



Universidad de Costa Rica
Sistema de Estudios de Posgrado
Programa de Posgrado en Estadística

NP-1602 Introducción al Análisis Multivariado

INFORMACIÓN GENERAL

Plan de estudios al que pertenece el curso:	730501 Maestría Académica en Estadística y 730506 Maestría Profesional en Estadística
Tipo de curso:	Nivelatorio para la Maestría Académica en Estadística y Maestría Profesional en Estadística
Modalidad:	Teórico
Número de créditos:	4 créditos
Horas presenciales:	4 horas semanales - K: 17:00 - 20:50
Horas de atención de estudiantes:	2 horas semanales - V: 17:00 - 18:50
Requisitos:	No tiene
Correquisitos:	No tiene
Profesor:	M.Sc. Oscar Centeno Mora



PROGRAMA

Descripción general del curso

El curso de Introducción al Análisis Multivariado está enfocado en el aprendizaje de los principales métodos de análisis para variables numéricas como no numéricas. El curso se centra en conocer la razón que justifica la aplicación de un determinado método, su objetivo, desarrollo e inmediata aplicación.

Justificación

Curso introductorio al análisis multivariado para estudiantes que ingresan a la Maestría Profesional o Académica en Estadística que provienen de disciplinas distintas a la Estadística o bien, que nunca hayan aprobado un curso similar previamente. Se pretende con este ofrecer el conocimiento mínimo necesario para nivelarse con los estudiantes que provienen de la carrera de Estadística. Con este curso, los estudiantes meta se integrarán exitosamente los cursos regulares del programa.

Objetivo general

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

Objetivos específicos

Al final del curso se espera que el estudiante haya aprendido a:

1. Resolver problemas con las técnicas estadísticas multivariantes básicas: análisis de componentes principales, análisis factorial, análisis discriminante, análisis multivariante de variancia y análisis de agrupamientos (“clusters”), .
2. Reconocer situaciones donde se puedan aplicar las técnicas aprendidas.
3. Verificar los supuestos en que se apoyan las técnicas antes de aplicarlas.
4. Evaluar las bondades y limitaciones de las técnicas.
5. Procesar datos multivariantes utilizando software estadístico e interpretar los resultados obtenidos.
6. Presentar los resultados obtenidos.
7. Aplicar un trabajo final en relación con una de las técnicas estudiadas en el curso



Metodología

El curso es teórico-práctico y exige el uso frecuente de la computadora. Se espera no solo 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 ilustraciones prácticas relacionadas en el laboratorio de cómputo de la Escuela de Estadística. 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 multivariantes.

Al ser un curso interactivo entre alumno-profesor, se solicita la participación de las y los estudiante para mejorar la calidad del curso y del aprendizaje conjunto de los participantes.

Presencia obligatoria

Se recuerda la presencia obligatoria a todas las clases. Si por razones mayores el estudiante no puede asistir, es imperativo justificar la ausencia.

Evaluación

Se realizarán dos 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 computadora, y los estudiantes deberán conocer el manejo del software para obtener los resultados. Por otra parte, los estudiantes presentarán varios trabajos de análisis de datos reales. Los trabajos deberán presentarse en forma de artículos cortos.

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.

La evaluación del curso se desglosa según los siguientes rubros:

Parcial 1	25%
Parcial 2	25%
Presentaciones e informes de las técnicas multivariadas	40%
Trabajo Final	10%
TOTAL	100%



Descripción del curso

TEMAS Y CONTENIDOS	MATERIALES
Tema 1 Visualización de datos multivariados Representación multidimensional: matrices de gráficos, colores, gráficos tridimensionales, contornos, etc. Representación de correlaciones: correlograma Detección de valores extremos multivariados: boxplot bivariado, envoltorio convexo, chi-plot	PC
Tema 2 Análisis de componentes principales (PCA) Elementos de álgebra lineal Objetivos del PCA Características de los componentes principales Construcción de los componentes principales Uso de covariancias o correlaciones Cálculo de los puntajes en los componentes principales Variancia explicada Representación gráfica: biplot Número de componentes principales Evaluación de resultados: 1) reproducción de matriz de variancias, 2) correlación entre componentes y variables originales.	PC
Tema 3 Análisis Factorial Modelos de factores comunes, modelos ortogonales Métodos de estimación: 1) componentes principales, 2) máxima verosimilitud Selección del número de factores Rotación de factores Estimación de puntajes factoriales Pensamiento a priori: el análisis factorial confirmatorio	PC
Tema 4 Análisis discriminante Separación y clasificación Regla basada en la minimización del costo esperado de clasificación errónea Procedimiento para poblaciones multinormales con matrices de covariancias iguales Evaluación de la eficacia de las reglas de clasificación Función discriminante de Fisher	PC



Parcial n° 1 Temas 1 - 4	PC
Tema 5 Análisis de agrupamiento (clusters) Objetivo del análisis de agrupamiento Selección de variables para el análisis Similitud entre objetos: 1) distancias para variables continuas, 2) distancias para variables binarias, 3) distancia para mezcla de variables. Estandarización Distancias entre grupos Agrupamientos jerárquicos: algoritmo y representación (dendograma) Método de k-medias: algoritmo y selección del número de clústers. Presentación de resultados: 1) componentes principales, 2) sombras y vecindarios, 3) rayas, 4) perfiles Clusters para variables categóricas Relación entre PCA-Discriminación-Agrupamiento Objetivo de los métodos Supuestos y comportamiento de los datos La reducción, clasificación y la discriminación	PC
Tema 6 Escalamiento Multidimensional (MSD) Objetivos del MDS Variables noma Procedimiento: 1) métodos métricos, 2) métodos no-métricos Representación gráfica Análisis de correspondencias	PC
*Tema 7 Árboles de decisión El análisis supervisado vs el análisis descriptivo ¿Por qué los árboles de decisión? Clasificación y predicción Reglas de partición El método CART El método CHAID Evaluación del árbol: nodos y profundidad	PC
Parcial n° 2 Temas 5 - 7	PC
Entrega de notas y trabajos finales	



Cronograma

Mes	Tema	Día	Actividad
Marzo	Presentación	13	
	Visualización	20	
	---	27	Semana santa
Abril	PCA	3	
		10	
	Factores	17	
		24	
Mayo	Discriminante	1	
		8	
	Examen 1	15	
	Agrupamiento	22	
		29	
Junio	Escalamiento	5	
		12	
	Arboles	19	
		26	
Julio	Examen 2 / TF	3	
	Entrega de notas	10	



Bibliografía

La siguiente bibliografía es la recomendada para el curso:

- 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**
- Sarkar, D (2008). Lattice: Multivariate Data Visualization with R. Springer. **BIBLIOTECA LUIS DEMETRIO TINOCO 006.6 S245L**