



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:	Teórico
Número de créditos:	4 créditos
Horas presenciales:	4 horas semanales
Horas de atención de estudiantes:	2 horas semanales (M 17-18, Oficina)
Requisitos:	No tiene
Correquisitos:	No tiene
Profesor:	Johnny Madrigal Pana. M.Sc.

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. Análisis de conglomerados. Métodos divisivos jerárquicos, métodos de distancia y semejanza, y otros métodos. La matriz de distancia. Distancia y semejanza entre objetos basadas en variables continuas, distancia y semejanza entre objetos basadas en variables categóricas, dendrograma, ejemplos.
- VI. Árboles de clasificación. Criterios para CHAID, criterios para CRT, criterios para QUEST, costos de clasificación errónea, probabilidades previas, puntuaciones, estadísticos del modelo, gráficos.



- VII. 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.
- VIII. Escalamiento multidimensional. Medidas de semejanza entre variables, escalamiento multidimensional clásico, ordinal y métrico, la evaluación del ajuste y la elección del número de dimensiones.
- IX. Correlación canónica y análisis discriminante. Definición, usos, supuestos, técnica, interpretación de los resultados del análisis por correlación canónica, cargas canónicas, proporción de variancia explicada, el coeficiente de redundancia, problemas de interpretación: ponderaciones canónicas y cargas, validación, recomendaciones.
- X. 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.
- XI. Relaciones lineales estructurales (LISREL). Los modelos, notación y simbología, modelos recursivos y no recursivos, el modelo estructural, el modelo de medición, supuestos del modelo, etapas de construcción, pruebas de ajuste.

Metodología

Se impartirán clases magistrales, con la participación activa 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 el profesor expone y resume los aspectos principales de las técnicas de análisis y revisa con los estudiantes los procedimientos en el computador. Para cada técnica que se explique en clase, el profesor 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, así como tampoco archivos de otros países sobre los que se ignoren detalles. Con el archivo, después de analizar cada técnica, el estudiante escribirá un artículo en el que presentará los principales resultados, entregará copia impresa del artículo al profesor y realizará una presentación que no exceda 10 minutos de duración ante el grupo. 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 multivariadas incluidas en el programa del curso, cada estudiante deberá especializarse en una técnica nueva, estudiarla, aplicarla, escribir un artículo en el que



resuma los resultados y presentarla 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 el profesor. La selección de la técnica que cada estudiante escogerá, así como la duración de cada exposición, se hará de mutuo acuerdo con el profesor. Las técnicas que pueden escoger los estudiantes son diversas: regresión no lineal, regresión con rangos, regresión en cresta, regresión cuantílica, regresión segmentada, círculo de correlación, análisis conjunto, escalamiento óptimo, análisis de correspondencia multivariante, redes neuronales, temas de minería de datos, u otros análisis categóricos no vistos en los cursos de la maestría, etc. Algunos de los temas incluidos en el programa podrán ser desarrollados por los estudiantes como parte de este trabajo de especialización.

Todos los trabajos escritos y las presentaciones orales deben llenar los requisitos que reúnen un trabajo de investigación universitario a nivel de maestría.

Cronograma

I(2), II(2), III(1), IV(1), V(2), VI(1), VII(2), VIII(2), IX(1), X(1) y XI(2). El número de semanas de cada tema (entre paréntesis) es un valor estimado.

Marzo

L12: Revisión del programa. Reflexión sobre las presentaciones orales y los trabajos escritos.
Construcción de índices y práctica.

L19: Revisión del trabajo de construcción de índices. Este tema se entrega en un documento impreso, se revisa grupalmente y no se expone formalmente porque el archivo de datos que deben usar es el mismo para todos los estudiantes.

Análisis de factores. Construcción de Índices con Análisis de Factores y práctica.

L26: Semana Santa.

Abril

L2: Entrega y exposición de trabajos del tema de Análisis de Factores y construcción de índices.

L9: Análisis de Conglomerados y práctica.

L16: Análisis de Correspondencias y práctica (Semana Universitaria).

L23: Semana Universitaria. Presentación de trabajos del tema de Conglomerados y Análisis de Correspondencias.

L30: Escalamiento Multidimensional y práctica.

Mayo

L7: Análisis Discriminante y práctica.

L14: Regresión Logística Politémica y práctica.

L21: Presentación de trabajos de los temas de Escalamiento Multidimensional, Análisis Discriminante y Regresión Logística.

L28: Relaciones lineales estructurales y práctica.



Junio

- L4: Presentación de trabajos del tema de Relaciones lineales estructurales.
- L11: Análisis de factores con datos binarios (tema de especialización de un estudiante). Análisis de factores para variables categóricas ordenadas (tema de especialización de un estudiante).
- L18: Árboles de clasificación (tema de especialización de un estudiante). Presentación de trabajos del tema de Análisis de Factores con datos binarios y Análisis de Factores con variables categóricas ordenadas.
- L25: Presentación del trabajo de Árboles de Clasificación y de otros trabajos de especialización.

Julio

- L2: Presentación de otros trabajos de especialización. **Final del curso lectivo.**
- L9: Entrega de notas.

Resumen de trabajos a entregar y presentar

1. Construcción de índices.
2. Análisis de Factores Exploratorio y construcción de índices.
3. Análisis de Conglomerados.
4. Análisis de Correspondencias.
5. Escalamiento Multidimensional.
6. Análisis Discriminante y Regresión Logística Politémica.
7. Relaciones lineales estructurales.
8. Análisis de Factores con datos binarios o Análisis de Factores para variables categóricas ordenadas.
9. Árboles de Clasificación.
10. Técnica adicional de investigación del estudiante.

Bibliografía

- Aldenderfer, M.S. y Blashfield, R.K (1984). *Cluster Analysis*. Sage University Paper.
- Asher. *Causal modeling*. Sage University Paper.
- 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.
- Escobar, Modesto. (1998). *Las aplicaciones del análisis de segmentación: El procedimiento Chaid*. EMPIRIA. Revista de Metodología de Ciencias Sociales. No.1, 1998, pp.13-49
- Hair, J. F., Anderson, R.E., Tatham, R.L., Black, W.C. (1999). *Análisis Multivariante*. 5a. edición.
- Johnson, R.A y Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. Sexta 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.



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.
SPSS. Manuales: SPSS Base 10.0: Applications Guide, Advanced models 12.0, Regression Models 12.0, Categories 11.0,
SAS. Manuales
Thompson, B.(1984) *Canonical Correlation Analysis: Uses and Interpretation*. Sage University Paper.
Villegas B, Greibin. (2007). *Quienes están a favor del si, quienes a favor del no... Perfil mediante CHAID*. Memoria X Jornada de Análisis Estadístico de Datos (2007).

Evaluación

TOTAL	100%
Presentación oral y escrita de informes de técnicas multivariadas	90%
Presentación oral y escrita del trabajo y artículo final de investigación	10%