



**XS-4110: INTRODUCCION AL ANALISIS MULTIVARIADO
PROGRAMA
I SEMESTRE 2019**

| | | |
|----------------------------|--|------------------------|
| Docente: | Ricardo Alvarado Barrantes. | Samuel Muñoz Trigueros |
| Oficina: | 17 Estadística | Cubículos profesores |
| Teléfono | 2511-6529 / 8402-1263 | 8706-8589 |
| Correo electrónico: | estad.ucr@gmail.com | samutriz@gmail.com |
| Horas de consulta: | K: 3:00-5:00/M:9:00-11:00 | M: 5:00-7:00pm |
| Horario de clases: | M: 1:00-4:50pm | M: 1:00-4:50pm |

1. Descripción

Curso introductorio de técnicas estadísticas multivariados para estudiantes de cuarto año del Bachillerato de Estadística, impartido con un enfoque teórico-práctico. Además de los conocimientos teóricos se brindará al estudiante la posibilidad de aplicar los métodos mediante el uso de lenguaje estadístico R.

- **Requisitos:** **XS-3310 Teoría Estadística y XS-2130 Modelos de Regresión Aplicados**
- **Correquisitos:** XS-4410 Práctica Profesional I
- **Horas:** 4 horas semanales (2 de teoría y 2 de práctica)
- **Créditos:** 4

2. Objetivo General

Ofrecer una visión general de las técnicas básicas, gráficas y cuantitativas, del análisis multivariado que involucra varias variables y múltiples casos, e ilustrar sus aplicaciones con datos provenientes de nuestro medio y de revistas científicas.

3. Objetivos Específicos

Al finalizar el curso el estudiante tendrá criterio y conocimiento básico para:

- Resolver problemas con las técnicas estadísticas multivariados básicas: análisis de componentes principales, análisis factorial, análisis discriminante, escalamiento multidimensional y análisis de agrupamientos.
- Reconocer situaciones donde se puedan aplicar las técnicas aprendidas.
- Verificar los supuestos en que se apoyan las técnicas antes de aplicarlas.
- Evaluar las bondades y limitaciones de las técnicas.
- Procesar datos multivariados utilizando software estadístico, e interpretar los resultados obtenidos.





4. Contenidos

| | |
|---|--|
| I. Visualización de datos multivariados – 2 clases | |
| 1.1 | Representación multidimensional: matrices de gráficos, colores, gráficos tridimensionales, contornos, shingles |
| 1.2 | Representación de correlaciones: correlograma |
| 1.3 | Detección de valores extremos multivariados |
| II. Análisis de componentes principales (PCA) – 2 clases | |
| 2.1 | Objetivos del PCA |
| 2.2 | Características de los componentes principales |
| 2.3 | Construcción de los componentes principales |
| 2.4 | Uso de covariancias o correlaciones |
| 2.5 | Cálculo de los puntajes en los componentes principales |
| 2.6 | Variación explicada |
| 2.7 | Representación gráfica: biplot |
| 2.8 | Número de componentes principales |
| 2.9 | Evaluación de resultados: 1) reproducción de matriz de variancias, 2) correlación entre componentes y variables originales. |
| III. Análisis de agrupamientos (clústers) – 2 clases | |
| 3.1 | Objetivos del análisis de agrupamientos |
| 3.2 | Selección de variables para el análisis |
| 3.3 | Similitud entre objetos: 1) distancias para variables continuas, 2) distancias para variables binarias, 3) distancia para mezcla de variables. |
| 3.4 | Estandarización |
| 3.5 | Distancias entre grupos |
| 3.6 | Agrupamientos jerárquicos: algoritmo y representación (dendograma) |
| 3.7 | Método de k-medias: algoritmo y selección del número de clústers. |
| 3.8 | Presentación de resultados |





IV. Clasificación – 4 clases

- 4.1 Análisis discriminante
- 4.2 Regresión logística binomial y multinomial
- 4.3 Árboles de decisión
- 4.4 K-vecinos más cercanos
- 4.5 Máquinas de vectores de soporte
- 4.6 Redes neuronales
- 4.7 Validación
- 4.8 Métricas de desempeño

V. Ensamblados de modelos – 2 clases

- 5.1 Agregación de bootstrap
- 5.2 Bosques aleatorios
- 5.3 Boosting
- 5.4 Stacking

5. Metodología

- El curso es teórico-práctico y exige el uso frecuente de la computadora.
- Se espera no sólo que el estudiante aprenda los fundamentos teóricos de las técnicas multivariantes, sino que también aplique las técnicas a archivos de datos utilizando paquetes estadísticos.
- Las lecciones son magistrales, con aplicaciones prácticas utilizando computadoras de la Escuela de Estadística o del propio estudiante.
- Para reforzar la asimilación de conceptos y la interpretación de los resultados obtenidos de los paquetes estadísticos, se dejarán ejercicios que el estudiante debe tratar de resolver en casa.
- Se dejarán algunas lecturas de artículos para familiarizar a los estudiantes con aplicaciones prácticas de las técnicas multivariadas.
- Los estudiantes deberán realizar análisis de datos reales que serán presentados en forma de artículos científicos.





6. Evaluación

- Se realizarán dos exámenes parciales, en ellos se evaluarán conceptos y la forma de interpretar resultados.
- Se realizará un examen práctico llamado "reto predictivo", en el cual cada estudiante deberá analizar una base de datos diferentes y presentar sus resultados en un tiempo determinado.
- Los estudiantes presentarán dos trabajos de análisis de datos reales. Los trabajos deberán presentarse en forma de artículos cortos y deberán ser presentados oralmente.
- Cada estudiante deberá realizar una presentación oral de visualización de datos.
- Si un estudiante faltase a algún examen por **causa justificada**, debe solicitar por escrito la reposición del examen indicando las razones de la ausencia, acompañada de los documentos justificantes. La misma debe entregarse ante el profesor que imparte el curso **a más tardar en cinco días hábiles del reintegro a lecciones.**

| | |
|---------------------------------|-----|
| Primer examen teórico | 15% |
| Segundo examen teórico | 15% |
| Tercer examen práctico (reto) | 20% |
| Primer artículo y presentación | 15% |
| Segundo artículo y presentación | 20% |
| Presentación individual | 15% |

7. Cronograma

| | M | Actividad |
|-----|----|-----------------------------|
| MAR | 13 | Visualización |
| | 20 | PCA |
| | 27 | Jornadas + PCA |
| ABR | 3 | Agrupamientos |
| | 10 | Presentación Visualización |
| | 17 | SEMANA SANTA |
| | 24 | Agrupamientos |
| MAY | 1 | FERIADO |
| | 8 | Clasificación |
| | 15 | Presentación Artículo 1 |
| | 22 | Examen No.1 + Clasificación |
| | 29 | Clasificación |
| JUN | 5 | Clasificación |
| | 12 | Ensamblaje de modelos |
| | 19 | Ensamblaje de modelos |
| | 26 | Presentación Artículo 2 |
| JUL | 3 | Examen No.2 |
| | 10 | Examen No.3 |



8. Referencias bibliográficas

- Cichosz, Pawel. (2015). Data Mining Algorithms: Explained Using R. Wiley.
- Everitt, B y Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R. Springer
BIBLIOTECA LUIS DEMETRIO TINOCO 519.535.028.5 E93i
- Hair, J.F. et al (2014). Multivariate Data Analysis. Pearson Education Limited. **BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 M958m7 2015**
- Hernández R, Óscar (1998). Temas de Análisis Estadístico Multivariado. Editorial UCR.
BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 H557t
- Hernández R, Óscar (2006). Notas adicionales a Temas de Análisis Estadístico Multivariado.
- Johnson, R. A. y Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice-Hall International, Inc. **BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 J68a6**
- Johnson, D. (1998). Métodos multivariados aplicados al análisis de datos. International Thompson Editores.
- Kleinbaum et al. (1998). Applied Regression Analysis and other Multivariate Methods. Duxbury Press.
- Mirkin, B (2005). Clustering for Data Mining: A Data Recovery Approach. Chapman & Hall.
- Mishra, P. (2016). R Data Mining Blueprints. Packt Publishing.
- Pan et al. (2013). Introduction to Data Mining. Pearson.
- Pla, E.L. (1986). Análisis Multivariado: Método de componentes principales. O.E.A. Washington.
(Cap 4: Caracterización de la producción lechera de un distrito). **BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 P696a**
- Olson et al. (2017). Predictive Data Mining Models. Springer.
- Ramasubramanian, K y Singh, A (2017). Machine Learning Using R Apress.
- Sarkar, D (2008). Lattice: Multivariate Data Visualization with R. Springer. **BIBLIOTECA LUIS DEMETRIO TINOCO 006.6 S245L**