



## **Curso Estadística Avanzada y Aplicaciones**

CURE Rocha

Primer semestre de 2016

**Docentes:** Dr. Mathias Bourel (Laboratorio de Probabilidad y Estadística, IMERL, Facultad de Ingeniería), Dra. Carolina Crisci y Dr. Ángel Segura (PDU Modelización y Análisis de Recursos Naturales, CURE)

**Lugar:** CURE, sede Rocha, salón 5

**Días y horario:** viernes de 13:30 a 17:00

**Fecha de inicio:** viernes 1° de abril

**Carga horaria total:** 112 horas (45 horas teóricos + 18 horas prácticos + 43 horas actividades individuales + 3 horas defensa de evaluaciones + 3 horas seminarios)

**Modalidad:** presencial

**Aprobación:** Evaluación continua, entrega de ejercicios prácticos y de un trabajo final de aplicación de técnicas con datos de interés para el participante. El trabajo final tiene defensa oral obligatoria.

**Cursos previos:** curso básico de Estadística

**Inscripción:** Los interesados pueden inscribirse por dos vías posibles: escribiendo a [posgrado@fing.edu.uy](mailto:posgrado@fing.edu.uy) o en la Bedelía de Facultad de Ciencias. En ambos casos se solicita además que se notifique de la inscripción escribiendo a [carocrisci@gmail.com](mailto:carocrisci@gmail.com), indicando porque vía se realizó la misma.

### **Objetivos y estrategia del curso**

El curso de Estadística Avanzada y Aplicaciones está destinado a docentes, investigadores y estudiante de diversos grupos de trabajo y carreras, en particular de posgrado. Se tratarán temas que pretenden dar herramientas para fortalecer el análisis estadístico en trabajos de grupos de investigación, así como de estudiantes en formación.

El curso constará mayormente de exposiciones por parte de los docentes sobre técnicas estadísticas de interés para el análisis de datos. Se dará un espacio importante a la ilustración de las técnicas mediante la presentación de aplicaciones y se promoverá la

participación activa de los asistentes. Habrá instancias prácticas en donde se aplicarán las técnicas expuestas utilizando el software R.

Por último, se contará con exposiciones (seminarios) por parte de investigadores nacionales sobre algunas aplicaciones de las técnicas vistas en el curso en distintas áreas de interés de los participantes.

### Actualización para docentes e investigadores del CURE

En esta edición del curso ofrecemos la asistencia libre a docentes e investigadores del CURE que puedan tener interés en asistir a alguna de las clases o módulos. Para este no hay ningún requisito, simplemente presentarse en el salón el o los días de las clases de interés (ver debajo el programa del curso). De ser posible se solicita que se envíe un mail a [carocrisci@gmail.com](mailto:carocrisci@gmail.com) uno o dos días antes de la clase de interés por si se presenta algún cambio en el programa.

## **Programa**

### Introducción

Clase 1 (1° de abril): presentación del curso, repaso el concepto de modelo estadístico, repaso de conceptos básicos (distribución de probabilidad, test de hipótesis, intervalos de confianza, etc.), repaso de álgebra lineal.

Clase 2 (8 de abril): Distribución Normal multivariada, introducción al software R.

### Módulo 1: Análisis de ordenación

Clase 2 (8 de abril): Análisis de Componentes Principales

Clase 3 (15 de abril): Análisis de Correspondencia y Análisis de Cluster

Clase 4 (22 de abril): Práctico en R

### Módulo 2: Análisis clásicos de Regresión y Clasificación

Clase 5 (29 de abril): Regresión lineal simple y múltiple

Clase 6 (6 de mayo): Modelos Lineales Generalizados (GLM)

Clase 7 (13 de mayo): Práctico en R

Clase 8 (20 de mayo): Análisis Discriminante + práctico en R

### Módulo 3: Técnicas avanzadas de Regresión y Clasificación

Clase 9 (27 de mayo): Introducción al Aprendizaje Automático

Clase 10 (3 de junio): Máquinas de Vectores de Soporte (SVM) + práctico en R

Clase 11 (10 de junio): Modelos Aditivos Generalizados (GAM) + práctico en R

Clase 12 (17 de junio): Árboles de Clasificación y Regresión (CART) + práctico en R

Clase 13 (24 de junio): Métodos de agregación de modelos: Bagging, Boosting, Random Forest, Modelos de consenso + práctico en R

## Seminarios y clases de consulta

Clase 14 (1º de julio): Seminarios de investigadores invitados

Clase 15 (8 de julio): Clase de consulta sobre trabajos finales

## **Bibliografía**

**Bourel, M. (2012)** Model aggregation methods and applications. Memorias de trabajos de difusión científica y técnica, Vol. 10, p. 19-32, 2012.

**Crawley, M. J. (2013)** The R Book, 2<sup>nd</sup> edition. Wiley & Sons

**Crisci, C.; Ghattas, B.; Perera, G. (2012).** A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling 240: 113-122

**De'Ath, G. et al (2000).** Classification and regression trees: a powerful yet simple technique for ecological data analysis

**Everitt, B.; Torsten, H. (2010).** A Handbook of Statistical Analyses Using R, 2nd edition, Brian S., Chapman & Hall/CRC

**Hastie, T.; Tibshirani, R. (1990)** Generalized Additive Models. Chapman and Hall, London.

**Hastie, T.; Tibshirani, R.; Friedman, J. (2011).** The Elements of Statistical Learning: Data Mining, Inference and Prediction. (5th. Edition). Springer Seris in Statistics

**Izenman, A. J. (2008).** Modern Multivariate Statistical Techniques. Regression, Classification and Manifold Learning. Springer Texts in Statistics

**James, G.; Witten, D.; Hastie, T.; Tibshirani, R. (2013)** An introduction to Statistical Learning with Applications in R, Springer

**Legendre, P.; Legendre, L. (1998).** Numerical Ecology. Elsevier Science

**McCullagh, P.; Nelder, J. A. (1989)** Generalized Linear Models. Chapman & Hall

**Peña, D. (2002).** Análisis de datos multivariantes, Daniel Peña, Mac Graw Hill

**Wood, S. N. (2006)** Generalized Additive Models, An Introduction with R. Chapman & Hall