

Guía docente

270950 - SIM - Inferencia Estadística y Modelización

Última modificación: 23/11/2023

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa.

Titulación: MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (Plan 2021). (Asignatura obligatoria).

Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: LIDIA MONTERO MERCADÉ

Otros: Primer quadrimestre:
JOSEP FRANQUET FÀBREGAS - 12
LIDIA MONTERO MERCADÉ - 11, 12

CAPACIDADES PREVIAS

Los estudiantes deben tener conocimientos suficientes de álgebra y análisis matemático para asimilar los conceptos relacionados con el álgebra de conjuntos, series numéricas, funciones de variables reales de una o más dimensiones, derivación e integración. Los alumnos deben haber cursado un curso de probabilidad y estadística

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

Genéricas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

Transversales:

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Básicas:

CB6. 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.

CB9. 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.

METODOLOGÍAS DOCENTES

El aprendizaje de la asignatura consta de tres fases diferenciadas:

1. Adquisición de los conocimientos específicos mediante el estudio de la bibliografía y del material proporcionado por los profesores.
2. La adquisición de las destrezas en las técnicas específicas de análisis de datos, selección del proceso de modelado estadístico y validación del modelo y
3. Integración de los conocimientos, destrezas y competencias (específicas y transversales) mediante la resolución de casos de estudio reales.

En las clases de Teoría exponen los fundamentos de las metodologías y técnicas propias de la asignatura. Las clases de laboratorio sirven para aprender la utilización de las técnicas específicas para la resolución de problemas, utilizando las herramientas informáticas adecuadas, en este sentido los alumnos deberán primero de repetir un problema solucionado por los profesores y luego solucionar uno similar al primer . Mientras que el Casos de Estudio, resuelto en grupos y en horas básicamente de autoaprendizaje, sirven para poner en práctica los conocimientos, destrezas y competencias en la resolución de casos reales.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

- 1.Saber realizar procesos de inferencia basados en los datos y de forma paramétrica tradicional para la toma de decisiones.
- 2.Saber hacer un informe sobre la calidad de los datos y pre-procesado
- 3.Determinación de las características significativas dirigidas a targets numéricos y categóricos en grupos de individuos
- 4.Estimación de los parámetros e interpretación de los modelos lineales de respuesta normal
- 5.Validación de los modelos de respuesta normal. Identificación de datos inusuales e influyentes. Análisis de los residuos
- 6.Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta normal
- 7.Estimación de los parámetros e interpretación de los modelos lineales de respuesta binaria
- 8.Validación de los modelos de respuesta binaria. Identificación de datos inusuales e influyentes. Tipología de los residuos
- 9.Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta binaria
- 10.Estimación de los parámetros e interpretación de los modelos lineales de respuesta polinómica nominal y ordinal
- 11.Validación de los modelos de respuesta polinómica nominal y ordinal. Identificación de datos inusuales e influyentes.
- 12.Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta polinómica nominal y ordinal
- 13.Estimación de los parámetros e interpretación de los modelos lineales para conteos
- 14.Validación de los modelos de por conteos. Identificación de datos inusuales e influyentes. Tipología de los residuos. Diagnóstico sobredispersión. Modelos probabilísticos paramétricos
- 15.Inferencia de hipótesis sobre parámetros simples y múltiples en modelos para conteos
- 16.Saber hacer un diseño de experimentos factorial y factorial fraccional

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo pequeño	27,0	18.00
Horas aprendizaje autónomo	96,0	64.00
Horas grupo grande	27,0	18.00

Dedicación total: 150 h

CONTENIDOS

Inferencia clásica versus fisheriana

Descripción:

Inferencia clásica. Función de verosimilitud. Propiedades de MLE. Prueba de razón de verosimilitud. Procedimientos inferenciales paramétricos vs no paramétricos. Usar datos históricos para probar hipótesis. Vínculos con la inferencia de Fisher y el bootstrapping.



Calidad de los datos

Descripción:

Outliers univariados y multivariados.
Datos perdidos. Procedimientos de imputación: deterministas, estocásticos.

Modelos lineales normales

Descripción:

Descripción del modelo lineal normal. Estimación por mínimos cuadrados. Comparación de modelos. Bondad de ajuste.
Diagnóstico: datos influyentes y valores atípicos. Uso de variables explicativas categóricas. Selección de modelo. Predicción.
Estimación de redes neuronales de modelos de regresión lineal.

Modelos lineales generalizados

Descripción:

Definición de los modelos lineales generalizados. Modelos para datos de respuesta binaria. Modelos para contar datos. Problemas de sobredispersión. Datos de respuesta multinomial. Comparación de modelos. Diagnóstico: datos influyentes y valores atípicos.
Comparación y selección de modelos.

Diseño de experimentos

Descripción:

Diseños experimentales factoriales y factoriales fraccionales.
Técnicas modernas de análisis de datos para diseño experimental

ACTIVIDADES

Inferencia clásica versus fisheriana

Descripción:

Saber diferenciar las condiciones de aplicabilidad de los distintos métodos de inferencia y saber escoger la más adecuada al proceso de Ciencia de Datos entre manos.

Realizar procesos de inferencia para generar conclusiones sobre las poblaciones. Usar los valores p, los intervalos de confianza y las pruebas por permutaciones para la toma de decisiones y la interpretación de los análisis en un problema de Ciencia de Datos recurrente o puntual.

Objetivos específicos:

1

Competencias relacionadas:

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

CB6. 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.

Dedicación: 18h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 12h

Calidad de los datos

Descripción:

Problemas en la calidad de los datos: Se trata de ver en el Caso de Estudio los problemas que presentan o pueden presentar los datos: inconsistencia, redundancia. Datos faltantes. Outliers. Cómo se hace un Informe de calidad de los datos. En que consiste la estandarización de los datos.

Objetivos específicos:

2

Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 7h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 3h

Determinación de las características significativas dirigidas a targets numéricos y categóricos en grupos de individuos

Descripción:

Aplicación de la inferencia estadística para determinar las relaciones entre las variables presentes en una BBDD y una variable de respuesta (numérica o categórica)

Objetivos específicos:

3

Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 5h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 1h

Estimación de los parámetros e interpretación de los modelos lineales de respuesta normal

Descripción:

Perspectiva del modelado por técnicas de regresión lineal: componentes estadísticas implicadas. Roles: variables de respuesta / explicativas. Estimación por mínimos cuadrados. Propiedades de los estimadores. Procesos inferenciales involucrados.

Objetivos específicos:

4

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 8h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 4h

Validación de los modelos de respuesta normal. Identificación de datos inusuales e influyentes. Análisis de los residuos

Descripción:

Elementos que intervienen en la validación del modelado por regresión. Valores influyentes y / o atípicos

Objetivos específicos:

5

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 8h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 4h

Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta normal

Descripción:

Inferencia sobre los estimadores de los parámetros en modelos lineales de respuesta normal. Intervalos de confianza, regiones de confianza. Contrastes de hipótesis simples, múltiples, combinaciones lineales. Inferencia sobre las predicciones y cálculos de intervalos de confianza.

Objetivos específicos:

6

Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 6h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 2h



Estimación de los parámetros e interpretación de los modelos lineales de respuesta binaria

Descripción:

Estimación máximo verosímil. Rol de la función de enlace. Función de enlace empleadas. Propiedades de los estimadores. Procesos inferenciales involucrados.

Objetivos específicos:

7

Competencias relacionadas:

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

Dedicación: 6h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 2h

Validación de los modelos de respuesta binaria. Identificación de datos inusuales e influyentes. Tipología de los residuos

Objetivos específicos:

8

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 6h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 2h



Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta binaria

Descripción:

Inferencia sobre estimadores de parámetros en modelos lineales de una respuesta binaria. Intervalos de confianza. Contrastes de hipótesis simples, múltiples, combinaciones lineales. Inferencia sobre predicciones y cálculos de intervalos de confianza.

Objetivos específicos:

9

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CB9. 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.

Dedicación: 3h

Grupo grande/Teoría: 1h

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h

Estimación de los parámetros e interpretación de los modelos lineales de respuesta polinómica nominal y ordinal

Descripción:

Estimación de máxima verosimilitud. Modelado nominal versus ordinal. Funciones de enlace utilizadas. Propiedades de los estimadores. Procesos inferenciales involucrados.

Objetivos específicos:

10

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 4h

Grupo grande/Teoría: 1h

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 2h

Validation of nominal and ordinal polytomous response models. Identification of unusual and influential data

Descripción:

Residuos de la devianza, Pearson. Residuos estudentitzats. Indicadores de datos inusuales e influyentes, mediante la extensión de los indicadores empleados en la regresión normal.

Objetivos específicos:

11

Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 2h 30m

Grupo grande/Teoría: 0h 30m

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h

Inferencia de hipótesis sobre parámetros simples y múltiples en modelos de respuesta politómica nominal y ordinal

Descripción:

Inferencia sobre estimadores de parámetros en modelos lineales de respuesta politómica. Intervalos de confianza. Contrastes de hipótesis simples, múltiples, combinaciones lineales. Inferencia sobre predicciones y cálculos de intervalos de confianza.

Objetivos específicos:

12

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 3h

Grupo grande/Teoría: 1h

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h



Estimación de los parámetros e interpretación de los modelos lineales para conteos

Descripción:

Estimación de máxima verosimilitud. Modelado Poisson, binomial negativo. Sobredispersión. Funciones de enlace utilizadas. Procesos inferenciales involucrados.

Objetivos específicos:

13

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

Dedicación: 2h 30m

Grupo grande/Teoría: 0h 30m

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h

Validación de los modelos de por conteos. Identificación de datos inusuales e influyentes. Tipología de los residuos. Diagnóstico sobredispersión. Modelos probabilísticos paramétricos

Descripción:

Indicadores de datos inusuales e influyentes. Comprobación de sobredispersión. Cómo superar la dispersión excesiva.

Objetivos específicos:

14

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 2h 30m

Grupo grande/Teoría: 0h 30m

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h



Inferencia de hipótesis sobre parámetros simples y múltiples en modelos para conteos

Descripción:

Inferencia sobre estimadores de parámetros en modelos lineales para conteos. Intervalos de confianza. Contrastes de hipótesis simples, múltiples, combinaciones lineales. Inferencia sobre predicciones y cálculos de intervalos de confianza.

Objetivos específicos:

15

Competencias relacionadas:

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 2h 30m

Grupo grande/Teoría: 0h 30m

Grupo pequeño/Laboratorio: 1h

Aprendizaje autónomo: 1h

Teoría y práctica del diseño de experimentos factorial y factorial fraccional

Objetivos específicos:

16

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

CB6. 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.

Dedicación: 8h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 4h



Examen Parcial

Objetivos específicos:

1, 2, 3, 4, 5, 6

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

CB6. 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.

Dedicación: 8h

Actividades dirigidas: 2h

Aprendizaje autónomo: 6h

Examen Final

Objetivos específicos:

7, 8, 9, 10, 11, 12, 13, 14, 15, 16

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

CB6. 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.

Dedicación: 10h

Actividades dirigidas: 2h

Aprendizaje autónomo: 8h

Práctica de Modelos Lineales

Objetivos específicos:

2, 3, 4, 5, 6

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. 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.

Dedicación: 20h

Aprendizaje autónomo: 20h



Práctica de Modelos Lineales Generalizados

Objetivos específicos:

7, 8, 9, 10, 11, 12, 13, 14, 15

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. 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 la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB9. 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.

CB6. 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.

Dedicación: 20h

Aprendizaje autónomo: 20h

SISTEMA DE CALIFICACIÓN

La evaluación de la asignatura integra las tres fases de aprendizaje descritas: conocimientos, destrezas y competencias.

Los conocimientos se evalúan mediante dos exámenes realizados a la mitad (T1, peso 1/3) y en la semana de exámenes finales del curso (T2, peso 2/3). En caso de suspender el examen parcial, el alumno podrá repetirlo como extensión del examen final. (Nota T).

Las destrezas se evalúan a partir de la entrega de 2 prácticas, Así mismo como las competencias transversales. Cada uno de los bloques 1, 2 y 3 para la primera práctica (P1) y 4 y 5 para la segunda (P2) comportarán una práctica que el alumno deberá realizar individualmente o en grupos de 2. El promedio de las notas da la nota P.

La Nota Final (NF) se calcula:

Examen Parcial (T1, 1/3) y Examen Final (T2, 2/3).

PRACT 1 (P1) y Práctica 2 (P2)

P: Nota de Prácticas $P = (P1 + P2) / 2$.

T: Nota Teoría = $\text{Max}(T2, (T1 + 2T2) / 3)$.

NF: Nota Final = $0.6T + 0.4P$.



BIBLIOGRAFÍA

Básica:

- Fox, J.; Weisberg, S. An R companion to applied regression. 3rd ed. Thousand Oaks, California: SAGE Publications, Inc, 2019. ISBN 9781544336473.
- Wickham, H. Ggplot2: elegant graphics for data analysis. 2nd ed. Cham: Springer, 2016. ISBN 9783319242774.
- Montgomery, D. Design and Analysis of Experiments. 10th ed. Hoboken, N.J: Wiley, 2020. ISBN 9781119722106.
- Box, George E. P; Hunter, J. Stuart; Hunter, William Gordon. Statistics for experimenters : design, innovation, and discovery. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2005. ISBN 9780471718130.
- Fox, John. Applied regression analysis and generalized linear models. 3rd edition. Los Angeles: SAGE, 2016. ISBN 9781452205663.

Complementaria:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, cop. 2009. ISBN 9780387952840.
- Trivedi, K.S,. Probability and statistics with reliability, queuing and computer science applications. 2nd ed. New York [etc.]: John Wiley and Sons, 2016. ISBN 1119285429.
- Mendenhall, W.; Wackerly, D.; Scheaffer, R. Mathematical Statistics with applications. Belmont, CA: Thomson Brooks/Cole, 2008. ISBN 9780495110811.

RECURSOS

Enlace web:

- <https://www.r-project.org/>- <https://www.rstudio.com/>