



CA-506 ANÁLISIS DE DATOS
Carta al Estudiante

- NOMBRE DEL CURSO: ANALISIS DE DATOS
- MODALIDAD DEL CURSO: Teórico-Práctico
- NUMEROS DE CREDITOS: 4
- NUMEROS DE HORAS SEMANALES: 5
- REQUISITOS: CA-0403
- PROFESOR: Dr. Oldemar Rodríguez R.
- Aula virtual del curso en: <https://piazza.com>
- Sitio WEB y correo: www.oldemarrodriguez.com info@oldemarrodriguez.com
- JUSTIFICACION: El análisis multivariado de datos es una de las herramientas estadísticas más importantes debido a su capacidad para sintetizar la información proveniente de observaciones complejas. No es necesario hacer hipótesis a priori de la distribución de las variables observadas, siguiendo el punto de vista de la Escuela Francesa de Analyse des Données.
- OBJETIVOS:
Que el estudiante conozca y domine las principales técnicas de base del análisis multivariado de datos, que distinga su aplicabilidad en las distintas situaciones y que aprenda a interpretar los resultados.
- CONTENIDOS:
 1. Programación en Lenguaje R
 - Uso de R como paquete de análisis
 - Programación en R como lenguaje funcional
 2. Introducción al análisis multidimensional de datos
 - Métrica de pesos
 - Proyecciones ortogonales
 - Varianza total y residual
 - Regresión Lineal
 3. Análisis en componentes principales - ACP
 - Aplicación e interpretación de un ACP
 - Algoritmo e implementación del ACP
 - Presentación geométrica y algebraica del ACP
 4. Escalamiento Multidimensional – Mutidimensional Scaling
 5. Análisis de Tablas Múltiples
 - Método STATIS
 6. Clasificación Automática



- Agrupación Jerárquica
- Método K-Medias
- 7. Análisis Discriminante
 - Método factorial
 - Método bayesiano
- 8. Análisis de correspondencias simples
 - Nubes de perfiles; solución; propiedades; ejemplos
- 9. Análisis de correspondencias múltiples
 - Tabla disyuntiva completa; solución; propiedades; ejemplos
- 10. Regresión
 - Regresión Lineal Generalizada, Regresión Ridge, Regresión LASSO y Redes Elásticas

- ACTIVIDADES: El curso se desarrollará mediante clases magistrales y tareas para trabajar en la casa.

- EVALUACION:

Dos exámenes	60%
Tareas	40%

- BIBLIOGRAFIA DE REFERENCIA:

1. Benzécri, J.P. (1973) Analyse des Données. Tome II: Correspondances. Dunod, Paris.
2. Bry X. "Analyses factorielles simples", Ed. Economica, Paris, 1995.
3. Gailliez, F.; Pagés, J.P. (1976) Introduction à l'Analyse des Données. SMASH, Paris.
4. Chatfield, Collins, A. (1988) Multivariate Analysis. Chapman & Hall. London.
5. Diday, E.; Lemaire, J.; Pouget, J.; Testu, F. (1982) Eléments d'Analyse des Données. Dunod. Paris.
6. Escofier, B.; Pagés, J. (1988) Analyse Factorielles Simples et Multiples. Dunod, Paris.
7. Gareth James, Daniela Witten, Trevor Hastie y Robert Tibshirani. "An Introduction to Statistical Learning: with Applications in R". Springer Texts in Statistics, 2015.
8. Greenacre, M. (1984) Theory and Applications of Correspondence Analysis. Academic Press, New York.
9. Jambu, M. (1989) Exploration Informatique et Statistique des Données. Dunod, Paris.
10. Jambu M. "Introduction au Data Mining: Analyse Intelligente des données". Eyrolles, Paris, 1999.
11. Lebart, L.; Morineau, A.; Piron, M. (1995) Statistique Exploratoire Multidimensionnelle. Dunod, Paris
12. R Development Core Team. "R: A Programming Environment for Data Analysis and Graphics". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>



13. R Development Core Team. "Writing R Extensions". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
14. Trejos, J. (1996) Introducción al Análisis de Datos. Notas de Curso, U.C.R.