



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

## ASIGNATURA / **COURSE TITLE**

Curso Avanzado de Estadística / [Advanced Course in Statistics](#)

### 1.1. **Código / Course number**

30075

### 1.2. **Materia / Content area**

Estadística / [Statistics](#)

### 1.3. **Tipo / Course type**

Formación optativa / [Elective course](#)

### 1.4. **Nivel / Course level**

Máster M2 / [Master M2](#)

### 1.5. **Curso / Year**

2014-2015

### 1.6. **Semestre / Semester**

2º / [2nd \(Spring semester\)](#)

### 1.7. **Idioma / Language**

Español e inglés. (El curso se podrá impartir en inglés siempre y cuando, al menos, un alumno internacional matriculado en la asignatura lo solicite). / [Spanish and English. \(The course can be taught in English if at least one officially registered international student requests so\).](#)

### 1.8. **Requisitos previos / Prerequisites**

Conocimientos de probabilidad a nivel de *Probabilidad I*.  
Conocimientos de estadística a nivel de *Estadística I* (ver Plan de Estudios de Grado en Matemáticas).

[Background on Probability Theory at the level of the course \*Probability I\*.](#)  
[Statistical training at the level of the course \*Statistics I\* \(see the](#)



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

corresponding Programmes in the webpage of the Mathematics Department).

### 1.9. Requisitos mínimos de asistencia a las sesiones presenciales / **Minimum attendance requirement**

**60% de las clases / 60% of the classes**

### 1.10. Datos del equipo docente / **Faculty data**

Docente(s) / **Lecturer(s)** Amparo Baíllo  
Departamento de Matemáticas / **Department of Mathematics**  
Facultad / **Faculty** Ciencias  
Despacho - Módulo / **Office - Module** 203 - 8  
Teléfono / **Phone**: +34 91 497 7698  
Correo electrónico/**Email**: amparo.baillo@uam.es  
Página web/**Website**: <http://www.uam.es/amparo.baillo>  
Horario de atención al alumnado/**Office hours**: Se fija a petición individual del alumno / **To be fixed individually with each student**

### 1.11. Objetivos del curso / **Course objectives**

El curso está centrado en algunos temas de Estadística Multivariante. El objetivo es conocer algunos aspectos y técnicas de la teoría estadística que permite analizar datos de dimensión finita mayor que 1 (vectores). El análisis estadístico multivariante es una herramienta de investigación muy útil, tanto desde el punto de vista teórico (es un punto de partida para estudiar datos funcionales, se emplea en análisis de series temporales y en modelos de matemática financiera) como desde el punto de vista práctico (la mayoría de los datos reales disponibles en la actualidad son como mínimo multivariados).

El propósito de este curso es presentar algunos temas esenciales de la Estadística Multivariante que permitan a los estudiantes comprender y trabajar en temas de investigación actual como, por ejemplo, el análisis de datos de muy alta dimensión o dimensión infinita. Se pondrá especial énfasis en aquellos aspectos que presentan una elaboración matemática más interesante. También se describirán problemas de investigación de actualidad en cada uno de los temas del programa. Las clases teóricas se completarán con la implementación de las técnicas en programas de R.

**This course focuses on some topics of Multivariate Statistics. The aim is to introduce some aspects and techniques of this statistical theory developed**



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

to handle data of finite dimension larger than 1 (vectors). Multivariate statistical analysis is a very useful research tool, both from the theoretical viewpoint (a starting point for Functional Data Analysis; useful in Time Series Analysis and Mathematical Finance) and from the practical viewpoint (the majority of real data available nowadays are at least multivariate).

The objective of this course is to introduce the students to some essential aspects of Multivariate Statistics, which will allow them to understand and work on current research topics such as, for instance, the analysis of very high- or infinite-dimensional data. There will be a special emphasis on mathematically more interesting aspects. Current research topics will be mentioned all along the course. A fundamental complement to the theoretical lectures will be the implementation of the multivariate techniques in R.

## 1.12. **Contenidos del programa / Course contents**

### 1. **Descripción de datos multivariantes / Description of multivariate data**

- 1.1. Descripción numérica: matriz de datos, vector de medias, matriz de covarianzas, medidas globales de variabilidad. Dependencia lineal.  
Numerical description: data matrix, mean vector, covariance matrix, global measures of variability. Linear dependence.
- 1.2. Análisis gráfico y transformaciones.  
Graphical analysis and transformations.

### 2. **Distribuciones multivariantes / Multivariate distributions**

- 2.1. Vectores aleatorios: Media, matriz de covarianzas, transformaciones lineales, medidas de dependencia. Distancia basada en la noción de "energía".  
Random vectors: mean, covariance matrix, linear transformations, dependence measures. "Energy"-based statistics.
- 2.2. La distribución normal multidimensional. Modelos de mixturas.  
Multidimensional normal distribution. Mixture models.

### 3. **Reducción de la dimensión / Dimension reduction**

- 3.1. Componentes principales.  
Principal components.
- 3.2. Análisis factorial. Aplicación a finanzas.  
Factor analysis. Applications fo finance.

### 4. **Análisis de conglomerados / Cluster analysis**



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

- 4.1. Métodos jerárquicos.  
[Hierarchical methods.](#)
- 4.2. El algoritmo de  $k$ -medias.  
[k-means algorithm.](#)
- 4.3. Clasificación de máxima verosimilitud  
[Classification maximum likelihood](#)
- 5. Clasificación o análisis discriminante / [Classification or discriminant analysis](#)**
  - 5.1. Introducción al problema matemático. Regla Bayes. Metodología *plug-in*.  
[Introduction to the mathematical problem. Bayes rule. Plug-in methodology.](#)
  - 5.2. La regla lineal de Fisher.  
[Fisher's linear rule.](#)
  - 5.3. La regla de  $k$  vecinos más próximos.  
[k-nearest neighbours rule.](#)
  - 5.4. Clasificación multiclase  
[Multiclass classification](#)
- 6. Espacios de Hilbert con núcleo reproductor y su utilidad en Estadística Multivariante / [Reproducing kernel Hilbert spaces \(RKHS\) and their utility in Multivariate statistics](#)**

## 1.13. [Referencias de consulta](#) / [Course bibliography](#)

- [1] Claeskens, G. y Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press.
- [2] Devroye, L., Györfi, L. y Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, New York.
- [3] Everitt, B.S., Landau, S., Leese, M. y Stahl, D. (2011). *Cluster Analysis*. Wiley.
- [4] Genton, M.G. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, 2, 299-312.
- [5] Hall, P., Park, B. U. y Samworth, R. J. (2008). Choice of neighbor order in nearest-neighbor classification. *The Annals of Statistics*, 36, 2135-2152.
- [6] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*. 2nd edition. Springer.
- [7] Hofmann, T., Schölkopf, B. y Smöla, A.J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 36, 1171-1220.
- [8] Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques. Regression, Classification and Manifold Learning*. Springer.
- [9] Johnson, R. A. y Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall.



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

- [10] Jolliffe, I.T. (2002). *Principal Component Analysis*, 2nd. edition. Springer.
- [11] Kaufman, L. y Rousseeuw, P.J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- [12] McLachlan, G. y Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker.
- [13] McLachlan, G. y Peel, D. (2000). *Finite Mixture Models*. Wiley.
- [14] Pollard, D. (1981). Strong consistency of k-means clustering. *The Annals of Statistics*, 9,135-140.
- [15] Schneeweiss, H. y Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, 55, 105-124.
- [16] Schölkopf, B. y Smola, A.J. (2002). *Learning with Kernels. Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press.
- [17] Székely, G.J. y Rizzo, M.L. (2013). Energy statistics: a class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143, 1249-1272.
- [18] Székely, G.J., Rizzo, M.L. y Bakirov, N.K. (2007). Measuring and testing independence by correlation of distances. *The Annals of Statistics*, 35, 2769-2794.

## 2. Métodos docentes / Teaching methodology

Clases presenciales y discusión personal a través de tutorías o lecturas dirigidas. Presentación de trabajos.

Classroom sessions and personal discussion via tutorship activities or guided readings. Written essays based on the classroom materials.

## 3. Tiempo de trabajo del estudiante / Student workload

		Nº de horas	
Presencial	Clases teóricas	42h(21%)	68h (33%)
	Clases prácticas	4 h (2%)	
	Tutorías	10 h (5%)	
	Seminarios y trabajos	10 h (5%)	
	Examen final / proyecto	2h (1%)	



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

		Nº de horas	
No presencial	Elaboración de problemas	80h(40%)	132h (66%)
	Estudio semanal	46h(23%)	
	Preparación del examen	6h(3%)	
Carga total de horas de trabajo: 25 horas x 8 ECTS		200 h	

#### 4. Métodos de evaluación y porcentaje en la calificación final / Evaluation procedures and weight of components in the final grade

Examen y/o elaboración y presentación oral de un trabajo: 50% de la nota final.

Entrega de ejercicios, presentaciones orales y participación en clase: 50% de la nota final.

Final exam (elaboration and oral presentation of a research or review work): 50% of final mark.

Assignments and active participation in lectures: 50% of final mark.

EVALUACIÓN EXTRAORDINARIA / Make up exam:

Examen ante tribunal de Máster / Examination by a committee.

#### 5. Cronograma\* / Course calendar

Semana Week	Contenido Contents	Horas presenciales Contact hours	Horas no presenciales Independent study time
1	Apartado 1	4	6
2	Apartado 2.1	4	7
3	Apartado 2.2	4	9
4	Apartado 3.1	5	10
5	Apartado 3.2	4	9
6	Apartado 4.1	5	10



Asignatura: Curso Avanzado de Estadística  
Código: 30075  
Centro: Facultad de Ciencias  
Titulación: Máster en Matemáticas y Aplicaciones  
Nivel: Máster M2  
Tipo: Optativa  
Nº de créditos: 8 ECTS

Semana Week	Contenido Contents	Horas presenciales Contact hours	Horas no presenciales Independent study time
7	Apartado 4.2	5	10
8	Apartado 4.3	4	7
9	Apartado 5.1	4	9
10	Apartado 5.2	5	10
11	Apartados 5.3	4	6
12	Apartado 5.4	5	10
13	Apartado 6	5	11
14	Exposición de trabajos por los alumnos / <a href="#">presentation of students' works</a>	4	9
15	Exposición de trabajos por los alumnos / <a href="#">presentation of students' works</a>	4	9

\*Este cronograma tiene carácter orientativo / [This course calendar should be considered just as an approximate guidance.](#)