

Información general del curso

Nombre de la asignatura	Métodos Estadísticos para Data Analytics
Créditos académicos	3
Horas presenciales por semana	3 horas
Horas de trabajo no presencial por semana	6 horas

MÉTODOS ESTADÍSTICOS PARA DATA ANALYTICS

La gran mayoría de fenómenos basados en datos están sujetos a situaciones de incertidumbre que modelamos mediante esquemas probabilísticos o aleatorios. La estadística, como ciencia que permite explicar condiciones de algún fenómeno de ocurrencia aleatoria o condicional, es una herramienta esencial para el análisis de datos, el ajuste de modelos y la comprensión de los factores que definen a los mismos. La inclusión de la aleatoriedad y su estudio a través de diversos métodos estadísticos es un componente formativo necesario para el estudiante del Doctorado en Ingeniería. Es necesario que el estudiante comprenda que la generación de un modelo no termina con el proceso de validación, sino que existen otra serie de análisis que permitirán establecer la estabilidad y robustez de un modelo probabilístico cualquiera que éste sea. De igual manera, es necesario hacer uso de métodos estadísticos con el fin de caracterizar el comportamiento de grupos de individuos o componentes de forma tal que el fenómeno a modelar sea basado en muestras lo suficientemente representativas y que las técnicas de estimación permitan llegar a conclusiones relevantes para el modelo a estudiar.

Objetivos de aprendizaje

Objetivo General

El objetivo de la asignatura MÉTODOS ESTADÍSTICOS PARA DATA ANALYTICS será el de promover y desarrollar en los estudiantes la comprensión e implementación de conceptos y métodos estadísticos, útiles en el campo analítica de datos, que permitan considerar la dimensión aleatoria de fenómenos de variabilidad inherente, su modelamiento y sus implicaciones en las diversas aplicaciones consideradas en este plan de estudios.

Objetivos específicos

- Dominar el uso de métodos de estadística descriptiva
- Entender los conceptos y principios básicos de las distribuciones probabilísticas, distribuciones de muestreo y estimación.
- Entender los conceptos básicos de modelos lineales, comparación de modelos, diseño de experimentos, análisis de errores y colinealidad
- Entender los conceptos básicos asociados a la estadística e inferencia Bayesiana, los modelos jerárquicos y el uso de técnicas de MCMC

Evaluación

Aspectos a evaluar:

- Identificación y apropiación conceptual de estadística y sus principios asociados
- Conocimiento y apropiación de los métodos de estadística descriptiva, modelos lineales y estadística bayesiana
- Adquisición y aplicación de métodos estadístico
- Capacidad de desarrollar proyectos de aplicación

CALIFICACIONES

50 % Tareas y Talleres

50 % Proyecto

Dinámica de clase/ Metodología

La dinámica para adelantar esta asignatura, está definida por las actividades que el estudiante define realizar con su profesor directos y que puede incluir:

- Clases presenciales
- Prácticas en laboratorio
- Lecturas asignadas por el profesor
- Proyecto sobre un caso práctico
- Talleres

Trabajo del estudiante

- Reforzar los temas vistos en clase realizando lectura de textos y artículos relacionados
- Completar los ejercicios prácticos propuestos
- Desarrollar un proyecto sobre un caso práctico
- Preparar un documento y presentación sobre el tema del proyecto

Se espera que, al finalizar esta asignatura, el estudiante finalice su doctorado.

Cronograma del curso

- Introducción
- Medidas de tendencia central y dispersión
- Medidas de posición relativa, correlación y regresión
- Variables aleatorias y distribuciones
- Distribuciones discretas
- Distribuciones continuas
- Leyes de grandes números y Teorema del Límite Central
- Modelos lineales, estimación de parámetros, comparación de modelos, intervalos de confianza
- Experimentos diseñados, datos observacionales
- Identificación de observaciones inusuales, verificación de supuestos del modelo, errores de escala
- Colinealidad
- Inferencia bayesiana
- Modelos jerárquicos bayesianos y modelos empíricos bayesianos
- Métodos de Monte Carlo (MCMC) para estimación de distribuciones posteriores
- Taller de estadística bayesiana usando R
- Parcial y entrega proyecto final

Bibliografía

1. Quinn G, Keough, M. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, New York, USA. 537 pp.
2. Robert C., Casella G. 2010. Introducing Monte Carlo Methods with R. Springer Science + Business Media, New York, USA. 283 pp.
3. Saltelli A., Chan K., Scott, M. 2008. Sensitivity Analysis. John Wiley & Sons, Chichester, England. 492 pp.
4. Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin. Bayesian Data Analysis, Second Edition. Chapman & Hall/CRC Texts in Statistical Science, 2009.

5. Dennis D. Boos, L.A. Stefanski. Essential Statistical Inference. Theory and Methods Springer, Nueva York, 2013
6. Larry A. Wasserman . All of Statistics: A Concise Course in Statistical Inference. Springer Texts in Statistics, Nueva York, 2004