El problema de la no respuesta: métodos de imputación.</p>	<ul style="list-style-type: none"> <li>▪ 1 semana</li> </ul>
<p><b>Tema 8: Análisis de Componentes Principales</b> Planteamiento del problema. Resolución. Propiedades. Criterios de selección de componentes.</p>	<ul style="list-style-type: none"> <li>▪ 1 semana</li> </ul>
<p><b>Tema 9: Análisis Factorial</b> Planteamiento del problema. Propiedades generales. Resolución y Métodos de resolución. Puntuaciones factoriales. Rotaciones. Análisis de Adecuación.</p>	<ul style="list-style-type: none"> <li>▪ 2 semanas</li> </ul>
<p><b>Tema 10: Análisis de la Proximidad</b> Introducción al análisis de la proximidad. Medidas de disimilaridad y de similaridad para casos y para variables.</p>	<ul style="list-style-type: none"> <li>▪ 1 semanas</li> </ul>
<p><b>Tema 11: Análisis de Conglomerados</b> Introducción al análisis de conglomerados. Métodos jerárquicos y no jerárquicos de conglomeración.</p>	<ul style="list-style-type: none"> <li>▪ 2 semanas</li> </ul>

## 4. METODOLOGÍAS DE ENSEÑANZA-APRENDIZAJE.-ACTIVIDADES FORMATIVAS

### 4.1. Distribución de créditos (especificar en horas)

Número de horas presenciales: 48	<ul style="list-style-type: none"> <li>Clases magistrales, Seminarios teórico-prácticos y clases prácticas en aula con ordenador (45 horas, en sesiones de hora y media)</li> <li>Examen final (3 horas)</li> </ul>
Número de horas del trabajo propio del estudiante: 102	<ul style="list-style-type: none"> <li>Tutorías ECTS: 3 horas.</li> <li>Trabajo autónomo: 99 horas.</li> </ul>
Total horas: 150	

### 4.2. Estrategias metodológicas, materiales y recursos didácticos

Clases presenciales en aula tradicional	<ul style="list-style-type: none"> <li>Clases teóricas con ayuda de pizarra y de medios audiovisuales.</li> <li>Seminarios con presentaciones y debates sobre análisis de casos prácticos.</li> </ul>
Clases presenciales en aula con ordenador	<ul style="list-style-type: none"> <li>Clases de carácter práctico-aplicado con uso intensivo de software estadístico y medios informáticos.</li> </ul>
Tutorías individualizadas	<ul style="list-style-type: none"> <li>Atención individualizada a los estudiantes, para asistirles en su proceso de aprendizaje</li> </ul>

## 5. EVALUACIÓN: Procedimientos, criterios de evaluación y de calificación<sup>1</sup>

### Criterios de evaluación.

Para la evaluación de las competencias de la asignatura adquiridas por los estudiantes se proponen dos alternativas, aunque la opción preferente será la primera de ellas:

#### a) Evaluación continua

La evaluación continua supone evaluar la adquisición de las competencias por parte de los alumnos de una forma continuada a lo largo del curso, valorando capacidades, actitudes y destrezas, entre otras cosas.

Para la evaluación continua de las competencias adquiridas por los estudiantes de la asignatura el alumno debe asistir regularmente a clase, participando de forma activa, resolviendo los casos prácticos que se propongan de forma individual o colectiva y realizando todas las pruebas presenciales de evaluación que se propongan a lo largo del curso (al menos el 90% de ellas).

#### b) Examen final.

El alumno podrá optar ser evaluado únicamente por la realización de un examen final de la asignatura, siempre que se solicite formalmente por escrito al principio del curso, atendiendo a lo dispuesto en el artículo 10-3 de la Normativa Reguladora de los Procesos de Evaluación y Aprendizaje.

### Criterios de calificación.

#### a) Evaluación continua

Los criterios de calificación que se emplearán, así como su peso en la calificación final, serán los siguientes:

- Se calculará una calificación correspondiente a los contenidos de la **Parte I de la asignatura** a partir de Pruebas presenciales y no presenciales propuestas a lo largo del curso para la evaluación del dominio de conocimientos prácticos (100%).
- Se calculará una calificación correspondiente a los contenidos de la **Parte II de la asignatura** a partir de Pruebas presenciales y no presenciales propuestas a lo largo del curso para la evaluación del dominio de conocimientos conceptuales teóricos (30%) y prácticos (70%).
- La **calificación final en la asignatura** se obtendrá como la media simple de las calificaciones obtenidas en cada una de las dos partes citadas de la asignatura.

---

<sup>1</sup> Siguiendo la **Normativa reguladora de los procesos de evaluación de los aprendizajes, aprobada en Consejo de Gobierno de 24 de Marzo de 2011**, es importante señalar los procedimientos de evaluación: por ejemplo evaluación continua, final, autoevaluación, co-evaluación. Instrumentos y evidencias: trabajos, actividades. Criterios o indicadores que se van a valorar en relación a las competencias: dominio de conocimientos conceptuales, aplicación, transferencia conocimientos. Para el sistema de calificación hay que recordar la **Normativa del Consejo de Gobierno del 16 de Julio de 2009**.

## b) Examen final.

Aquellos alumnos que opten por una única evaluación final tendrán que realizar un examen final consistente en una prueba escrita teórico-práctica con la siguiente estructura donde figuran las calificaciones de cada parte:

### Parte I:

**Ejercicio práctico (100%):** consistente en la realización de un programa R en el que se le solicitarán diferentes tareas.

### Parte II:

**Ejercicio Teórico (30%):** No se permite ningún tipo de material. Constaría de una batería de preguntas de tipo test y/o cuestiones de índole teórica.

**Ejercicio Práctico (70%):** Se compondrá uno o varios supuestos prácticos, de naturaleza similar a los desarrollados en clase, sobre los que se deberán realizar ciertos análisis estadísticos basándose en el software utilizado durante las clases.

La calificación final en la asignatura se obtendrá como la media simple de las calificaciones obtenidas en cada una de las dos partes citadas anteriormente.

Según el artículo 10-5 de la Normativa Reguladora de los procesos de evaluación y aprendizaje, los alumnos que suspendan la evaluación continua no podrán acogerse a un examen final de convocatoria ordinaria.

Nota: Los alumnos que no superen la asignatura en el periodo ordinario, podrán presentarse a la convocatoria extraordinaria, que consistirá en una prueba escrita teórico-práctica con la estructura descrita anteriormente

## 6. BIBLIOGRAFÍA

### I. Introducción al R

[CRA] Crawley, M.: *The R Book*. 2nd edition. John Wiley & Sons. 2013

[VEN] Venables, W.N., Smith, D.M. and the R Core Team: *An Introduction to R*. Version 3.1.0. 2014

[EFR] Efron B. and Tibshiran, R.J.: *An introduction to the Bootstrap*, Chapman and Hall. 1993

[JAC] Jackel, P.: *Monte Carlo Methods in Finance*. Wiley Finance. 2002

### II. Análisis Multivariante de datos

#### BIBLIOGRAFÍA BÁSICA:

[M-U] Manzano, J.; Uriel Jimenez, E.: *Análisis multivariante aplicado con R*. 2ª ed. Paranifo, 2017.

[E-H] Everitt, B.; Hothorn, T.: *An Introduction to Applied Multivariate Analysis with R*. Springer. 2011.

Además, se recurrirá a los manuales y ayudas de referencia proporcionados por los programas informáticos que se emplearán en la docencia de esta materia.



**BIBLIOGRAFÍA COMPLEMENTARIA:**

- [CUA] Cuadras, C. M.: *Métodos de Análisis Multivariante*. Ed. EUB, 1996.
- [EVc] Everitt, B. S.: *Cluster Analysis*. 3ª Ed. John Wiley & Sons, New York, 1993.
- [HTB] Hair, J.F. Tatham, R.L. Black, W.C. *Análisis Multivariante*. Prentice Hall, 1999.
- [JyW] Johnson, R. A.; Wichern, D. W.: *Applied Multivariate Statistical Analysis*. Ed. Prentice-Hall, 1992.
- [L-V] Levy, J.; Varela, J.: *Análisis Multivariable para las Ciencias Sociales*. Pearson Prentice-Hall. 2003.
- [MKB] Mardia, K.V. Kent, J.T. Bibby, J.M. *Multivariate Analysis*. Academic Press, 1979.
- [PEÑ] Peña, D.: *Análisis de Datos Multivariantes*. McGraw Hill. 2002.
- [SHA] Sharma, S *Applied Multivariate Techniques*. John Wiley & Sons, 1996.