

MÁSTER UNIVERSITARIO EN ESTUDIOS AVANZADOS EN BOTÁNICA

Obtención y Análisis Avanzado de Datos	
Código	610632
Módulo	Obligatorio
Materia	Plantas en un mundo cambiante: cambio global y ecología vegetal
Carácter	Asignatura obligatoria
Créditos ECTS	6
Curso	Primero
Semestre	primero
Profesor responsable	Miguel Berdugo Vega, miberdug@ucm.es
Otros profesores	Francisco Cabezas Fuentes, Alberto López Teixido, Jesús Rojo Úbeda, David Sánchez Pescador, Jorge Romero Morte

SINOPSIS

DESCRIPTOR

Esta asignatura capacitará al estudiantado para diseñar los muestreos y experimentos necesarios que permitan testar una hipótesis científica dada en el ámbito de la botánica. El estudiantado conocerá las principales fuentes de información en botánica en la era del BigData y la Inteligencia Artificial, desde las de colecciones científicas hasta las que permiten el uso de datos de satélite. Finalmente, el estudiantado estará capacitado para reconocer, seleccionar, realizar e interpretar los resultados de análisis estadístico inferencial de nivel avanzado en botánica y ecología vegetal.

CONOCIMIENTOS PREVIOS RECOMENDADOS

Aunque la asignatura cubre un abanico suficientemente amplio de conceptos (desde los más básicos a los más complejos) como para que los estudiantes puedan cursarla, estos se beneficiarían de conocimientos previos básicos sobre estadística (distribuciones y contraste de hipótesis simples), diseño experimental y manejo básico de herramientas de programación y Sistemas de Información Geográfica.

OBJETIVOS FORMATIVOS

La asignatura está configurada en dos partes. La primera pretende ofrecer al estudiantado un conocimiento detallado sobre la obtención de datos en campo y la existencia, importancia, acceso y manipulación de bases de datos de interés en botánica. Aquí se emplearán herramientas de libre acceso como QGIS o Google Earth Engine y repositorios vinculados a colecciones científicas. El objetivo es dotar al estudiantado con la capacidad y herramientas para obtener datos científicos con el fin de testar cualquier hipótesis en botánica.

En la segunda parte se buscará que el estudiantado adquiera conocimientos avanzados de análisis estadístico en diversas técnicas. El objetivo aquí es tanto mostrar una gran variedad de técnicas (su interpretabilidad y sus detalles técnicos) como enseñar a diferenciar su utilidad de modo que, dado un diseño experimental y una hipótesis, el estudiante sepa cuál es la manera más conveniente de analizar los datos.

RESULTADOS DE APRENDIZAJE

Contenidos:

RA1 - Haber adquirido conocimientos avanzados sobre los principales aspectos teóricos, prácticos y metodológicos en el ámbito de la Botánica.

RA6 - Comprender los principales modelos estadísticos y aproximaciones computacionales necesarios para su aplicación en el análisis de datos y diseño y modelización de procesos en el ámbito de la Botánica.

RA9 - Conocer las principales fuentes de información científica y métodos de difusión en el ámbito de la Botánica.

RA10 - Conocer las nuevas tendencias y perspectivas de investigación en Botánica, así como los principios del método científico y sus condicionantes éticos.

Habilidades:

RA15 - Aplicar las herramientas experimentales y de análisis cualitativo y cuantitativo de los rasgos vegetativos y reproductivos en plantas.

RA17 - Aplicar técnicas estadísticas avanzadas en el estudio de la flora y vegetación

RA18 - Aplicar herramientas en la obtención, análisis e integración de información bibliográfica o bases de datos en el ámbito de la Botánica.

Competencias:

RA21 - Capacidad para diseñar, planificar y realizar el trabajo de laboratorio y de campo en Botánica dominando el uso de herramientas y software informático para la interpretación de los datos.

RA22 - Capacidad para valorar y discutir de forma crítica las distintas teorías, modelos o procesos implicados en la evolución de los vegetales y sus comunidades.

RA28 - Ser capaz de integrar la Botánica dentro de sus propias necesidades de desarrollo personal y entorno profesional.

RA29 - Ser capaz de desarrollar un trabajo de investigación o informe técnico o científico dentro del ámbito de la Botánica con un grado significativo de independencia y originalidad en un entorno laboral.

RA30 -Ser capaz de organizar sus propios caminos de aprendizaje de manera autónoma en el ámbito de la Botánica.

Ver descripción de resultados de aprendizaje en: [https:// www.ucm.es/mab/presentación](https://www.ucm.es/mab/presentación).

METODOLOGÍA DOCENTE

Para la primera parte en obtención de datos el estudiantado recibirá clases de índole teórico-práctico sobre diseño experimental, así como distintas técnicas de muestreo en campo. Después serán introducidas bases de datos de carácter general y libre disposición, como las de las colecciones científicas, las bases de datos de rasgos funcionales de plantas (e.g., TRY), inventarios forestales, las bases de datos sobre diversidad e inventarios vegetales (GBIF, SIVIM, AFLIBER), inventarios nacionales, entre otras. También se le enseñará al estudiantado a explorar la información contenida en mapas y la que puede conseguirse con teledetección, explorando la funcionalidad de los sistemas de información geográfica (SIG) como QGIS y de Google Earth Engine.

La segunda parte se compone de sesiones eminentemente prácticas en las que se exploran diversas técnicas de análisis estadístico inferencial. Las clases están diseñadas para que el estudiantado adquiera conocimientos avanzados en el lenguaje de programación de R y para que vaya incrementando poco a poco la complejidad de las técnicas de análisis hasta cubrir un

amplio abanico de las más usadas en estudios botánicos (desde modelos lineales a modelos de inteligencia artificial). En estas clases el estudiantado adquirirá los fundamentos teóricos de los análisis, aprenderá a interpretarlos y saber bajo qué supuestos usar cada análisis, alcanzando independencia y razonamiento crítico sobre los distintos tipos de análisis estadístico.

CONTENIDO TEMÁTICO

PROGRAMA TEÓRICO-PRÁCTICO

BLOQUE I. OBTENCIÓN DE DATOS EN BOTÁNICA

1. Diseño de muestreos y experimentos.
2. Sesgos y errores comunes de muestreo.
3. Métodos de muestreo en botánica.
4. Data Mining I: Recursos en botánica. Inventarios forestales; colecciones científicas. Bases de datos en botánica de interés global (e.g., TRY, SPLOT, GBIF).
5. Data Mining II: Cartografía y Sistemas de Información Geográfica (GIS). Fundamentos de GIS. Bases de datos cartográficas de interés: climáticas (e.g., ERA; Worldclim, TERRACLIMATE); hábitats y tipos de vegetación (e.g., CORINE; CGLS; Directiva Hábitats).
6. Data Mining III: Teledetección. Fundamentos de teledetección. Principales fuentes de datos: LANDSAT, MODIS, SENTINEL, EAN. Técnicas de teledetección: LiDAR, NDVI, entre otras. Fundamentos de Google Earth Engine.

BLOQUE II. ANÁLISIS ESTADÍSTICO

7. Fundamentos de R. Introducción, Lenguaje, RStudio, Objetos de R, operadores, Funciones básicas, Librerías. Manipulación de bases de datos (funciones de fusión: merge, apply, tapply, de ciclo y condicionales: for, if, ifelse, while...). Universo tidyverse.
8. Ordenación y clasificación. Algoritmos de clasificación por k-means. Análisis de componentes principales (PCA). Árboles filogenéticos y sistemática.
9. Correlación y regresión. Modelos lineales (ANOVA, ANCOVA y regresiones). Generalización de modelos (GLMs y extensiones, GLZz). Obtención de parámetros: Bayesiano vs frecuentista. Calidad del modelo y bondad de ajuste. Selección de modelos por criterio de Akaike (AIC).
10. Violación de asunciones y su solución. Ortogonalidad (medida con variance inflation factor; multicolinealidad), normalidad (estadística no paramétrica), puntos influyentes (regresiones robustas), no linealidad (modelos generalizados aditivos; GAMs; regresión local; LOESS), pseudo-replicación y medidas repetidas (Modelos generalizados mixtos y uso de random factors; GLMMs).
11. Análisis espacial y temporal. Autocorrelación temporal y espacial (I de Moran; semivariogramas, modelos ARIMA). Análisis de patrones fenológicos con transformada de Fourier. Análisis de espectros espaciales, descomposición estacional y de tendencia de series temporales.
12. Inferencia causal. Modelos de ecuaciones estructurales (SEM) y grafos acíclicos dirigidos (DAGs). Efectos directos e indirectos.
13. Modelos de inteligencia artificial. Fundamentos de machine learning. Random forest, Gradient boosting machines y Neural networks.

ACTIVIDADES DOCENTES

Actividad	Número de horas	% respecto presencialidad
Clases teóricas	12	25
Clases prácticas	30	62,5
Prácticas de Campo	6	12,5
Estudio autónomo	102	

SISTEMA DE EVALUACIÓN

La distribución porcentual de la nota final de la asignatura será:

Examen Teórico

- Examen teórico (30%), en el que se evalúa los contenidos teóricos de la asignatura.

Evaluación Continua

- Resolución de ejercicios y cuestionarios en las sesiones prácticas (30%)
- Memorias de prácticas (30%) que se evaluará realizando un visu estadístico basado en los apuntes de la memoria de prácticas.
- Evaluación de la actitud y participación en clase del estudiante (10%)

CRITERIOS DE EVALUACIÓN

Para superar la asignatura es imprescindible obtener un 5 de media entre todos los apartados, así como aprobar todos por separado.

RECURSOS

BIBLIOGRAFÍA BÁSICA

- Zuur, A., Ieno, E.N., Meesters, E. 2009. A beginner's guide to R. Springer ed.
- Bali, R., Sarkar, D. R. 2016. Machine learnign by example Understand the fundamentals of machine learning with R and build your own dynamic algorithms to tackle complicated real-world problems successfully. PAKT publishing ed.
- Maindonald, J., Braun, J. 2003. Data Analysis and Graphics using R. Cambridge University Press.
- Martínez Vega, J., & Martín Isabel, M. P. (Eds.). (2010). Guía didáctica de teledetección y medio ambiente. Red Nacional de Teledetección Ambiental; Centro de Ciencias Humanas y Sociales (CSIC).
- Quinn, GP., Keough, M.J. 2023. Experimental design and data analysis for biologists. 2nd Edition. Cambridge University Press.

OTROS RECURSOS

Recursos en el campus virtual.



- Easy web publishing from R <https://rpubs.com/>
- Write HTML, PDF, ePub, and Kindle books with R Markdown <https://bookdown.org/>
- Sistemas de Información Geográfica <https://volaya.github.io/libro-sig/index.html>