



**UNIVERSIDAD DE COSTA RICA  
SISTEMAS DE ESTUDIO DE POSTGRADO  
PROGRAMA DE POSTGRADO EN  
ESTADÍSTICA**



<b>Materia Regular:</b>	Introducción al Análisis Multivariado
<b>Código:</b>	NP-1602
<b>Periodo:</b>	I 2017
<b>Grupo:</b>	001
<b>Profesor a cargo:</b>	Oscar Centeno Mora
<b>Correo electrónico:</b>	oscar.centenomora@ucr.ac.cr / oscarcenteno86@gmail.com
<b>Horas de teoría:</b>	2
<b>Horas de práctica:</b>	2
<b>Créditos:</b>	4
<b>Horario del curso:</b>	K: 17:00 - 20:50
<b>Horario de consulta:</b>	V: 17:00 - 18:50

#### **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**

El aprendizaje de los métodos multivariados es una destreza esencial para el profesional que desea aplicar la Estadística. De lo anterior, se resalta la importancia de recibir un curso donde se deba analizar información en abundancia, donde se desee llegar a un resumen de la misma y a una contestación del problema en causa. El conocer tanto los métodos, su desarrollo y aplicación es un requisito indispensable para el futuro profesional analista de la información.

#### **Objetivo General**

El objetivo general del curso es enseñar al estudiante los conceptos y principales aplicaciones de los análisis multivariados, además de brindarle los conocimientos necesarios para poder desenvolverse como profesional. El estudiante aprenderá tanto técnicas multivariadas para datos cualitativas, cuantitativos, así como de ambos tipos de variables. La modalidad práctica del curso permitirá que el estudiante pueda aplicar en cualquier contexto las técnicas del presente curso.

#### **Objetivos específicos**

El presente curso tiene por objetivos de:

- Estudiar los principales métodos de visualización multivariada.

- Aprender las principales técnicas multivariadas: análisis por componentes principales, el análisis factorial, el análisis discriminante, el análisis por agrupamiento, la relación entre reducción de variables – discriminación – agrupamiento, y el Escalamiento multidimensional
- Apoyar el estudio de las técnicas con artículos ilustrativos
- Aplicar las principales técnicas multivariadas
- Procesar datos multivariados utilizando software estadístico, e interpretar los resultados obtenidos
- Evaluar las bondades y limitaciones de las técnicas.
- Exponer y realizar pequeño informe de cada método.
- Aplicar un trabajo final en relación con una de las técnicas estudiadas en el curso.

### **Estrategias de enseñanza-aprendizaje (metodología)**

El presente curso se divide en dos modalidades: teoría y práctica. En una primera parte se desarrollará la teoría de cada uno de los temas, para luego pasar a apoyar ese conocimiento adquirido con la práctica. Con respecto a la parte práctica, en una primera parte el profesor entregará los ejemplos ilustrativos del método estudiado. Seguidamente los alumnos deberán aplicar y presentar la técnica multivariada expuesta en clase. 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. Los estudiantes deberán realizar análisis de datos reales que serán presentados en forma de artículos científicos.

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.

### **Evaluación del curso**

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%
<b>TOTAL</b>	<b>100%</b>

## 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.

## Contenido del curso

TEMAS Y CONTENIDOS	MATERIALES
<b>Tema 1 Visualización de datos multivariados</b> 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
<b>Tema 2 Análisis de componentes principales (PCA)</b> 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
<b>Tema 3 Análisis Factorial</b> 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
<b>Tema 4 Análisis discriminante</b> 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
<b>Parcial n° 1</b> <b>Temas 1 - 4</b>	PC

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

\*Optativo

## Cronograma

Mes	Tema	Día	Actividad
Marzo	Presentación	14	
	Visualización	21	
	PCA	28 4	
Abril		11	Semana santa
	Factores	18 25	
	Discriminante	2 9	
Mayo	Examen 1	16	
	Agrupamiento	23 30	
	Escalamiento	6 13	
Junio	Arboles	20 27	
	Examen 2 / TF	4	
	Entrega de notas	11	
		18	

## 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**