

19-20

MÁSTER UNIVERSITARIO EN
INGENIERÍA Y CIENCIA DE DATOS

GUÍA DE ESTUDIO PÚBLICA



MODELADO ESTADÍSTICO DE DATOS

CÓDIGO 31110018

UNED

19-20

MODELADO ESTADÍSTICO DE DATOS

CÓDIGO 31110018

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

EQUIPO DOCENTE

HORARIO DE ATENCIÓN AL ESTUDIANTE

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

RESULTADOS DE APRENDIZAJE

CONTENIDOS

METODOLOGÍA

SISTEMA DE EVALUACIÓN

BIBLIOGRAFÍA BÁSICA

BIBLIOGRAFÍA COMPLEMENTARIA

RECURSOS DE APOYO Y WEBGRAFÍA

ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19

Nombre de la asignatura	MODELADO ESTADÍSTICO DE DATOS
Código	31110018
Curso académico	2019/2020
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	SEMESTRE 1
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

Los modelos estadísticos son técnicas fundamentales que permiten extraer información de los datos, determinando si ciertas variables explicativas (predictores o variables independientes) influyen en una variable respuesta (variable resultado o dependiente). Esta asignatura se enmarca, por tanto, dentro, del aprendizaje supervisado, donde se modeliza una variable respuesta cuantitativa (modelos lineales de regresión) o una variable respuesta cualitativa (modelos lineales de clasificación).

La asignatura pretende hacer asequible los contenidos matemáticos mínimos necesarios para entender el funcionamiento interno de los modelos. Y, por supuesto, sin olvidar la parte práctica de utilización de los modelos mediante software estadístico libre.

Esta asignatura forma parte del Máster Universitario en Ingeniería y Ciencia de Datos. Consta de 6 ECTS y es de carácter obligatorio.

La presente guía contiene información de carácter general sobre la asignatura: requisitos y recomendaciones, equipo docente, horario de atención, resultados de aprendizaje, contenidos, metodología, plan de trabajo, sistema de evaluación, bibliografía básica, bibliografía complementaria, recursos de apoyo y glosario.

Es de destacar que en esta asignatura, la interacción con los alumnos es constante en la plataforma virtual y que se envían noticias sobre dicha interacción al correo de la UNED del alumno, por lo que conviene acceder a la plataforma y a dicho correo de forma regular (al menos una vez por semana).

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Es conveniente haber estudiado previamente una asignatura básica de Estadística a nivel de grado y manejar con soltura el cálculo matricial. No obstante, se repasarán conceptos básicos estadísticos en los primeros temas y se darán ayudas sobre cálculo matricial cuando se necesite. Por otra parte, es recomendable conocer (leer y escribir) el inglés técnico.

Como recomendación general para esta asignatura conviene tener en cuenta lo siguiente:

- El foro de la plataforma virtual es un espacio vivo en el que el equipo docente está presente de forma continua para ayudar al alumno a progresar en su estudio.
- En el foro de la plataforma virtual, se va marcando un ritmo de estudio semanal para poder llevar al día la asignatura (para aquél que pueda y/o desee).

- En todo momento es factible reengancharse a la asignatura gracias a los resúmenes que se van poniendo en la plataforma virtual dentro de las FAQ.

Por último, conviene remarcar que es importante que el estudiante rellene las encuestas de satisfacción de esta asignatura (y de todas) dando sugerencias porque éstas nos ayudan a ir mejorando la docencia en ella.

EQUIPO DOCENTE

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

EMILIO LETON MOLINA (Coordinador de asignatura)
emilio.leton@dia.uned.es
91398-9473
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
INTELIGENCIA ARTIFICIAL

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

OLGA C. SANTOS MARTÍN-MORENO
ocsantos@dia.uned.es
91398-9388
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
INTELIGENCIA ARTIFICIAL

HORARIO DE ATENCIÓN AL ESTUDIANTE

- D. Emilio Letón Molina
- Tfno: 91 398 9473 (lunes lectivos, de 14:30 a 18:30 h.)
- Despacho 3.04
- Email: emilio.leton@dia.uned.es
- D^a. Olga C. Santos Martín
- Tfno: 91 398 9388 (lunes lectivos, de 11:00 a 15:00 h.)
- Despacho 3.01
- Email: ocsantos@dia.uned.es

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS BÁSICAS

CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones

últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS GENERALES

CG1 - Identificar los métodos apropiados para la solución de problemas asociados a la ciencia de datos y la analítica de información

CG2 - Ser capaz de aplicar diferentes técnicas de aprendizaje máquina, seleccionando el algoritmo óptimo que genere modelos precisos y permita el desarrollo de soluciones predictivas en diferentes ámbitos de uso

CG5 - Utilizar las habilidades de científico de datos y/o ingeniero de datos en entornos de trabajo multidisciplinares y ser capaz de distinguir/organizar las diferentes actividades de los roles en dicho entorno

COMPETENCIAS TRANSVERSALES

CT1 - Ser capaz de abordar y desarrollar proyectos innovadores en entornos científicos, tecnológicos y multidisciplinares.

CT2 - Ser capaz de tomar decisiones y formular juicios basados en criterios objetivos (datos experimentales, científicos o de simulación disponibles).

COMPETENCIAS ESPECÍFICAS

CE1 - Conocer los fundamentos de la inferencia estadística y el análisis probabilístico y desarrollar diferentes tipos de modelos probabilísticos

CE2 - Desarrollar aplicaciones/servicios/scripts orientados a la analítica de datos y analizar el uso de diferentes librerías para el desarrollo e implementación de métodos numéricos, algoritmos y modelos asociados a los datos

CE6 - Diseñar mecanismos de evaluación de modelos de aprendizaje y comprender las métricas usadas para dicha evaluación

RESULTADOS DE APRENDIZAJE

Los resultados más relevantes que se pretenden alcanzar con el estudio de esta asignatura son los siguientes:

- Explicar los conceptos principales de los modelos estadísticos.
- Distinguir las diferencias y las relaciones entre los modelos lineales de regresión y clasificación.
- Explicar los fundamentos matemáticos de los modelos.
- Aplicar los modelos a datos reales.
- Interpretar los resultados obtenidos de los modelos.
- Evaluar la calidad de ajuste de los modelos.
- Contrastar el cumplimiento de los requisitos de aplicación de los modelos.

CONTENIDOS

Tema 1: Principales técnicas estadísticas

Este tema tiene como objetivo repasar las técnicas estadísticas principales. Se contemplan aquellas en las que hay una variable respuesta y una variable explicativa (t-Student, ANOVA 1 factor, Ji-Cuadrado, Fisher, Mann-Whitney y Kruskal-Wallis).

Tema 2: Índices de riesgo

Este tema explica las medidas básicas de riesgo en una tabla de contingencia (diferencia de riesgos, riesgo relativo y odds ratio).

Tema 3: Técnicas diagnósticas

En el caso de que una variable explicativa se utilice para diagnosticar aparecen nuevos conceptos estadísticos como: sensibilidad y especificidad, valores predictivos, likelihood ratios y curvas ROC.

Tema 4: Introducción al aprendizaje supervisado

Se introducen los conceptos básicos del aprendizaje estadístico en general y del aprendizaje supervisado en particular.

Tema 5: Regresión lineal

En este tema se estudia el modelo básico de regresión lineal que modeliza una respuesta cuantitativa.

Tema 6: Variantes del modelo de regresión lineal + Práctica 1

En este tema se tratan algunas variantes de modelos de regresión lineal como: ridge regression, LASSO, PCR o PLS.

Tema 7: Análisis discriminante lineal

En este tema se estudia el primero de los modelos básicos de clasificación que modeliza una respuesta cualitativa.

Tema 8: Regresión logística

En este tema se estudia el segundo de los modelos básicos de clasificación que modeliza una respuesta cualitativa.

Tema 9: Evaluación y selección de modelos + Práctica 2

En este tema se describen los métodos principales de evaluación de modelos y cómo usar dichos métodos para elegir entre varios modelos.

Tema 10: Fundamentos de la inferencia bayesiana

En la estadística llamada clásica o frecuentista, los parámetros son constantes desconocidas que hay que estimar. Sin embargo, en la llamada estadística bayesiana, a los parámetros se les considera variables aleatorias. En este tema se dan los conceptos básicos de la inferencia bayesiana.

METODOLOGÍA

Esta asignatura ha sido diseñada para la enseñanza a distancia. Por tanto, el sistema de enseñanza-aprendizaje estará basado en gran parte en el estudio independiente o autónomo del estudiante. Para ello, el estudiante contará con un Manual Didáctico que permitirá su trabajo autónomo. Asimismo, mediante la plataforma virtual de la UNED existirá un contacto continuo entre el equipo docente y los estudiantes, así como una interrelación entre los propios estudiantes a través de los foros, importantísimo en la enseñanza no presencial. Las actividades formativas para el estudio de la asignatura son las siguientes:

- Estudios de contenidos: 54 horas.
- Tutorías: 6 horas.
- Actividades en la plataforma virtual: 15 horas.
- Prácticas informáticas: 60 horas.
- Trabajos: 15 horas.

Además, en esta asignatura la evaluación continua se hace a través de la participación en el foro y de dos prácticas. Estas dos formas de evaluación continua se detallan a continuación.

Participación en el foro

La puntuación de la participación en el foro será de 0 a 10.

Por cada participación relevante, el equipo docente dará un punto al alumno. Se entiende por participación relevante cuando se contesta de manera acertada la pregunta de otro alumno o cuando se comunica un aspecto interesante relacionado con la asignatura. No se considera relevante preguntar simplemente una pregunta con una duda en el foro.

El equipo docente para favorecer que cualquier estudiante pueda participar en el foro irá

también haciendo preguntas cada semana, por lo que todo el mundo podrá optar a puntuar en este apartado, incluso aunque se incorpore tarde al estudio de la asignatura. No se tendrán en cuenta contestaciones repetidas de otros alumnos. Señalar, por último, que la experiencia demuestra que no es difícil conseguir la máxima puntuación y que no hay que agobiarse por conseguirla en las primeras semanas: ¡¡hay tiempo suficiente!!

El último día que se considera para poder puntuar en la participación en el foro es el 13-Ene. En la plataforma virtual estará publicada una tarea desde el 31-Oct para que el estudiante rellene la fecha de su última aportación relevante. Esto permite que el equipo docente pueda decirle su nota con anterioridad a examinarse y en caso de que hubiera tiempo, incluso podría mejorarla.

Prácticas

La puntuación de cada práctica será de 0 a 10. La primera práctica se podrá entregar desde el 18-Nov al 9-Dic y la segunda práctica desde el 16-Dic al 7-Ene.

Finalmente, conviene remarcar que el equipo docente sólo evaluará la participación en el foro y las prácticas durante el cuatrimestre en el que se imparte la asignatura. No obstante, debe tenerse en cuenta que para la convocatoria de septiembre, se mantendrá la nota obtenida en dicha evaluación continua durante el cuatrimestre en el que se imparte la asignatura.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen de desarrollo
Preguntas desarrollo	4
Duración del examen	120 (minutos)
Material permitido en el examen	

Calculadora.

Criterios de evaluación

El examen consta de 4 preguntas de desarrollo similares a las que componen las prácticas.

% del examen sobre la nota final	50
Nota del examen para aprobar sin PEC	10
Nota máxima que aporta el examen a la calificación final sin PEC	5
Nota mínima en el examen para sumar la PEC	0
Comentarios y observaciones	

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad

Si

Descripción

La prueba presencial se describe en el apartado anterior. Los únicos trabajos evaluables en esta asignatura se describen en el apartado "PRUEBAS DE EVALUACIÓN CONTINUA (PEC)" y son no presenciales.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?

Si, PEC no presencial

Descripción

La evaluación continua en esta asignatura se hace a través de la participación en el foro y de dos prácticas. La información detallada se encuentra en el apartado de "Metodología"

Criterios de evaluación

La participación en el foro y de cada práctica se evalúan de 0 a 10. Los criterios detallados se encuentran en el apartado de "Metodología".

Ponderación de la PEC en la nota final

La ponderación de la participación en el foro es de un 10% y la de cada práctica de un 20%.

Fecha aproximada de entrega

La fecha última para optar a la nota de participación en el foro es el 13-Ene, para entregar la primera práctica el 9-Dic y para entregar la segunda el 7-Ene.

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?

No

Descripción

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

Nota examen (Ex): 0-10.

Nota práctica 1 (P1): 0-10.

Nota práctica 2 (P2): 0-10.

Nota participación curso virtual (F): 0-10.

Nota final $NF=0,5*NE+0,2*P1+0,2*P2+0,1*F$

BIBLIOGRAFÍA BÁSICA

En el espacio de esta asignatura, en la plataforma de cursos virtuales de la UNED, se encuentra un Manual Didáctico específico donde se detalla todo el material de estudio.

BIBLIOGRAFÍA COMPLEMENTARIA

La bibliografía complementaria de esta asignatura es:

- Hosmer Jr, D. W., Lemeshow, S., &Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.
- Letón, E, &Pedromingo, A. (2001). Introducción al análisis de datos en meta-análisis. Ediciones Díaz de Santos.
- Peña, D. (2013). Análisis de datos multivariantes. McGraw-Hill España.
- Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. Oxford University Press.

RECURSOS DE APOYO Y WEBGRAFÍA

En esta asignatura se utiliza el programa estadístico R, disponible en la página web de R **ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19**

<https://app.uned.es/evacaldos/asignatura/adendasig/31110018>

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.