

**Guía docente de la asignatura**

<b>Asignatura</b>	ANÁLISIS DE DATOS MULTIVARIANTES		
<b>Materia</b>	SISTEMAS INTELIGENTES Y BASADOS EN CONOCIMIENTO		
<b>Módulo</b>	TECNOLOGÍAS INFORMÁTICAS		
<b>Titulación</b>	MÁSTER EN INGENIERÍA INFORMÁTICA (510)		
<b>Plan</b>	510	<b>Código</b>	53179
<b>Periodo de impartición</b>	2º CUATRIMESTRE	<b>Tipo/Carácter</b>	OPTATIVA
<b>Nivel/Ciclo</b>	MÁSTER	<b>Curso</b>	1º
<b>Créditos ECTS</b>	3 ECTS		
<b>Lengua en que se imparte</b>	CASTELLANO		
<b>Profesor/es responsable/s</b>	LUIS AUGUSTO SAN JOSÉ NIETO		
<b>Datos de contacto (E-mail, teléfono...)</b>	TELÉFONO: 983 423000 ext. 5707 E-MAIL: <a href="mailto:augusto@mat.uva.es">augusto@mat.uva.es</a>		
<b>Horario de tutorías</b>	Véase <a href="http://www.uva.es">www.uva.es</a> → Centros → Campus de Valladolid → Escuela Técnica Superior de Ingeniería Informática → Tutorías		
<b>Departamento</b>	MATEMÁTICA APLICADA		



## 1. Situación / Sentido de la Asignatura

### 1.1 Contextualización

El objetivo general de la asignatura, que tiene carácter instrumental, es dar a conocer al alumno los métodos y las técnicas más habituales que se utilizan, dentro del ámbito de la ingeniería informática, en el análisis multivariante de datos. Más concretamente, en esta asignatura se proporcionarán las herramientas básicas que se usan en el análisis de datos con el objeto de reducir el número de variables disponibles, estudiar las posibles relaciones de dependencia entre las variables, clasificar y segmentar los datos observados, y todo ello tanto si el carácter de los datos disponibles es cuantitativo como si es cualitativo. Las técnicas del análisis de datos multivariante que se estudiarán tienen aplicaciones en diferentes sectores (industria, economía, sociología, etc.) y han alcanzado una gran utilización en ingeniería informática. Así estas técnicas se utilizan como herramientas para resumir la información y diseñar sistemas de clasificación automática y de reconocimiento de patrones.

### 1.2 Relación con otras materias

Las técnicas que se estudian en Análisis de Datos Multivariantes se utilizan en diferentes asignaturas de la materia Sistemas Inteligentes y Basados en Conocimiento.

### 1.3 Prerrequisitos

Para cursar esta asignatura se recomienda que el alumno posea unos conocimientos básicos de Estadística y de cálculo matricial.



## 2. Competencias

### 2.1 Generales

Código	Descripción
CG4	Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática.

### 2.2 Específicas

Código	Descripción
CET9	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

## 3. Objetivos

Código	Descripción
CET9.1	Conocer y saber aplicar los conceptos básicos y resultados.
CET9.2	Adquirir aptitudes para aplicar las técnicas multivariantes más adecuadas en el planteamiento y resolución de problemas propios de la ingeniería informática.
CET9.3	Utilizar software estadístico para resolver problemas de análisis multivariante.
CET9.4	Interpretar los resultados obtenidos con las distintas técnicas de análisis multivariante.
CET9.5	Conocer la presencia de esta materia en otras disciplinas de la titulación.



#### 4. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	14	Estudio y trabajo autónomo individual	30
Laboratorios (L)	8	Estudio y trabajo autónomo grupal	15
Seminarios (S)	8		
<b>Total presencial</b>	<b>30</b>	<b>Total no presencial</b>	<b>45</b>





## 5. Bloques temáticos

### Bloque 1: PRELIMINARES

Carga de trabajo en créditos ECTS: 

#### a. Contextualización y justificación

El primer paso de cualquier análisis estadístico es la exploración inicial de los datos a través de sencillas técnicas descriptivas. En este primer tema se presentan algunas de estas herramientas y, además, se introduce la distribución normal multivariante (generalización para vectores de la variable del modelo normal). Esta distribución de probabilidad multivariante desempeña un papel importante, ya que cuando se trabaja en problemas prácticos, la hipótesis más habitual es que las observaciones proceden de una distribución normal. Además, algunos de los procedimientos que se desarrollarán posteriormente necesitan la hipótesis de normalidad multivariante.

#### b. Objetivos de aprendizaje

Código	Descripción
CET9.1	Conocer y saber aplicar los conceptos básicos y resultados.
CET9.3	Utilizar software estadístico para resolver problemas de análisis multivariante.

#### c. Contenidos

##### TEMA 1: Introducción a las técnicas de análisis estadístico multivariante

- 1.1 El análisis de datos multivariantes. Repaso de conceptos de cálculo matricial
- 1.2 Descripción de datos multivariantes
- 1.3 La distribución normal multivariante

#### d. Métodos docentes

- Clases magistrales participativas y expositivas
- Aprendizaje basado en problemas
- Resolución de problemas y casos prácticos con apoyo informático

#### e. Plan de trabajo

Ver Anexo

#### f. Evaluación

Ver punto 7 de esta guía

#### g. Bibliografía básica

- C. M. Cuadras, Nuevos Métodos de Análisis Multivariante. Barcelona: CMC Editions, 2012. Capítulos 1 y 2.
- D. Peña, Análisis de datos multivariantes. Madrid: McGraw-Hill, 2002. Capítulos 1-4, y 9. ISBN 978-84-481-3610-9.



### h. Bibliografía complementaria

- W. Härdle and L. Simar, Applied multivariate statistical analysis, second edition. Berlín: Springer, 2007. Capítulos 1-4. ISBN 978-3-540-72243-4.
- R.A. Jonson and D.W. Wichern, Applied multivariate statistical analysis, sixth edition. New Jersey: Pearson Prentice Hall, 2007. Capítulos 1-4. ISBN 978-0-13-187715-3.

### i. Recursos necesarios

El profesor de la asignatura hará accesible a los alumnos el conjunto de materiales y recursos de apoyo que considere adecuados utilizar en la preparación de la asignatura, tanto para los aspectos teóricos como los computacionales, a través de la página web de la asignatura o del servicio de reprografía del centro.

## Bloque 2: MÉTODOS DE EXPLORACIÓN DE DATOS

Carga de trabajo en créditos ECTS:

### a. Contextualización y justificación

En este bloque de la asignatura se presentan los métodos multivariantes de interdependencia. Estos métodos (descriptivos) se utilizan para estudiar las interrelaciones entre variables y su estructura subyacente cuando entre las variables no existe ningún tipo de distinción (es decir, todas ellas tienen un papel similar). Se abordan técnicas de reducción de la dimensión (análisis de componentes principales, escalado multidimensional y análisis de correspondencias) y de agrupación en grupos homogéneos (análisis de conglomerados).

### b. Objetivos de aprendizaje

Código	Descripción
CET9.1	Conocer y saber aplicar los conceptos básicos y resultados.
CET9.2	Adquirir aptitudes para aplicar las técnicas multivariantes más adecuadas en el planteamiento y resolución de problemas propios de la ingeniería informática.
CET9.3	Utilizar software estadístico para resolver problemas de análisis multivariante.
CET9.4	Interpretar los resultados obtenidos con las distintas técnicas de análisis multivariante.
CET9.5	Conocer la presencia de esta materia en otras disciplinas de la titulación.

### c. Contenidos

#### TEMA 2: Análisis de componentes principales

- 2.1 Introducción
- 2.2 Cálculo de las componentes principales. Propiedades
- 2.3 Interpretación. Número de componentes

#### TEMA 3: Escalado multidimensional

- 3.1 Introducción
- 3.2 Escalado métrico
- 3.3 Escalado no métrico





**TEMA 4: Análisis de correspondencias**

- 4.1 Introducción
- 4.2 Representación
- 4.3 Interpretación

**TEMA 5 Análisis de conglomerados (clusters)**

- 5.1 Introducción
- 5.2 Medidas de similitud
- 5.3 Algoritmos de clasificación

---

**d. Métodos docentes**

- Clases magistrales participativas y expositivas
- Aprendizaje basado en problemas
- Resolución de problemas y casos prácticos con apoyo informático

---

**e. Plan de trabajo**

Ver Anexo

---

**f. Evaluación**

Ver punto 7 de esta guía

---

**g. Bibliografía básica**

- C. M. Cuadras, Nuevos Métodos de Análisis Multivariante. Barcelona: CMC Editions, 2012. Capítulos 5, 8-10.
- D. Peña, Análisis de datos multivariantes. Madrid: McGraw-Hill, 2002. Capítulos 5-8. ISBN 978-84-481-3610-9.

---

**h. Bibliografía complementaria**

- W. Härdle and L. Simar, Applied multivariate statistical analysis, second edition. Berlín: Springer, 2007. Capítulos 9, 11, 13 y 15. ISBN 978-3-540-72243-4.
- A.C. Rencher, Methods of multivariate analysis, second edition. New York: John Wiley, 2002. Capítulos 12, 14-15. ISBN 0-471-41889-7.

---

**i. Recursos necesarios**

El profesor de la asignatura hará accesible a los alumnos el conjunto de materiales y recursos de apoyo que considere adecuados utilizar en la preparación de la asignatura, tanto para los aspectos teóricos como los computacionales, a través de la página web de la asignatura o del servicio de reprografía del centro.

**Bloque 3: TÉCNICAS DE ANÁLISIS DE DEPENDENCIA**Carga de trabajo en créditos ECTS: **a. Contextualización y justificación**

La última parte de la asignatura está dedicada a exponer algunas de las técnicas que se utilizan cuando las variables analizadas están divididas en dos grupos: variables dependientes y variables independientes. El objetivo de estas técnicas, que se denominan métodos explicativos o de dependencia, es determinar si el conjunto de variables dependientes puede ser explicado a través de la información suministrada por el conjunto de variables independientes. El primer problema que abordaremos es el de la clasificación de un nuevo elemento (observación) de la que se desconoce su grupo (población) de procedencia mediante el análisis del valor que en ella toman las variables independientes. Este tipo de problemas aparece en muchas situaciones en que se necesita clasificar elementos con información incompleta. Finalmente, analizaremos el problema de encontrar indicadores (combinaciones lineales) de un conjunto de variables independientes que explique lo máximo posible a los indicadores de las variables del grupo de variables dependientes. Es decir, se trata de obtener un conjunto de ponderaciones tanto para las variables dependientes como para las independientes que maximice la correlación entre el grupo de variables dependientes y el grupo de variables independientes.

**b. Objetivos de aprendizaje**

Código	Descripción
CET9.1	Conocer y saber aplicar los conceptos básicos y resultados.
CET9.2	Adquirir aptitudes para aplicar las técnicas multivariantes más adecuadas en el planteamiento y resolución de problemas propios de la ingeniería informática.
CET9.3	Utilizar software estadístico para resolver problemas de análisis multivariante.
CET9.4	Interpretar los resultados obtenidos con las distintas técnicas de análisis multivariante.
CET9.5	Conocer la presencia de esta materia en otras disciplinas de la titulación.

**c. Contenidos****TEMA 6: Análisis discriminante**

- 6.1 Introducción
- 6.2 Clasificación en dos poblaciones
- 6.3 Clasificación en  $k$  poblaciones

**TEMA 7: Análisis de correlación canónica**

- 7.1 Introducción
- 7.2 Variables canónicas
- 7.3 Interpretación

**d. Métodos docentes**

- Clases magistrales participativas y expositivas
- Aprendizaje basado en problemas
- Resolución de problemas y casos prácticos con apoyo informático





### **e. Plan de trabajo**

---

Ver Anexo

### **f. Evaluación**

---

Ver punto 7 de esta guía

### **g. Bibliografía básica**

---

- C. M. Cuadras, Nuevos Métodos de Análisis Multivariante. Barcelona: CMC Editions, 2012. Capítulos 4 y 11.
- D. Peña, Análisis de datos multivariantes. Madrid: McGraw-Hill, 2002. Capítulos 13 y 16. ISBN 978-84-481-3610-9.

### **h. Bibliografía complementaria**

---

- W. Härdle and L. Simar, Applied multivariate statistical analysis, second edition. Berlín: Springer, 2007. Capítulos 12 y 14. ISBN 978-3-540-72243-4.
- R.A. Jonson and D.W. Wichern, Applied multivariate statistical analysis, sixth edition. New Jersey: Pearson Prentice Hall, 2007. Capítulos 10-11. ISBN 978-0-13-187715-3.
- A.C. Rencher, Methods of multivariate analysis, second edition. New York: John Wiley, 2002. Capítulos 9 y 11. ISBN 0-471-41889-7.

### **i. Recursos necesarios**

---

El profesor de la asignatura hará accesible a los alumnos el conjunto de materiales y recursos de apoyo que considere adecuados utilizar en la preparación de la asignatura, tanto para los aspectos teóricos como los computacionales, a través de la página web de la asignatura o del servicio de reprografía del centro.



## 6. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1: PRELIMINARES	0.3 ECTS	Semana 1
Bloque 2: MÉTODOS DE EXPLORACIÓN DE DATOS	2 ECTS	Semanas 1 a 6
Bloque 3: TÉCNICAS DE ANÁLISIS DE DEPENDENCIA	0.7 ECTS	Semanas 6 a 8

## 7. Sistema de calificaciones – Tabla resumen

De los sistemas de evaluación descritos en la memoria de verificación de la titulación se utilizan los siguientes:

- Evaluación continua
  - Observación sistemática de actividad
  - Trabajos individuales y en grupo
  - Prácticas de Laboratorio
  - Solución de problemas
- Exámenes escritos
  - Pruebas de preguntas cortas
  - Pruebas de desarrollo
  - Solución de problemas

de acuerdo a la siguiente tabla

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Entrega de trabajos individuales, en grupo y exposición.	70%	Se propondrá la entrega de ejercicios y una práctica resuelta con datos reales. El profesor podrá solicitar al alumno que defienda oralmente alguno de los trabajos entregados.
Prueba final.	30% o 100%	Se realizará una prueba final con preguntas acerca del contenido de la asignatura y/o con casos prácticos a resolver. En esta prueba el alumno podrá utilizar, pero no compartir, la documentación que considere oportuna. <sup>(1)</sup> Ver criterios de calificación para detalle sobre el peso de la prueba escrita

### CRITERIOS DE CALIFICACIÓN

- **Convocatoria ordinaria:**
  - La calificación final será el máximo de la obtenida tal y como especifica en la tabla anterior (30% prueba final, 70% otras calificaciones) y la obtenida únicamente con la prueba final (100% prueba final).
- **Convocatoria extraordinaria:**
  - La calificación será la obtenida con la prueba final.



### 8. Anexo: Métodos docentes

Actividad	Metodología
<b>Clases teórico-prácticas</b>	<ul style="list-style-type: none"> <li>• Clase magistral participativa</li> <li>• Resolución de problemas y casos prácticos</li> <li>• Aprendizaje basado en problemas</li> </ul>
<b>Laboratorio</b>	<ul style="list-style-type: none"> <li>• Resolución de problemas y casos prácticos con apoyo informático</li> <li>• Aprendizaje basado en problemas</li> <li>• Realización de prácticas siguiendo un enfoque colaborativo</li> </ul>
<b>Seminarios</b>	<ul style="list-style-type: none"> <li>• Se realizarán seminarios cada dos semanas en las que los alumnos trabajarán y discutirán sobre la solución a problemas propuestos por el profesor</li> </ul>

### 9. Anexo: Cronograma de actividades previstas

#### CRONOGRAMA

- (T/M) Clases teórico-prácticas
- (L) Laboratorio
- (S) Seminario

Semana	Contenido	Actividades previstas	Evaluación	Presenciales	No Presenciales
<b>1)</b> 1ª	Temas 1 y 2	2h (T/M);2h (L)		4	4
<b>2)</b> 2ª	Tema 2	2h (T/M); 2h (S)		4	7
<b>3)</b> 3ª	Temas 2 y 3	2h (T/M);2h( L)		4	5
<b>4)</b> 4ª	Temas 3 y 4	2h (T/M); 2h (S)		4	7
<b>5)</b> 5ª	Temas 4 y 5	2h (T/M); 2h (L)		4	5
<b>6)</b> 6ª	Temas 5 y 6	2h (T/M);2h (S)		4	7
<b>7)</b> 7ª	Temas 6 y 7	2h (T/M); 2h (L)		4	5
<b>8)</b> 8ª	Tema 7	2h (S)		4	5

NOTA: La distribución de los temas por semana, y el número de horas que se ha de dedicar a los mismos es orientativo. Este podrá ser modificado si así lo demanda el desarrollo de la materia.