

Curso Académico: ( 2022 / 2023 )

Fecha de revisión: 16-05-2022

Departamento asignado a la asignatura: Departamento de Estadística

Coordinador/a: GARCIA PORTUGUES, EDUARDO

Tipo: Obligatoria Créditos ECTS : 3.0

Curso : 1 Cuatrimestre : 1

**REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)**

Matemáticas para el Análisis de Datos  
Estadística para el Análisis de Datos

**OBJETIVOS****\* Competencias básicas**

- CB6: Poseer y comprender los conocimientos que proporcionan una base u oportunidad para ser original en el desarrollo y/o aplicación de las ideas, a menudo en un contexto de investigación.
- CB9: Que los estudiantes sepan cómo comunicar sus conclusiones y el conocimiento y las razones últimas que las sustentan a audiencias especializadas y no especializadas de una manera clara e inequívoca.
- CB10: Que los estudiantes tengan las habilidades de aprendizaje que les permitan continuar estudiando de manera que sean en gran medida autodirigidos o autónomos.

**\* Competencias generales**

- CG1: Capacidad de aplicar las técnicas de análisis y representación de la información, para adaptarla a problemas reales.
- CG4: Sintetizar las conclusiones obtenidas del análisis de datos y presentarlas de forma clara y convincente en un entorno bilingüe (español e inglés) tanto escrito como oral.
- CG5: Ser capaz de generar nuevas ideas (creatividad) y anticiparse a nuevas situaciones, en los contextos de análisis de datos y toma de decisiones.
- CG6: Utilizar las habilidades sociales para el trabajo en equipo y para relacionarse con los demás de forma autónoma.
- CG7: Aplicar técnicas avanzadas de análisis y representación de información, para adaptarla a problemas reales.

**\* Competencias específicas**

- CE1: Aplicar en el desarrollo de métodos de análisis de problemas reales, conocimientos avanzados de inferencia estadística.
- CE2: Usar software libre como R y Python para la implementación de análisis estadísticos.
- CE5: Aplicar los fundamentos estadísticos avanzados para el desarrollo y análisis de problemas reales, que implican la predicción de una respuesta variable.
- CE6: Aplicar modelos no paramétricos para la interpretación y predicción de fenómenos aleatorios.
- CE10: Aplicar la modelización estadística en el tratamiento de problemas relevantes en el campo científico.

**\* Resultados del aprendizaje**

Adquisición de conocimientos sobre: 1) fundamentos estadístico-matemáticos del modelo de regresión lineal; 2) comparación y selección de modelos de regresión; 3) extensiones del modelo de regresión lineal (penalización, modelos no lineales, modelos con reducción de la dimensión, modelos lineales generalizados, etc.); 4) adaptaciones de los modelos lineales generalizados para trabajar con big data; 5) aprendizaje automático automatizado.

## DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Este curso está diseñado para dar una visión panorámica de varias herramientas disponibles para la modelización predictiva, a un nivel intermedio-avanzado. Esta visión abarca en profundidad los principales conceptos de los modelos lineales y los modelos lineales generalizados (con sus versiones penalizadas), y más superficialmente el enfoque del aprendizaje automático automatizado. Se hace hincapié en proporcionar las principales ideas sobre los fundamentos estadísticos/matemáticos de los modelos y en mostrar la aplicación efectiva de los métodos mediante el uso de software estadístico. Esto se logra mediante una mezcla de teoría y código reproducible.

1. Introducción
  - 1.1. Resumen del curso
  - 1.2. Notación general y repaso
  - 1.3. ¿Qué es la modelización predictiva?
2. Modelos lineales I: modelo lineal múltiple
  - 2.1. Formulación de modelos y mínimos cuadrados
  - 2.2. Supuestos del modelo
  - 2.3. Inferencia para los parámetros del modelo
  - 2.4. Predicción
  - 2.5. ANOVA
  - 2.6. Ajuste del modelo
3. Modelos lineales II: selección de modelos, extensiones y diagnósticos
  - 3.1. Selección del modelo
  - 3.2. Utilización de predictores cualitativos
  - 3.3. Relaciones no lineales
  - 3.4. Diagnóstico de modelos
  - 3.5. Técnicas reductoras de la dimensión
4. Modelos lineales III: penalización y big data
  - 4.1. Penalización
  - 4.2. Consideraciones sobre big data
5. Modelos lineales generalizados
  - 5.1. Formulación y estimación de modelos
  - 5.2. Inferencia para los parámetros del modelo
  - 5.3. Predicción
  - 5.4. Deviance
  - 5.5. Selección de modelos
  - 5.6. Diagnóstico de modelos
  - 5.7. Penalización
  - 5.8. Consideraciones sobre big data
6. Aprendizaje automático automatizado
  - 6.1. Introducción
  - 6.2. Explicabilidad
  - 6.3. Ejemplos en regresión
  - 6.4. Ejemplos en clasificación binaria
  - 6.5. Ejemplos en clasificación multiclase

El programa está sujeto a modificaciones menores debido al desarrollo del curso y/o al calendario académico.

## ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Las clases consisten en una mezcla de teoría (descripción de los métodos) y práctica (implementación y aplicación de los métodos). Se emplea el lenguaje estadístico R. Se espera que los estudiantes traigan sus propios portátiles para experimentar con el código durante algunas partes de las clases.

### \* Actividades formativas

- AF1: Clase teórica.
- AF2: Clase práctica.
- AF5: Tutorías.
- AF6: Trabajo en grupo.
- AF7: Trabajo individual.

- AF8: Pruebas de evaluación presenciales.

\* Metodologías docentes

- MD1: Exposiciones en clase del profesor con soporte de medios informáticos y audiovisuales, en las que se desarrollan los conceptos principales de la materia y se proporciona la bibliografía para complementar el aprendizaje de los alumnos.

- MD3: Resolución de casos prácticos, problemas, etc. planteados por el profesor de manera individual o en grupo.

- MD4: Exposición y discusión en clase, bajo la moderación del profesor de temas relacionados con el contenido de la materia, así como de casos prácticos.

- MD5: Elaboración de trabajos e informes de manera individual o en grupo.

## SISTEMA DE EVALUACIÓN

La evaluación ordinaria se realiza enteramente mediante evaluación continua. Esta se hace con una mezcla de:

- (a) dos cuestionarios sobre ideas clave y conceptos teóricos;
- (b) un conjunto de ejercicios prácticos a realizar en grupo;
- (c) participación activa en las clases y realización de ejercicios voluntarios.

La calificación de la evaluación continua (en la escala 0-10) es

$\min(0.5 * A + 0.5 * B + 0.10 * C, 10)$ ,

donde

- A (en la escala 0-10) es la calificación ponderada de los cuestionarios;

- B (en la escala 0-10) es la calificación de los ejercicios prácticos;

- C (en la escala 0-10) es la calificación correspondiente a (c).

Los estudiantes que no hayan seguido la evaluación continua podrán realizar un examen final con un valor del 60% de la calificación final.

La calificación en la convocatoria extraordinaria se establece por medio de un cuestionario y una entrega de un conjunto de ejercicios prácticos.

Se proporcionan más detalles en Aula Global. La evaluación está sujeta a modificaciones debido al desarrollo del curso y/o al calendario académico.

<b>Peso porcentual del Examen Final:</b>	0
<b>Peso porcentual del resto de la evaluación:</b>	100

## BIBLIOGRAFÍA BÁSICA

- James, G., Witten, D., Hastie, T. y Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer-Verlag, 2013

## BIBLIOGRAFÍA COMPLEMENTARIA

- Kuhn, M. y Johnson, K. Applied Predictive Modeling, Springer, 2013

- Li, Q. y Racine, J. S. Nonparametric Econometrics, Princeton University Press, 2007

- Peña, D. Regresión y Diseño de Experimentos, Alianza Editorial, 2002

- Wasserman, L. All of Statistics, Springer-Verlag, 2004

- Wood, S. N. Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC, 2006