

Codigo de asignatura: 4773B4			
Nombre del programa académico	Maestría en Ingeniería Eléctrica		
Nombre completo de la asignatura	Análisis de Datos Multivariantes		
Número de créditos ECTS por categoría	Ciencias naturales y matemáticas	Módulos profesionales y especiales	Humanidades y ciencias sociales y económicas
	3	3	1
Semestre y año de actualización	2024-2		
Semestre y año en que se imparte	2025-2		
Tipo de asignatura	[ ] Obligatoria [X] Electiva		
Director o contacto del programa	Andrés Escobar Mejía		
Coordinador o contacto de la asignatura	Genaro Daza Santacoloma		
Descripción y contenidos			
<b>1. Breve descripción</b>			
La asignatura Análisis Multivariado introduce al estudiante en el estudio y aplicación de técnicas estadísticas orientadas al análisis simultáneo de múltiples variables. Se abordan métodos de exploración, reducción de dimensión, análisis de relaciones entre variables, clasificación y agrupamiento, con un enfoque práctico y aplicado. El curso combina fundamentos teóricos con ejercicios computacionales que permiten interpretar y modelar fenómenos complejos en contextos de ingeniería, economía, salud, entre otros. Se enfatiza el desarrollo de competencias para aplicar estas herramientas a datos reales utilizando software estadístico.			
<b>2. Objetivo del curso:</b>			
Proporcionar al estudiante una comprensión sólida de los principios, métodos y aplicaciones del análisis estadístico multivariado, con el fin de modelar, interpretar y resolver problemas que involucren múltiples variables observadas simultáneamente. El curso busca desarrollar competencias para seleccionar, aplicar y evaluar técnicas multivariantes de forma crítica, apoyándose en herramientas computacionales para el análisis de datos reales en diversos contextos de la ingeniería.			
<b>3. Resultados de aprendizaje. Los propósitos de formación en el estudiante de posgrado son:</b>			
<b>RAA1:</b> Interpretar la estructura y las relaciones entre múltiples variables mediante técnicas estadísticas multivariantes, considerando sus supuestos y limitaciones. Se corresponde con los RAP8, RAP11			
<b>RAA2:</b> Aplicar métodos de análisis multivariado —como análisis de componentes principales, análisis discriminante, análisis de clústeres y correlación canónica— para explorar y modelar conjuntos de datos complejos. Se corresponde con los RAP6, RAP7, RAP8, RAP11			
<b>RAA3:</b> Utilizar herramientas computacionales para implementar procedimientos de análisis multivariado y visualizar sus resultados de forma efectiva. Se corresponde con los RAP6, RAP7, RAP10, RAP11			
<b>RAA4:</b> Formular conclusiones técnicas sustentadas en evidencia estadística multivariada, comunicando sus hallazgos de manera clara y estructurada en contextos académicos o profesionales. Se corresponde con los RAP7, RAP8, RAP9, RAP10, RAP11, RAP12, RAP13			
<b>4. Contenido</b>			
<b>Unidad 1 – Introducción al análisis multivariado:</b> Naturaleza de los datos multivariantes. Aplicaciones en ingeniería, economía, salud, entre otros contextos. Tipos de variables y escalas de medición. Comparación con análisis univariado y bivariado.			
<b>Unidad 2 – Fundamentos estadísticos y algebraicos:</b> Matrices de datos y notación matricial. Medidas de centralización y dispersión multivariada. Matriz de covarianza y matriz de correlación. Distancia euclídea y distancia de Mahalanobis. Supuestos del análisis multivariado. Pruebas de normalidad multivariada (Mardia, Royston). Transformaciones de variables (log, Box-Cox, estandarización).			
<b>Unidad 3 – Visualización y análisis exploratorio:</b> Diagramas de dispersión múltiples. Gráficos de perfil y gráficos paralelos. Biplots y representación gráfica de estructuras. Identificación de observaciones atípicas.			
<b>Unidad 4 – Análisis de Componentes Principales (PCA):</b> Fundamentos geométricos y algebraicos. Autovalores, autovectores y varianza explicada. Selección del número de componentes. Interpretación de componentes y proyección de datos. Visualización y aplicación práctica.			
<b>Unidad 5 – Análisis Factorial Exploratorio (AFE):</b> Comparación con PCA. Criterios de extracción y comunales. Métodos de rotación (Varimax, Promax). Aplicaciones prácticas.			
<b>Unidad 6 – Inferencia multivariada:</b> Hotelling's $T^2$ . Pruebas sobre medias multivariadas. MANOVA: fundamentos y comparación con ANOVA. Aplicaciones prácticas.			
<b>Unidad 7 – Relaciones entre conjuntos de variables:</b> Análisis de correlación canónica. Análisis de correspondencias simples y múltiples. Escalamiento multidimensional (MDS).			
<b>Unidad 8 – Análisis de conglomerados (clústeres):</b> Agrupamiento jerárquico: métodos de vinculación y dendrogramas. Agrupamiento no jerárquico: k-medias y k-medoides. Determinación del número óptimo de grupos (método del codo, coeficiente de silueta, gap statistic).			
<b>5. Requisitos.</b> Los definidos en requisito de admisión de la IES.			
<b>6. Recursos</b>			

Durante el desarrollo del curso se utilizarán diferentes recursos tecnológicos y computacionales que apoyan tanto el análisis estadístico como la presentación y documentación de resultados. Entre ellos se incluyen:

**MATLAB:** software de cálculo numérico y visualización gráfica, utilizado para la implementación práctica de las técnicas multivariantes vistas en clase.

**Python** (con bibliotecas como numpy, pandas, matplotlib, seaborn y scikit-learn): como alternativa de código abierto para análisis de datos multivariantes.

**Overleaf:** plataforma en línea para redacción de documentos científicos en LaTeX, recomendada para la elaboración del informe técnico final del curso.

**Google Suite:** herramientas como Google Docs, Google Sheets y Google Slides para elaboración de presentaciones, reportes y documentos colaborativos.

**Google Classroom:** plataforma institucional para la gestión del curso, entrega de actividades, comunicación con el docente y acceso al material de clase.

**Sitios web para descarga de conjuntos de datos:**

- UCI Machine Learning Repository – amplia colección de datasets multivariantes clásicos y contemporáneos (<https://archive.ics.uci.edu>).

- Kaggle Datasets – plataforma colaborativa con conjuntos de datos diversos, descargables en formatos comunes (<https://www.kaggle.com/datasets>).

- DataHub – repositorio abierto de datos estructurados en múltiples dominios (<https://datahub.io>).

- OpenML – repositorio que permite explorar datasets y modelos de aprendizaje automático (<https://www.openml.org>).

- StatLib Data Archive (CMU) – colección académica clásica útil para ejercicios y demostraciones estadísticas (<http://lib.stat.cmu.edu/datasets/>).

**Bibliografía**

- Peña, D. (2002). Análisis de datos multivariantes. McGraw-Hill.

- Johnson, R. A., & Wichern, D. W. (2013). Applied Multivariate Statistical Analysis (6th ed.). Pearson.

- Flury, B. (2013). A First Course in Multivariate Statistics. Springer Science & Business Media.

- Greenacre, M. (2017). Correspondence Analysis in Practice (3rd ed.). Chapman & Hall/CRC.

- Hjellbrekke, J. (2018). Multiple Correspondence Analysis for the Social Sciences. Routledge.

- Rencher, A. C., & Christensen, W. F. (2012). Methods of Multivariate Analysis (3rd ed.). Wiley.

- Shaitino, L. (2013). Representación gráfica de información multivariante. Centro de Investigaciones Sociológicas – CIS.

**7. Herramientas técnicas de soporte para la enseñanza**

Tema	Actividad	Herramienta de soporte
Fundamentos del análisis multivariado	Lecturas orientadas y discusión guiada en clase	Google Classroom, Google Docs
Matriz de covarianza, distancia, supuestos	Resolución de ejercicios numéricos individuales	MATLAB, Python (Jupyter Notebook), pizarra digital
Visualización y análisis exploratorio	Taller grupal: análisis gráfico de un dataset	MATLAB, Python (matplotlib, seaborn), Google Slides
Análisis de Componentes Principales (PCA)	Ejercicio práctico con datos reales, presentación de resultados por equipos	MATLAB, Google Sheets, Google Slides, UCI Repository
Análisis Factorial Exploratorio (AFE)	Comparación de modelos factoriales entre estudiantes	Python (factor_analyzer), MATLAB, Google Docs
Pruebas de inferencia (Hotelling, MANOVA)	Taller individual con interpretación de resultados en contexto aplicado	MATLAB, Python (statsmodels), Overleaf, Google Docs
Correlación canónica, correspondencias, MDS	Actividad colaborativa: análisis e interpretación de estructuras latentes	Python, MATLAB, Overleaf, Kaggle Datasets
Agrupamiento (clústeres)	Actividad grupal: comparación entre agrupamientos jerárquicos y k-medias	MATLAB, Python, Google Sheets, visualizaciones interactivas
Proyecto final	Desarrollo guiado por etapas: análisis completo sobre datos reales	MATLAB, Python, Overleaf, Google Classroom, Google Docs, Google Slides

**8. Trabajos en laboratorio y proyectos**

**Tarea 1:** Trabajo colaborativo. Análisis de lectura. 1 hora

**Tarea 2:** Trabajo individual. Actividades de programación en torno a los fundamentos estadísticos y algebraicos. 6 horas estudiante.

**Tarea 3:** Trabajo individual. Análisis exploratorio gráfico de datos reales multivariados. 4 horas estudiante.

**Tarea 4:** Trabajo individual. Aplicación de análisis de componentes principales a un conjunto de datos. 4 horas estudiante.

**Tarea 5:** Trabajo individual. Implementación de un modelo de análisis factorial exploratorio y discusión de resultados. 6 horas.

**Tarea 6:** Trabajo individual. Taller de interpretación de resultados de pruebas de Hotelling y MANOVA en un contexto aplicado. 6 horas estudiante.

**Tarea 7:** Trabajo colaborativo. Actividad de análisis de correspondencias y escalamiento multidimensional. 4 horas estudiante.

**Tarea 8:** Trabajo colaborativo. Taller comparativo entre métodos de agrupamiento jerárquico y no jerárquico. 6 horas estudiante.

**Tarea 9:** Trabajo integrador de investigación formativa. Proyecto individual de análisis multivariado con datos reales. Incluye formulación del problema, análisis completo y entrega escrita. 24 horas estudiante.

**9. Métodos de aprendizaje**

**Cátedra magistral:** Se introducen conceptos fundamentales del análisis multivariado, acompañados de ejemplos explicativos y discusión dirigida.

**Aula extendida:** Se asignan lecturas especializadas y ejercicios prácticos para ser desarrollados de forma autónoma fuera del aula.

**Aprendizaje basado en problemas:** Se analizan conjuntos de datos reales mediante la formulación de preguntas y retos que requieren aplicar diferentes técnicas multivariantes.

**Trabajos colaborativos:** Se desarrollan actividades prácticas en grupo orientadas a la exploración, visualización y modelado multivariado de datos.

**Investigación formativa:** Se promueve la integración de conocimientos a través de un proyecto individual que requiere estructurar un análisis completo, desde la formulación del problema hasta la interpretación de resultados.

#### **10. Métodos de evaluación**

La evaluación de la asignatura se basa en la realización de tareas prácticas individuales y colaborativas, así como en un proyecto final integrador. Estas actividades permiten evidenciar la comprensión teórica, el dominio técnico y la capacidad de análisis del estudiante en el contexto del análisis multivariado. Cada tarea está alineada con los resultados de aprendizaje propuestos y tiene un valor definido dentro de la calificación final:

**Actividad 1 (20%):** Bitácora técnica y sustentación acerca de las tareas de fundamentos estadísticos y algebraicos y análisis exploratorio gráfico RAA1, RAA2, RAA3

**Actividad 2 (20%):** Bitácora técnica y sustentación de las tareas de análisis de componentes principales y análisis factorial exploratorio: RAA1, RAA2, RAA3

**Actividad 3 (15%):** Bitácora técnica y sustentación de las tareas en torno a la inferencia multivariada. Resultados de aprendizaje: RAA1, RAA2, RAA3, RAA4

**Actividad 4 (15%):** Bitácora técnica y sustentación de las tareas en torno al análisis de correlación canónica, el análisis de correspondencias y el escalamiento multidimensional. RAA2, RAA3, RAA4

**Actividad 5 (15%):** Bitácora técnica y sustentación de las tareas en torno al análisis de conglomerados. RAA2, RAA3, RAA4

**Actividad 6 (15%):** Trabajo integrador de investigación formativa. Proyecto individual de análisis multivariado con datos reales. Incluye formulación del problema, análisis completo y entrega escrita. Resultados de aprendizaje: RAA1, RAA2, RAA3, RAA4