



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

SP-1628 Análisis Estadístico 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:	Obligatorio para la Maestría Académica en Estadística y Maestría Profesional en Estadística
Modalidad:	Bimodal (mediación virtual, https://udecr.zoom.us/j/89709609762)
Número de créditos:	4 créditos
Horas de clase:	4 horas semanales (M 17-20:50; 300B CE)
Horas de atención de estudiantes:	2 horas semanales (L 18-20 virtual)
Requisitos:	No tiene
Correquisitos:	No tiene
Profesora:	Shirley Rojas Salazar (shirleyelena.rojas@ucr.ac.cr ; ucrsrs@gmail.com)
Ciclo y año:	II ciclo del 2023

PROGRAMA

Justificación

El curso pretende proporcionar al estudiante las herramientas para aplicar el análisis multivariado en contextos diversos. Se plantea como un curso introductorio de Maestría, en el que se pretende brindar al estudiante técnicas más complejas de las que ya ha conocido en el grado o profundizar las ya conocidas, para que logre resolver las dificultades que se presentan a la hora de analizar datos en varias dimensiones.



Objetivo general

Desarrollar técnicas multivariantes no tratadas en los cursos de grado, o profundizar en algunas de ellas, prestando atención a los principios, supuestos, usos y aplicaciones con el uso del software estadístico pertinente.

Objetivos específicos

1. Aplicar diversas técnicas multivariantes a conjuntos de datos específicos para obtener resultados interpretables bajo un contexto dado.
2. Analizar el efecto de los supuestos y de los datos faltantes en la aplicación de técnicas multivariantes.
3. Escribir los resultados obtenidos con las técnicas multivariantes utilizadas bajo un contexto dado.

Descripción del curso

El curso consiste en la exposición de la teoría y la aplicación práctica del análisis multivariado con las siguientes técnicas:

- I. Construcción de índices a partir de encuestas. Marco conceptual, conceptos y definiciones operacionales, relaciones bivariadas y multivariadas, puntuación de los índices, tratamiento de valores faltantes, fórmula de cálculo, validación interna y externa.
- II. Análisis de factores exploratorio. El modelo lineal general de factores, generalidades sobre la estimación del modelo, las cargas factoriales, las comunalidades, la adecuación del modelo y la elección del número de factores, rotación ortogonal y oblicua, puntajes factoriales, ejemplos.
- III. Análisis factorial de datos binarios. Modelos de rasgos latentes, matriz de datos binarios, la invalidez del modelo factorial para variables métricas cuando se le aplica a respuestas binarias, el modelo factorial para datos binarios, el modelo de Rasch, el ajuste del modelo logito, interpretación de los parámetros del modelo, bondad de ajuste, puntajes factoriales, rotación, ejemplos.
- IV. Análisis factorial para variables categóricas ordenadas. El enfoque de la función de respuesta ítem: ajuste del modelo, bondad de ajuste, puntajes factoriales, ejemplos. El enfoque de la variable subyacente, su relación con el enfoque de la función de respuesta ítem, variables observadas ordenadas y parcialmente ordenadas, ejemplos.
- V. Árboles de decisión. Árboles de regresión y árboles de clasificación. Ajuste y poda de árboles.



- VI. Análisis de correspondencias. Objetivos, la matriz de datos, la realización de un análisis de correspondencia mediante un ejemplo (perfiles fila, perfiles promedio, inercia, representación bidimensional), el método general, residuos de Pearson, escalamiento dual, el biplot, ploteos simétricos y asimétricos, interpretación de las dimensiones, la elección del número de dimensiones.
- VII. Escalamiento multidimensional. Medidas de semejanza entre variables, escalamiento multidimensional clásico, ordinal y métrico, la evaluación del ajuste y la elección del número de dimensiones.
- VIII. Análisis discriminante. Definición, usos, supuestos, técnica, interpretación de los resultados. Distribución normal multivariada.
- IX. Regresión logística. El modelo múltiple, el ajuste, prueba de significancia del modelo, las razones de ventaja, los intervalos de confianza de los estimadores, tipos de variables, interacciones, selección de variables, el modelo politómico.
- X. Modelos de ecuaciones estructurales. Los modelos, notación y simbología, el modelo estructural, el modelo de medición, supuestos del modelo, etapas de construcción, pruebas de ajuste.

Metodología

Se impartirán clases magistrales, con la participación de los estudiantes. El curso es teórico- práctico y requiere el uso intensivo de la computadora. El procedimiento que se aplicará para desarrollar el curso se describe a continuación:

1. En las lecciones teóricas la profesora expone y resume los aspectos principales de las técnicas de análisis y revisa con los estudiantes los procedimientos en el computador. Para cada técnica explicada en clase, la profesora llevará un archivo de datos que utilizará con los estudiantes para practicar.
2. Después de conocer sobre una técnica específica, cada estudiante buscará un archivo de datos para aplicarla. El archivo de datos debe ser de una encuesta o un registro y debe ser conocido por el estudiante o debe adquirir ese conocimiento en el corto plazo. No se admiten archivos con datos ficticios o tomados de libros. Con este archivo, después de analizar cada técnica, el estudiante escribirá un artículo, entregará copia a la profesora por medio de la plataforma de mediación virtual, y realizará una presentación de los principales resultados ante el grupo que no exceda 10 minutos. Detalles como la estructura del artículo, redacción, ortografía, confección de cuadros y gráficos, y citas bibliográficas, serán especialmente tomados en cuenta para la calificación del trabajo realizado.

Otros elementos que también se tomarán en cuenta son la aplicación de la técnica a los datos presentados, la revisión de los supuestos para utilizar la técnica de análisis y las estrategias para manipular las variables del archivo de datos (uso de variables dummy, manejo de los missing values, etc).



3. Después de estudiar las técnicas multivariantes incluidas en el programa del curso, cada estudiante deberá especializarse en una técnica nueva: estudiarla, aplicarla, escribir un artículo y presentar los principales resultados ante el grupo. Además, durante la presentación, el estudiante debe proporcionar el archivo de datos y mostrar los procedimientos que utilizó. Los participantes del curso, como práctica, obtendrán los mismos resultados encontrados por el estudiante expositor. La duración de cada exposición será de común acuerdo con la profesora. La selección de la técnica que cada estudiante escogerá, así como la duración de cada exposición, se hará de mutuo acuerdo con la profesora. Algunos de los temas incluidos en el programa podrán ser desarrollados por los estudiantes como parte de este trabajo de especialización. Las técnicas que pueden escoger los estudiantes son diversas:
- Análisis de correspondencia multivariante
 - Análisis de componentes principales funcional
 - Análisis de correlación canónica
 - Regresión no lineal
 - Regresión cuantílica
 - Regresión segmentada
 - Análisis conjunto
 - Análisis de redes
 - Redes neuronales
 - Temas de minería de datos (incluyendo minería de texto) u otras técnicas no vistas en los cursos de la maestría.
4. Habrá un examen (presencial) al final del curso en el que se presentarán dos casos. Se debe responder a los objetivos asociados a estos casos utilizando las técnicas multivariantes vistas a lo largo del curso, incluyendo las que sean presentadas por los estudiantes en el trabajo de especialización.

Evaluación

Evaluación	Porcentaje
Artículo 1 - Análisis de Factores Exploratorio	10
Artículo 2 - Análisis de Correspondencias	10
Artículo 3 - Escalamiento Multidimensional	10
Artículo 4 - Análisis Discriminante y Regresión Logística Politémica	10
Artículo 5 - Trabajo de especialización (exposición y práctica)	30
Examen final	30



Cronograma

Fecha	Actividad
16-Ago	Revisión del programa. Revisión del documento sobre redacción de artículos.
23-Ago	Tema 1: Construcción de índices y práctica. Tema 2: Análisis de factores. Construcción de Índices o Escalas con Análisis de Factores y práctica.
30-Ago	Entrega y exposición de trabajos del tema de Análisis de Factores.
6-Set	Tema 3: Análisis de Correspondencias y práctica.
13-Set	Entrega y exposición de trabajos de Análisis de Correspondencias.
20-Set	Tema 4: Escalamiento Multidimensional y práctica
27-Set	Entrega y exposición de trabajo del tema de Escalamiento Multidimensional.
4-Oct	Tema 5: Análisis Discriminante, Regresión Logística Politémica y práctica
11-Oct	Entrega y exposición de trabajos de los temas de Análisis Discriminante y Regresión Logística.
18-Oct	Tema 6: Árboles de decisión con variables categóricas (tema de especialización de un estudiante). Presentación y práctica. Tema 7: Árboles de decisión con variables métricas (tema de especialización de un estudiante). Presentación y práctica.
25-Oct	Tema 8: Análisis de factores con datos binarios (tema de especialización de un estudiante). Presentación y práctica. Tema 9: Análisis de factores con variables ordinales (tema de especialización de un estudiante). Presentación y práctica.
1-Nov	Tema 10: Modelos de ecuaciones estructurales (tema de especialización de un estudiante). Presentación y práctica. Presentación de otro trabajo de especialización.
8-Nov	Presentación de dos trabajos de especialización.
15-Nov	Presentación de dos trabajos de especialización.
22-Nov	Presentación de dos trabajos de especialización.
29-Nov	Presentación de dos trabajos de especialización.
6-Dic	Examen final



Bibliografía

- Agresti, A. (2019). *An Introduction to Categorical Data Analysis* (3era edición). John Wiley and Sons, New York.
- Bartholomew, D.J, Steele, F, Moustaki, I, Galbraith J.I. (2002). *The analysis and interpretation of Multivariate Data for Social Scientists*. Chapman and Hall.
- Dillon, W.R, y Goldstein, M (1984). *Multivariate Analysis: Methods and Applications*. Wiley.
- Hair, J. F., Anderson, R.E., Tatham, R.L., Black, W.C. (1999). *Análisis Multivariante* (5a. edición). Prentice Hall, Madrid.
- Husson, F., Le, S., & Pagès, J. (2017). *Exploratory Multivariate Analysis by Example Using R* (2da edición). Chapman and Hall/CRC.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.
- Johnson, R.A y Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis* (6ta edición). Prentice-Hall.
- Kim, J and Mueller, C.W. (1978). *Introduction to Factor Analysis: What it is and How to do with it*. Sage University Paper.
- Kim, J and Mueller, C.W. (1978) *Factor Analysis: Statistical Methods and Practical uses*. Sage University Paper.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford Press
- Long, J.S. (1983). *Confirmatory Factor Analysis: A Preface to LISREL*. Sage University Paper. Long, J.S. (1983). *Covariance Structure Models: An Introduction to LISREL*. Sage University Paper.
- Thompson, B. (1984) *Canonical Correlation Analysis: Uses and Interpretation*. Sage University Paper.