



## Proyecto/Guía docente de la asignatura

<b>Asignatura</b>	Métodos de Suavizado en el Análisis de Datos		
<b>Materia</b>	Probabilidad y Estadística		
<b>Módulo</b>	Formación Avanzada		
<b>Titulación</b>	Máster en Matemáticas		
<b>Plan</b>	645	<b>Código</b>	55041
<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		
<b>Lengua en que se imparte</b>	Español. Se manejarán materiales teóricos y programas informáticos en inglés		
<b>Profesor/es responsable/s</b>	Bonifacio Salvador González		
<b>Datos de contacto (E-mail, teléfono...)</b>	<a href="mailto:bonifacio.salvador@uva.es">bonifacio.salvador@uva.es</a>		
<b>Departamento</b>	Estadística e Investigación Operativa		



## 1. Situación / Sentido de la Asignatura

### 1.1 Contextualización

El Plan de Estudios se ha configurado en base a materias, módulos y asignaturas, eligiendo la afinidad disciplinar por grandes áreas de las Matemáticas como criterio unificador dentro de cada materia. La materia "Probabilidad y Estadística" está compuesta de seis asignaturas, tres en cada cuatrimestre:

Métodos Probabilísticos (Primer cuatrimestre)

Seminario de Procesos Estocásticos (Primer cuatrimestre)

Programación y Análisis de Datos con R (Primer cuatrimestre)

Métodos de Suavizado en el Análisis de Datos (Segundo cuatrimestre)

Seminario de Modelización Estadística (Segundo cuatrimestre)

Modelos de Optimización (Segundo cuatrimestre)

Con esta oferta se debería conseguir la formación adecuada para profundizar, desde un planteamiento matemático, en temas de Probabilidad y Estadística.

### 1.2 Relación con otras materias

### 1.3 Prerrequisitos

Conocimientos de Probabilidad y Estadística a nivel de Grado en Matemáticas o Grado en Estadística.

## 2. Competencias

### 2.1 Generales

Las que aparecen en la memoria de verificación del Master.

### 2.2 Específicas

Las que aparecen en la memoria de verificación del Master.



### 3. Objetivos

El objetivo general de la asignatura es introducir técnicas para la estimación no paramétrica de curvas, como la función de densidad y la función de regresión, basadas en suavizadores tipo núcleo, y sus aplicaciones al análisis de datos.

Desarrollar la intuición y habilidades matemáticas necesarias para la comprensión del suavizador núcleo, en el contexto de la estimación de la función de densidad y de la función de regresión.

Conocer y manejar de la librería “sm” de R (web R: <http://www.r-project.org/>).





#### 4. Contenidos y/o bloques temáticos

##### Bloque 1: Métodos de Suavizado en el Análisis de Datos

Carga de trabajo en créditos ECTS: 3

##### a. Contextualización y justificación

Estimación no paramétrica de las funciones de densidad y de regresión, basadas en suavizadores tipo núcleo, como alternativa a los procedimientos paramétricos estudiados en otras asignaturas.

##### b. Objetivos de aprendizaje

Conocer técnicas de suavizado en la estimación no paramétrica de la función de densidad.

Conocer técnicas de suavizado en la estimación no paramétrica de la función de regresión.

Proporcionar técnicas, para el análisis exploratorio de datos y construcción de modelos, que pueden aplicarse a una amplia variedad de tipos de datos, sin imponer suposiciones restrictivas, con el objeto de poner de manifiesto estructuras en los datos que pueden quedar enmascaradas con otras técnicas de la estadística paramétrica.

Desarrollar la intuición y habilidades matemáticas necesarias para la comprensión del suavizador núcleo, en el contexto de la estimación de las funciones de densidad y de regresión.

Proporcionar resúmenes de datos en dimensiones 1,2 y 3, mediante representaciones gráficas que complementan las de la estadística descriptiva.

Conocer y manejar de la librería "sm" de R (web R: <http://www.r-project.org/>).

##### c. Contenidos

Estimador núcleo de la función de densidad. Aplicaciones: test de normalidad, test de independencia, test de modalidad, bootstrap suavizado, análisis discriminante. Estimador núcleo de la función de regresión, en especial el estimador polinómico local. Aplicaciones.

Los temas a desarrollar son:

###### 1. ESTIMACIÓN DE FUNCIONES DE DENSIDAD UNIVARIANTES

El estimador núcleo. Propiedades. Métodos de elección del parámetro de suavizado. Otros estimadores. Parámetro de suavizada variable, métodos basados en vecinos más próximos.

###### 2. ESTIMACIÓN DE FUNCIONES DE DENSIDAD MULTIVARIANTES.

Matriz de suavizado. Núcleo producto y radialmente simétrico. Visualizaciones. Dificultades en grandes dimensiones.

###### 3. APLICACIONES. Test de normalidad, Test de independencia. Bootstrap suavizado. Regla de clasificación núcleo.

###### 4. ESTIMACIÓN DE LA FUNCIÓN DE REGRESIÓN

Estimador núcleo de Nadaraya-Watson. Estimador basado en el ajuste local de un polinomio. Estimador basado en vecinos más próximos. Estimador Loess. Regresión no paramétrica en dos dimensiones.

###### 5. VALIDACIÓN DE MODELOS DE REGRESIÓN PARAMÉTRICOS

Contraste de no efecto. Contraste de linealidad

###### 6. COMPARANDO CURVAS Y SUPERFICIES

Comparando estimadores de la densidad en una dimensión.

Comparando curvas y superficies de regresión.

###### 7. VEROSIMILITUD LOCAL Y SUAVIZADO MODELOS LINEALES GENERALIZADOS.



#### d. Métodos docentes

---

La asignatura se desarrollará mediante la realización de diversas actividades, clases teóricas en el aula, clases prácticas en el laboratorio de Estadística, tutorías individualizadas, pruebas puntuables y examen final.

Clases en el aula: La teoría básica necesaria será expuesta por el profesor en clases en el aula. Se ilustrará su aplicación mediante ejemplos.

Clases prácticas en el laboratorio. Los estudiantes realizarán prácticas con ordenador en el Laboratorio de Estadística para familiarizarse con el manejo de R y la librería sm. Se analizarán conjuntos de datos reales y simulados. Mediante el análisis de datos simulados pondremos de manifiesto el funcionamiento de los métodos introducidos. Con el análisis de conjuntos de datos reales y su interpretación se pretende que el estudiante sea capaz de extraer el conocimiento, sobre el modelo subyacente, que los métodos aprendidos ponen de manifiesto.

Tutorías. Concertadas en el despacho del profesor y tutorías vía correo electrónico.

#### e. Plan de trabajo

---

El 50% de las horas presenciales se desarrollarán en el aula, en las que se exponen todos los contenidos teóricos del programa de la asignatura y se dan las indicaciones necesarias para facilitar la posterior labor de estudio de los alumnos.

El 50% de las horas presenciales se desarrollarán en el Laboratorio de Estadística (Aula de Informática). En estas clases de laboratorio se introduce la librería "sm" de la plataforma R, con el objeto de implementar, sobre conjuntos de datos reales y/o simulados, todos y cada uno de los procedimientos desarrollados en las clases en el aula.

A lo largo del desarrollo de la asignatura, se propondrá a los estudiantes la realización de trabajos evaluables

#### f. Evaluación

---

##### Convocatoria ordinaria:

La calificación de cada alumno se hará mediante **evaluación continua** de los diferentes trabajos evaluables.

##### Convocatoria extraordinaria:

Los alumnos suspensos en la convocatoria ordinaria realizarán un examen presencial, de tres horas de duración máxima, cuya nota equivale al 100% de la evaluación. El examen se resolverá con ordenador, utilizando R y con entrega en Moodle.

#### g Material docente

---

*Es fundamental que las referencias suministradas este curso estén actualizadas y sean completas. Los profesores tienen acceso, a la plataforma Leganto de la Biblioteca para actualizar su bibliografía recomendada ("Listas de Lecturas"). Si ya lo han hecho, pueden poner tanto en la guía docente como en el Campus Virtual el enlace permanente a Leganto.*



**g.1 Bibliografía básica**

Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford.

Hastie, T and Tibshirani, R..(1990). *Generalized Additive Models*. Chapman and Hall.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall.

**g.2 Bibliografía complementaria**

Scott, D.W. (1992) *Multivariate Density Estimation*. Wiley.

Silverman B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag

**g.3 Otros recursos telemáticos (píldoras de conocimiento, blogs, videos, revistas digitales, cursos masivos (MOOC), ...)**

**h. Recursos necesarios**

**i. Temporalización**

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
3	2º cuatrimestre

*Añada tantas páginas como bloques temáticos considere realizar.*

**5. Métodos docentes y principios metodológicos**

La asignatura se desarrollará mediante la realización de diversas actividades, clases teóricas en el aula, clases prácticas en el laboratorio de Estadística, tutorías individualizadas, pruebas puntuables y examen final.

Clases en el aula: La teoría básica necesaria será expuesta por el profesor en clases en el aula. Se ilustrará su aplicación mediante ejemplos.

Clases prácticas en el laboratorio. Los estudiantes realizarán prácticas con ordenador en el Laboratorio de Estadística para familiarizarse con el manejo de R y la librería sm. Se analizarán conjuntos de datos reales y

simulados. Mediante el análisis de datos simulados pondremos de manifiesto el funcionamiento de los métodos introducidos. Con el análisis de conjuntos de datos reales y su interpretación se pretende que el estudiante sea capaz de extraer el conocimiento, sobre el modelo subyacente, que los métodos aprendidos ponen de manifiesto.

Tutorías. Concertadas en el despacho del profesor y tutorías vía correo electrónico.

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

ACTIVIDADES PRESENCIALES o PRESENCIALES A DISTANCIA <sup>(1)</sup>	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases Teóricas	14	Estudio y trabajo individual	35
Clases Prácticas en Laboratorio	14	Realización de trabajos evaluables	10
Realización de Exámenes	2		
Total presencial	<b>30</b>	Total no presencial	<b>45</b>
TOTAL presencial + no presencial			<b>75</b>

(1) Actividad presencial a distancia es cuando un grupo sigue una videoconferencia de forma síncrona a la clase impartida por el profesor.

## 7. Sistema y características de la evaluación

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Evaluación continua (Trabajos evaluables)	100%	

### CRITERIOS DE CALIFICACIÓN

- **Convocatoria ordinaria:**
  - Evaluación continua 100%
- **Convocatoria extraordinaria:**
  - Los alumnos suspensos en la convocatoria ordinaria realizarán un examen presencial, de tres horas de duración máxima, cuya nota equivale al 100% de la evaluación. El examen se resolverá con ordenador, utilizando R y con entrega en Moodle.

## 8. Consideraciones finales

