

ANÁLISIS DE DATOS

CICLO

SEMESTRE 1

CLAVE DE LA ASIGNATURA

C16ADD1

OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA

El objetivo del curso es cubrir la base de la probabilidad, la estadística matemática y la ciencia de datos. En primer lugar se trata de aprender a usar los diferentes conceptos, aprender a reconocerlos en problemas de análisis de datos, y aprender a dominar herramientas matemáticas y algorítmicas para solucionar estos problemas.

Enseñamos y usamos el lenguaje R como plataforma principal para el análisis de datos.

TEMAS Y SUBTEMAS

Parte 1: Probabilidad

- **Conceptos básicos:** probabilidades, variables aleatorias, distribuciones, momentos, independencia, regla de Bayes, variables multidimensionales, transformaciones de variables aleatorias, medidas de dependencia, teoremas de límite, simulación; aplicaciones ilustrativas en computación matemática.

Parte 2: Análisis de datos

- **Panorama general:** conceptos generales de ciencia de datos, minería de datos y reconocimiento estadístico de patrones.
- **Técnicas básicas de exploración y visualización de datos.** Errores más comunes, ejemplos de herramientas informáticas para el análisis de datos (dashboards, notebooks, etc.).
- **Introducción a la inferencia estadística:** pruebas de hipótesis, intervalos de confianza y estimaciones puntuales, métodos computacionalmente intensivos.



- **Modelación estadística y modelación predictiva;** introducción a la regresión lineal; ejemplos de métodos de predicción y clasificación.

ACTIVIDADES DE APRENDIZAJE

Cursos presenciales.

Resolución de ejercicios (de manera individual y en sesiones semanales de ejercicios).

Desarrollo de mini proyectos.

Lectura de artículos de divulgación.

CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN Y ACREDITACIÓN

Tareas semanales (40%)

Tres exámenes parciales (40%)

Dos tareas-proyectos (20%)

BIBLIOGRAFÍA

J. Van Horebeek, *Introducción a Métodos Estocásticos para Ciencias de la Computación*, Notas propias CIMAT, 2018.

F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa, L.E. Meester. *A Modern Introduction to Probability and Statistics*. Springer, 2005.

S. Skiena, *The Data Science Design Manual*, Springer, 2017.

C. Grinstead & L. Snell, *Introduction to Probability*, American Mathematical Society, 1998.

G. James, D. Witten, T. Hastie, R. Tibshirani, *Introduction to Statistical Learning*, Springer, 2017.

D. Spiegelhalter, *The Art of Statistics: How to Learn from Data*, Basic Books, 2019.

C.Wilke, *Fundamentals of Data Visualization*, O'Reilly Media, 2019.