

**Proyecto/Guía docente de la asignatura**

Asignatura	Métodos de Suavizado en Estadística		
Materia	2- Técnicas Estadísticas		
Módulo			
Titulación	Grado en Estadística		
Plan	549	Código	47108
Periodo de impartición	2º Cuatrimestre	Tipo/Carácter	Optativa
Nivel/Ciclo	Grado	Curso	4º
Créditos ECTS	6		
Lengua en que se imparte	Español. Se manejarán materiales teóricos y programas informáticos en Inglés		
Profesor/es responsable/s	Bonifacio Salvador González		
Datos de contacto (E-mail, teléfono...)	bonifacio.salvador@uva.es		
Departamento	Estadística e Investigación Operativa		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

Con las asignaturas de probabilidad y estadística de los tres primeros cursos, en especial Inferencia estadística I y II, Regresión y Anova, Análisis de Datos, etc. se sientan las bases conceptuales de la estadística y las técnicas paramétricas más importantes. En esta asignatura se proponen alternativas no paramétricas a dichas técnicas.

1.2 Relación con otras materias

Esta asignatura se encaja en el plan de estudios dentro de la materia 2, denominada "Técnicas Estadísticas", en la que se reúnen los contenidos estadísticos avanzados de la titulación, articulados en 17 asignaturas, siendo "Métodos de Suavizado en Estadística" una de las optativas. En ella se presentan técnicas de suavizado núcleo, que constituyen alternativas no paramétricas al análisis de regresión, tests de ajuste o análisis discriminante entre otras.

1.3 Prerrequisitos

Conocimientos de probabilidad, inferencia estadística y análisis de regresión.





2. Competencias

Se incluyen a continuación las que aparecen en la memoria de verificación del plan de estudios.

2.1 Generales

- G1.** Capacidad para la gestión de la información: Incluye la capacidad para la búsqueda, manejo y exposición de información relevante de diversas fuentes, así como el manejo de las herramientas TIC necesarias.
- G2.** Capacidad para la abstracción y el razonamiento crítico: El modelado y análisis de datos de problemas reales exige una elevada capacidad de abstracción, y el razonamiento crítico es necesario para hacer interpretaciones y establecer conclusiones y soluciones con rigor científico.
- G3.** Capacidad para la puesta al día y el auto-aprendizaje: Incluye la capacidad para la búsqueda de recursos que permitan la solución de nuevos problemas o de nuevas técnicas, en un medio científico y tecnológico en continua evolución.

2.2 Específicas

- E1.** Recogida y tratamiento de datos: Incluye la capacidad para decidir sobre el diseño del procedimiento de obtención de datos. Capacidad para la búsqueda de información de fuentes diversas y para la elaboración de cuestionarios. Capacidad para manejar bases de datos y para llevar a cabo el tratamiento de los mismos.
- E2.** Descripción y síntesis de datos: Esta competencia es la que permitirá describir numéricamente y mediante gráficos diferentes características de interés de variables e individuos de los datos objeto de estudio, localización, escala, diferentes tipos de asociación, outliers,...
- E3.** Ajuste de modelos estadísticos y de investigación operativa: Incluye la capacidad para crear o reconocer un modelo adecuado al problema objeto de estudio. Capacidad para el ajuste del modelo mediante las herramientas adecuadas.
- E4.** Análisis de resultados, interpretación y validación de modelos: Incluye la capacidad para la selección y validación de un modelo. Capacidad para la interpretación de los modelos ajustados y las diferencias entre ellos.
- E5.** Extracción de conclusiones: Incluye la capacidad para interpretar los resultados del ajuste de un modelo seleccionado en términos del problema objeto de estudio, viendo su utilidad y/o proponiendo la necesidad de otras orientaciones del estudio.
- E6.** Presentación y comunicación de resultados: Incluye la capacidad para presentar los resultados de los análisis realizados, junto a las posibles soluciones a los problemas planteados por los demandantes del estudio en contextos diversos.

Competencias transversales

Instrumentales

- I1.** Capacidad de análisis y síntesis
- I2.** Capacidad de gestión de la información
- I3.** Capacidad de organización y planificación
- I4.** Conocimientos de informática relativos al ámbito de estudio
- I5.** Resolución de problemas
- I6.** Comunicación oral y escrita en lengua nativa
- I7.** Conocimiento de lenguas extranjeras
- I8.** Toma de decisiones

Personales

- P2.** Razonamiento crítico
- P4.** Compromiso ético

Sistémicas



- S1. Aprendizaje autónomo
- S2. Adaptación a nuevas situaciones
- S3. Motivación por el trabajo bien hecho
- S4. Iniciativa y espíritu emprendedor

3. Objetivos

El objetivo general de la asignatura es introducir técnicas no paramétricas, basadas en el suavizado núcleo, y su aplicación al análisis de datos.

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 la función de densidad y de la función de regresión.

Destacar la importancia del parámetro de suavizado en el compromiso entre el error sistemático y el error aleatorio.

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/>).



4. Contenidos y/o bloques temáticos

Bloque 1: Métodos de Suavizado en Estadística

Carga de trabajo en créditos ECTS: 6

a. Contextualización y justificación

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

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.

Destacar la importancia del parámetro de suavizado en el compromiso entre el error sistemático y el error aleatorio.

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. Selectores del parámetro de suavizado. 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. Selectores del parámetro de suavizado. Aplicaciones.

Los temas a desarrollar son:

1.INTRODUCCIÓN A LOS MÉTODOS DE SUAIVIZADO

Un compromiso entre los modelos paramétricos y no paramétricos.

Aplicaciones. Análisis de datos preliminares. Construcción de modelos. Bondad de ajuste. Estimación paramétrica. Modificando la metodología estándar.

2. ESTIMACIÓN DE FUNCIONES DE DENSIDAD UNIVARIANTES

El histograma. Propiedades: El sesgo y la varianza. Amplitud óptima del intervalo. Inconvenientes. El polígono de frecuencias.

El estimador núcleo. Propiedades: El sesgo y la varianza, el error cuadrático medio integrado. Núcleo y parámetro de suavizado óptimos. Inconvenientes: Sesgo en la frontera, Suavizado constante.

Métodos de elección del parámetro de suavizado. Suavizado óptimo normal. Suavizado por validación cruzada. Reglas plug-in.

Otros estimadores. Parámetro de suavizado variable, métodos basados en vecinos más próximos, métodos basados en series ortogonales.

3.INFERENCIA BASADA EN LA ESTIMACIÓN DE DENSIDADES

Bandas de variabilidad. Banda de variabilidad bootstrap. Test de normalidad. Banda de referencia normal.

4.ESTIMACIÓN DE FUNCIONES DE DENSIDAD MULTIVARIANTES.

Matriz de suavizado. Núcleo producto y radialmente simétrico. Visualizaciones.

Dificultades en grandes dimensiones.



Aplicaciones: Test de independencia. Regla de clasificación núcleo.

5. ESTIMACIÓN DE LA FUNCIÓN DE REGRESIÓN

Ideas básicas. Promedio móvil local. Estimador núcleo de Nadaraya-Watson. Estimador basado en el ajuste local de un polinomio. Estimador basado en vecinos más próximos. Estimador Loess. Estimador basado en series ortogonales. Regresión no paramétrica en dos dimensiones.

6. INFERENCIA EN REGRESIÓN NO PARAMÉTRICA

El sesgo y la varianza. Bandas de variabilidad.

7. VALIDACIÓN DE MODELOS DE REGRESIÓN PARAMÉTRICOS

Contraste de no efecto. Banda de referencia.

Validando una relación lineal. Test de pseudo razón de verosimilitud. Banda de referencia.

8. COMPARANDO CURVAS Y SUPERFICIES

Comparando estimadores de la densidad en una dimensión.

Riesgo relativo en dos dimensiones.

Comparando curvas y superficies de regresión.

Contraste de paralelismo.

9. APLICACIONES

Verosimilitud local y suavizado modelos lineales generalizados.

Aplicaciones al Análisis de Supervivencia.

10. INTRODUCCIÓN A LOS MODELOS ADITIVOS

Modelos aditivos y algoritmo de "backfitting".

Modelos semiparamétricos.

Modelos aditivos generalizados.

d. Métodos docentes

La asignatura se desarrollará mediante la realización de diversas actividades, clases en el aula, tanto teóricas como prácticas, clases prácticas en el laboratorio, 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. Esto hace difícil diferenciar claramente entre clases de teoría y clases prácticas en el aula. No obstante, podemos estimar que la "teoría" ocupará un 80% del tiempo total dedicado a las clases en el aula.

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, como se especifica en el calendario de actividades, se propondrá a los estudiantes la realización de trabajos evaluables, y se realizarán dos exámenes parciales. La evaluación tanto de los trabajos como de los exámenes parciales contribuye a la calificación final.



f. Evaluación

Criterios de evaluación y calificación para la asignatura de "Métodos de Suavizado en Estadística":

Convocatoria ordinaria (JUNIO):

La calificación de cada alumno se hará mediante **evaluación continua** de las diferentes actividades programadas en la asignatura: realización de prácticas y realización de exámenes parciales.

- En las fechas señaladas en el calendario general de actividades de 4º del Grado en Estadística, (ver <http://www.eio.uva.es/docencia/grado/>) se recogerán los **trabajos prácticos**, propuestos a los estudiantes con la anterioridad suficiente.

- Cada uno de los **trabajos prácticos** se calificará, en una escala de 0 a 10, informando a los estudiantes del resultado obtenido, así como de los errores cometidos, en su caso. La nota media **A** de todos ellos se utilizará directamente en la calificación final de la asignatura, en la convocatoria ordinaria (JUNIO), según se indica más abajo.

- En las fechas señaladas en el calendario general de actividades de 4º del Grado en Estadística, (ver <http://www.eio.uva.es/docencia/grado/>) Se realizarán, siempre dentro de las horas de clase, exámenes **parciales** que consistirán en dar respuesta a cuestiones sobre la materia estudiada hasta el momento, habitualmente con ayuda del ordenador.

- La nota media **B** de todos los **exámenes parciales**, en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura solo en la convocatoria ordinaria (JUNIO), según se indica más abajo.

El examen final (JUNIO) se realizará en el día y hora fijados en la primera convocatoria oficial de examen de la asignatura. La nota **C** del examen final (JUNIO), en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura en la convocatoria (JUNIO), según se indica más abajo.

La calificación final de la asignatura será:

$A*0.3 + B*0.7$ si $A*0.3 + B*0.7 \geq 5$

C si $A*0.3 + B*0.7 \leq 5$ o si $A*0.3 + B*0.7 \geq 5$ y el alumno renuncia a esta calificación.

Se hará sobre 10, con un decimal y su correspondencia cualitativa.

Convocatoria extraordinaria (JULIO):

- El examen final (JULIO) se realizará en el día y hora fijados en la segunda convocatoria oficial de examen de la asignatura. La nota **D** del examen final (JULIO), en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura en la convocatoria (JULIO), según se indica más abajo.

La calificación final de la asignatura **D**

Se hará sobre 10, con un decimal y su correspondencia cualitativa.



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
6	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 en el aula, tanto teóricas como prácticas, clases prácticas en el laboratorio, 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. Esto hace difícil diferenciar claramente entre clases de teoría y clases prácticas en el aula. No obstante, podemos estimar que la "teoría" ocupará un 80% del tiempo total dedicado a las clases en el aula.

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.

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, como se especifica en el calendario de actividades, se propondrá a los estudiantes la realización de trabajos evaluables, y se realizarán dos pruebas. La evaluación tanto de los trabajos como de las pruebas contribuye a la calificación final.

6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES o PRESENCIALES A DISTANCIA ⁽¹⁾	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases Teóricas	26	Estudio y trabajo autónomo individual	75
Clases Prácticas		Estudio y trabajo autónomo grupal	15
Laboratorios	26		
Realización de exámenes	8		
Total presencial	60	Total no presencial	90
TOTAL presencial + no presencial			150

(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)	30%	Se realizarán diversos trabajos
Evaluación continua (Exámenes parciales)	70%	Se realizarán dos exámenes parciales de dos horas de duración.



CRITERIOS DE CALIFICACIÓN

- **Convocatoria ordinaria:**
 - Evaluación continua 100%
 - Los alumnos que no sigan la evaluación continua o suspendan la evaluación continua o renuncien a la calificación por evaluación continua, 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....
- **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



