



XS-41100: INTRODUCCION AL ANALISIS MULTIVARIADO
PROGRAMA
II SEMESTRE 2018

Docente: Ricardo Alvarado Barrantes.

Oficina: 17 Estadística.

Correo electrónico: estad.ucr@gmail.com

Teléfono: 2511-6529 / 8402-1263

Horario de clases: L: 1:00-4:50pm

Horas de consulta: K: 1:00-5:00pm

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 paquetes estadísticos (principalmente 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	
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)	
2.1	Elementos de álgebra lineal
2.2	Objetivos del PCA
2.3	Características de los componentes principales
2.4	Construcción de los componentes principales
2.5	Uso de covariancias o correlaciones
2.6	Cálculo de los puntajes en los componentes principales
2.7	Variación explicada
2.8	Representación gráfica: biplot
2.9	Número de componentes principales
2.10	Evaluación de resultados: 1) reproducción de matriz de variancias, 2) correlación entre componentes y variables originales. El tamaño de muestra
III. Análisis factorial	
3.1	Modelos de factores comunes, modelos ortogonales
3.2	Métodos de estimación: 1) componentes principales, 2) máxima verosimilitud
3.3	Selección del número de factores
3.4	Rotación de factores
3.5	Estimación de puntajes factoriales
3.6	Análisis factorial confirmatorio
IV. Análisis de agrupamientos (clústers)	
4.1	Objetivos del análisis de agrupamientos
4.2	Selección de variables para el análisis
4.3	Similitud entre objetos: 1) distancias para variables continuas, 2) distancias para variables binarias, 3) distancia para mezcla de variables.
4.4	Estandarización
4.5	Distancias entre grupos
4.6	Agrupamientos jerárquicos: algoritmo y representación (dendograma)
4.7	Método de k-medias: algoritmo y selección del número de clústers.
4.8	Presentación de resultados
4.9	Evaluación interna y externa





V. Clasificación

- 5.1 Separación y clasificación
- 5.2 Regla basada en la minimización del costo esperado de clasificación errónea
- 5.3 Procedimiento para poblaciones multinormales con matrices de covariancias iguales
- 5.4 Evaluación de la eficacia de las reglas de clasificación
- 5.5 Función discriminante de Fisher
- 5.6 Regresión logística binomial y multinomial

VI. Árboles de decisión (MDS)

- 6.1 Tipos de árboles de decisión
- 6.2 Medidas de decisión
- 6.3 Métodos de aprendizaje
- 6.4 Árboles de conjunto

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 multivariados.
- Los estudiantes deberán realizar análisis de datos reales que serán presentados en forma de artículos científicos.





6. Evaluación

- Se realizarán tres exámenes parciales, en ellos se evaluarán los conceptos, el uso del software para obtener resultados y la forma de interpretar resultados. Los exámenes se realizarán en el laboratorio de computadoras y los estudiantes deberán conocer el manejo del software para obtener los resultados.
- Los estudiantes presentarán varios trabajos de análisis de datos reales. Los trabajos deberán presentarse en forma de artículos cortos.
- En grupos se deberán hacer dos presentaciones cortas sobre el análisis realizado en los trabajos.
- 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 parcial	20%
Segundo examen parcial	20%
Tercer examen parcial	20%
Artículos	40%

7. Cronograma

	Módulo	L	Actividad
MAR	I. Visualización	12	
		19	
		26	SEMANA SANTA
ABR	II. PCA	2	
		9	Artículo Visualización
	III. Factorial	16	
		23	
		30	FERIADO
MAY	IV. Agrupamientos	7	
		14	Artículo Dimensionalidad
		21	Examen No.1
JUN	V. Clasificación	28	
		4	Artículo Agrupamiento
JUL	VI. Árboles de decisión	11	Examen No.2
		18	
		25	Artículo Clasificación
		2	Examen No.3





8. Referencias bibliográficas

- 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.
- 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**
- 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**