

Guía docente

270210 - PIE2 - Probabilidad y Estadística 2

Última modificación: 19/07/2023

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa.
Titulación: GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura obligatoria).
Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Catalán

PROFESORADO

Profesorado responsable: MARTA PÉREZ CASANY
Otros: Primer quadrimestre:
VÍCTOR PEÑA PIZARRO - 11, 13
MARTA PÉREZ CASANY - 11, 12, 13

CAPACIDADES PREVIAS

Para seguir satisfactoriamente este asignatura, el estudiante tiene que haber comprendido los contenidos de las asignaturas PiE1 y Cálculo.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.

Genéricas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

Transversales:

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Básicas:

CB1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB3. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

METODOLOGÍAS DOCENTES

La mitad de las clases serán exposiciones de contenidos teóricos, y la otra mitad serán de tipo práctico y se realizarán o bien en aula informática o bien en aula electrificada con el propio portátil del estudiante. Al acabar las clases prácticas se propondrán ejercicios de análisis a los estudiantes para éstos que puedan trabajar de forma autónoma. dichos ejercicios se discutirán en la próxima sesión.



OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Aprender a contruir modelos estadísticos para sintetizar o información, explicar una variable respuesta en función de otras y hacer predicciones.
2. Entender los conceptos básicos y la filosofía alrededor de la estadística bayesiana
3. Aprender un paquete estadístico y su utilidad para analizar datos reales.
4. Aprender a validar modelos estadísticos, y a mejorarlos cuando estos no sean adecuados al problema.
5. Aprender a hacer documentos con los resultados obtenidos tras un análisis de datos.
6. Entender la diferencia entre la estadística Bayesiana y frecuentista
7. Identificar el modelo estadístico mas apropiado para cada situación
8. Aprender a interpretar los resultados de un modelo ajustado
9. entender los conceptos de validación cruzada, y de sobre y sub ajustes.
10. Utilizar los modelos obtenidos para hacer predicciones.
11. Entender la diferencia entre parámetro y estimador de un parámetro. Saber resolver problemas de inferencia estadística en el entorno de los modelos lineales y lineales generalizados.
12. aprender a incorporar variables categóricas en los modelos lineales y lineales generalizados.
13. Analizar con sentido crítico y de forma autónoma temas y datos relevantes para la sociedad.
14. Llevar a cabo la estimación de parámetros con intervalos de confianza
15. Entender la importancia de los test de hipótesis. Saber como realizar los tests clásicos y también técnicas para encarar posibles nuevos test.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	30,0	20.00
Horas aprendizaje autónomo	90,0	60.00
Horas grupo pequeño	30,0	20.00

Dedicación total: 150 h

CONTENIDOS

Distribuciones relacionadas con la Normal. Estimación por intervalos de confianza.

Descripción:

Distribucions Chiquadrat, t-d'Student i F-Fisher-Snedecor. Definició d'interval de confiança. IC per un valor esperat, per una variància, per una probabilitat i per la diferència de dos valors esperats i dues probabilitats. Quantitats pivotals.

Test de hipótesis

Descripción:

Conceptes generals en l'entorn dels test d'hipòtesis. Comparacions d'una esperança i una variància amb un valor concret. Comparació de dos valors esperats, comparació de dues variàncies. Comparació d'una probabilitat amb un valor concret. Comparació de dues probabilitats.

Modelo lineal

Descripción:

Definición de modelo lineal. Estimación de parámetros. Tabla anova y medidas de bondad de ajuste. Inferencia sobre los parámetros. predicción. Validación del modelo. Selección del modelo. Interpretación del modelo. Sesgo, colinealidad, causalidad. Uso de variables explicativas categóricas.



Modelos lineal generalizado

Descripción:

Definición de modelo lineal generalizado. Modelos para respuesta binaria. Estimación de los parámetros. Validación del modelo. Selección de modelos. Interpretación de modelos.

Introducción a la estadística Bayesiana.

Descripción:

Teorema de Bayes. Modelo Bayesiano. distribución predictiva a priori y a posteriori. elección de la distribución a priori.

ACTIVIDADES

Modelos lineales

Descripción:

Definición de modelo lineal. Estimación e inferencia en un modelo lineal. Predicción. Selección del modelo. Interpretación del modelo. Uso de variables categóricas en modelos lineales. Modelos no lineales con respuesta Normal.

Objetivos específicos:

1, 3, 4, 5, 7, 8, 10, 11, 12, 13

Competencias relacionadas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CB3. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 59h

Grupo grande/Teoría: 12h

Grupo pequeño/Laboratorio: 12h

Aprendizaje autónomo: 35h



Modelo lineal generalizado

Descripción:

Definición de modelo lineal generalizado. Modelos con respuesta binaria. Estimación e inferencia en un modelo lineal generalizado. Predicción, interpretación y selección del modelo.

Objetivos específicos:

1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13

Competencias relacionadas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CB3. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 27h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 6h

Aprendizaje autónomo: 15h



Estadística Bayesiana

Descripción:

Modelo estadístico. Inferencia basada en la verosimilitud. Modelo Bayesiano. Distribución "a posteriori". distribución predictiva "a priori" y "a posteriori". Elección de la "a priori". Inferencia Bayesiana. Validación del modelo.

Objetivos específicos:

2, 6

Competencias relacionadas:

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CB1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 14h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 10h

Distribuciones relacionadas con la Normal. Estimación por intervalos de confianza.

Descripción:

Se definen las distribuciones ji-cuadrado, t-d'Student y F-Fisher-Snedecor. Se define intervalo de confianza y cantidad pivotal. Los IC más útiles se computan en clase.

Objetivos específicos:

13, 14

Competencias relacionadas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CB3. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 23h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 15h

Test de hipòtesis

Descripción:

Se definen los conceptos relacionados con los test de hipótesis. Se llevan a cabo las comparaciones de un valor esperado y una varianza a un valor concreto. De dos valores esperados. De dos varianzas y de dos probabilidades.

Objetivos específicos:

5, 13, 15

Competencias relacionadas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CB3. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 27h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 6h

Aprendizaje autónomo: 15h

SISTEMA DE CALIFICACIÓN

Habrà un examen parcial y un final. Asimismo durante el curso se realizaràn diversos ejercicios de análisis de datos.

El examen parcial corresponderà a la parte de intervalos de confianza y tests de hipótesis.

El examen final corresponderà al resto de contenidos de la asignatura.

La nota de curso se obtiene como la media aritmética de los ejercicios realizados durante el curso.

Nota Asignatura = $0,25 \cdot N_{\text{Curso}} + 0,25 \cdot \text{Parcial} + 0,5 \cdot \text{Final}$

Para el caso de los estudiantes que vayan a la reevaluación, la nota final se calcularà de la forma siguiente:

Nota Asignatura = $\max(N_{\text{Reevaluación}}, 0,25N_{\text{Curso}} + 0,75N_{\text{Reevaluación}})$

BIBLIOGRAFÍA

Básica:

- James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An introduction to statistical learning. Springer, 2013. ISBN 97-1461471370.
- Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.. Bayesian data analysis. 3rd ed. Chapman & Hall, 2014. ISBN 978-1439840955.
- Weisberg, S. Applied linear regression. 4th ed. John Wiley and Sons, 2014. ISBN 9780471704096.
- Dobson, A.J.; Barnett, A.G. An introduction to generalized linear models. 4th ed. Chapman & Hall, 2018. ISBN 978-1138741515.
- Dobrow, R.P. Introduction to stochastic processes with R. John Wiley and Sons, 2016. ISBN 978-1118740651.