Código de la Asignatura: IO6F3

Nombre del programa académico	Maestría en Investigación Operativa y Ciencia de Datos
Nombre completo de la asignatura	Análisis de Datos Multivariados
Área académica o categoría	Estadística Enfocada a la Ciencia de Datos
Semestre y año de actualización	Segundo semestre de 2023
Tipo de asignatura	[x] Obligatoria [ ] Electiva
Número de créditos ECTS	5.22 ECTS
Intensidad horaria	10 Horas Semanales – 40 Horas Mensuales
Director o contacto del programa	Eliana Mirledy Toro Ocampo
Coordinador o contacto de la asignatura	Orlando Antonio Sabogal Cardona - oasabogal@utp.edu.co Google Scholar Orlando Antonio Sabogal

Descripción y Contenidos

1. Breve descripción

La asignatura aborda, primero, los fundamentos del análisis estadístico multivariado (variables, medias, varianzas, pruebas de hipótesis y MANOVA) y, luego, técnicas de uso común como Componentes Principales, Análisis de Factores, Cluster, Discriminante, Clasificación y Correlación Canónica

2. Objetivos del programa académico

OP1. Analizar y combinar conocimientos y técnicas de la estadística y la optimización matemática bajo enfoques sostenibles, con el objetivo de tomar decisiones informadas, esenciales para abordar los desafíos actuales y futuros en ambientes académicos y empresariales. OP3. Analizar y extraer conocimientos significativos a partir de enfoques estadísticos multivariados, tanto cualitativos como cuantitativos, para examinar simultáneamente múltiples variables o factores en un conjunto de datos, con el objetivo de obtener información más completa y profunda sobre las relaciones y patrones entre ellas, aplicados en diferentes campos, como la investigación científica, la ingeniería, la economía, la medicina, entre otras.OP4. Integrar la investigación de operaciones y la estadística para promover un enfoque más completo y sólido, con el fin de abordar desafíos investigativos al considerar la eficiencia operativa, así como la incertidumbre inherente a los datos y los procesos bajo un marco que contemple el rigor ético, moral y científico. OP5. Fomentar dimensiones humanas que permitan generar espacios participativos de diálogo y reflexión con un enfoque sostenible, que contribuyan a la formación de profesionales capaces de abordar desafíos de manera integral y responsable.

Objetivos de la asignatura

OA1. Diseñar estructuras estadísticas para sistemas con múltiples variables interdependientes, empleando el pensamiento crítico para simplificar la complejidad y proponer diagnósticos que aseguren la sostenibilidad operativa en procesos industriales. OA2. Implementar técnicas multivariadas (MANOVA, Componentes Principales, Análisis de Conglomerados) mediante rigor científico, garantizando que el tratamiento de datos y la validación de supuestos reflejan la ética y la formación humana. OA3. Transformar grandes volúmenes de datos en decisiones estratégicas claras, comunicando patrones complejos de manera honesta y transparente para fortalecer la responsabilidad social

3. Resultados de aprendizaje de la asignatura

RAA1: Contrasta hipótesis en escenarios de múltiples variables simultáneas mediante el pensamiento crítico, permitiendo una resolución de problemas experimentales que optimice el uso de recursos y garantice la sostenibilidad de los procesos evaluados. RAA2: Sintetiza estructuras de datos masivas mediante el análisis de componentes principales y factores, validando la pérdida de información bajo criterios de rigor científico y la integridad técnica propia de la formación humana, evitando interpretaciones subjetivas sin sustento estadístico. RAA3:Diseña agrupamientos y clasificaciones de datos para identificar segmentos críticos en la organización, comunicando los hallazgos con un juicio profesional que priorice la transparencia y el impacto social de las decisiones derivadas del análisis.

4. Contenido : ●T1: Variables multivariadas, media y varianza, propiedades (15 h: 4 h con docente y 11 h independientes). ●T2: Estimadores de medias y varianzas. (15 h: 4 h con docente y 11 h independientes). ●T3: Pruebas de hipótesis sobre una, dos y más medias (21 h: 6 h con docente y 15 h independientes). ●T4: Análisis de varianza multivariada (21 h: 6 h con docente y 15 h independientes). ●T5: Análisis de Componentes Principales y Análisis de Factores (21 h: 6 h con docente y 15 h independientes). ●T6: Similaridades y Análisis de Conglomerados (21 h: 6 h con docente y 15 h independientes). ●T7: Análisis Discriminante (15 h: 4 h con docente y 11 h independientes). ●T8:
Análisis de Correlación Canónica (15 h: 4 h con docente y 11 h independientes)

5.Requisitos: Manejo de matrices y arreglos, Fundamentos del Álgebra Lineal, Teoría de probabilidad, Inferencia Estadística univariada

6. Recursos Libros de texto: ● Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford publications. ● Kline, R. B. (2023). Principles and practice of structural equation modeling. Guilford publications. ● Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). Multivariate data analysis. ● Finch, W. H. (2019). Exploratory factor analysis (Vol. 182). Sage Publications.● Gana, K., & Broc, G. (2019). Structural equation modeling with lavaan. John Wiley & Sons.● Zelterman, D. (2015). Applied multivariate statistics with R (pp. 393-393). Basel, Switzerland: Springer International Publishing.● Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. Springer Science & Business Media. ● Artículos académicos y lecturas complementarias proporcionadas durante el curso.

7. Herramientas técnicas de soporte para la enseñanza: Clases Magistrales ● Lectura fuera del aula ● Talleres en clase sobre pruebas de hipótesis para medias, Manovas, Componentes principales. ● Talleres fuera del aula sobre los diferentes temas del curso.

8. Trabajos en laboratorio y proyectos: Solución de problemas seleccionados sobre cada uno de los temas del curso mediante talleres en R (T1 a T8: 40 horas). Proyecto aplicado de análisis multivariado con datos reales (T5 a T8: 20 horas), que incluye un análisis de ética de datos: privacidad, sesgos y uso responsable de la segmentación

(6 horas), que se califica dentro de los talleres fuera de clase.
9. Métodos de aprendizaje: Trabajo en grupos. ● Exposiciones magistrales. ● Talleres. ● Solución de problemas. ● Resolución de problemas ejemplo en clase
10.Métodos de evaluación: Calificación de talleres fuera de clase: 40% · Evaluación sobre conceptos básicos teóricos del análisis multivariado: 30% · Evaluación sobre la segunda parte del curso (Técnicas multivariadas): 30%