

Nombre de la Materia	<b>Ciencia de datos con R: Herramientas avanzadas de análisis estadístico</b>
Departamento/Intituto	<b>Instituto de Cálculo</b>
Fecha de Inicio	<b>2do. Cuatrimestre de 2017</b>
Duración	<b>Curso Teórico-práctico 54 hs. totales (18 encuentros de 3hs. cada uno, 2 veces por semana)</b>
Puntaje	
Forma de Evaluación	<b>Examen final</b>
Docentes a Cargo	<b>Andrés Farall</b>
Docente Auxiliar	<b>Marina Valdora</b>
Cupo	<b>30</b>
Arancel	<b>1200 módulos</b>

# Ciencia de datos con R: Herramientas avanzadas de análisis estadístico

Noviembre, 2016

La existencia de grandes y complejas estructuras de información exige disponer de nuevas técnicas de análisis y visualización de datos, a la vez de mutar de un enfoque metodológico centrado en la teoría, a un enfoque centrado en la preeminencia del dato. La fuerte dinámica en la evolución de las estructuras de información generadas por la ciencia, las empresas y las distintas dependencias del estado, requieren contar con un entorno de creación y utilización de herramientas de análisis que comparta esa misma dinámica. En este contexto, el entorno de programación R se ha convertido en líder entre los paradigmas libres y comunitarios para el análisis y el modelado de datos. Así, la Ciencia de Datos conjuga en un contexto amplio las técnicas más exitosas disponibles, provenientes de la estadística aplicada, la minería de datos y el aprendizaje automático.

**Objetivos:** El curso procura brindar herramientas aplicadas, provenientes principalmente de la estadística, útiles para la descripción, el análisis y el modelado de datos en un contexto científico interdisciplinario. Todas las técnicas estudiadas son implementadas en R, y ejemplificadas con conjuntos de datos reales. El enfoque particular del curso se centra en los datos y en la selección de métodos de análisis que se ajustan a ellos. Se intentará dotar al estudiante del conocimiento de una amplia variedad de técnicas útiles en el contexto de la ciencia de datos. Un énfasis especial se dedica a las novedosas técnicas de descripción y visualización de datos disponibles en el entorno R. Asimismo, en el curso se darán las motivaciones estadísticas de las distintas herramientas consideradas sin profundizar en los fundamentos teóricos subyacentes.

**Modalidad:** Las clases serán teórico-prácticas y se desarrollarán en el laboratorio de computación o en un espacio adecuado para que cada participante disponga de una computadora. El curso constará de 54 horas, distribuidas en 18 encuentros de 3 horas cada uno, dos veces por semana. Adicionalmente, se requiere que el espacio previsto para el desarrollo de las

clases cuenta con proyector, con el cual se expndrán los contenidos teóricos del curso.

**Contenidos:**

- Técnicas supervisadas
  - Modelos Lineales Mixtos, Regresión Ridge, Lasso y Logística.
  - Regression Quantile
  - Análisis Discriminante (LDA y QDA)
  - Naive Bayes
  - KNN (vecinos más cercanos)
  - CART y Random Forests
  - Redes Neuronales
  - Support Vector Machines (SVM)
  - Reglas de Asociación
- Técnicas de segmentación (Clustering)
  - K-means y PAM
  - Mclust
  - Clusterización jerárquica
  - Mean Shift
  - DBSCAN
- Revisión de aplicaciones al Análisis Multivariado
  - FA (Análisis Factorial)
  - PCA (Componente Principales)
  - MCA (Análisis de Correspondencia)
  - Correlación Canónica
  - Mapping (MDS)
- Revisión de más métodos
  - Herramientas de Estadística Espacial (Leaflet, Rasters, Krigging)
  - Manejo y análisis de datos satelitales
  - Algoritmos genéticos
  - Herramientas para el control de calidad

**Prerrequisitos:** Para ser admitido al curso se requerirá haber cursado la materia *Ciencia de datos con R: Fundamentos Estadísticos*, o acreditar conocimientos de estadística y de R.

**Sistema de evaluación:** Se realizará un examen final individual.

**Cupo:** Se aceptará un máximo de 30 alumnos.

**Bibliografía:**

- Chi Yau (2015). The R Tutorial. Online material available. Link: <http://www.r-tutor.com/>
- Data Mining Algorithms In R (2014). Link: [https://en.wikibooks.org/wiki/Data\\_Mining\\_Algorithms\\_In\\_R](https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R)
- Kelly Black (2015). R Tutorial. Link: <http://www.cyclismo.org/tutorial/R/index.html>.
- Lantz, B. (2013). Machine learning with R. Packt Publishing Ltd.
- Peng, R. D. (2015). R Programming for Data Science. Lulu. com.
- R Development Core Team (2016). The R Manuals. Link: <https://cran.r-project.org/manuals.html>.
- Wickham H. (2014) Advanced R (The R series). Link: <http://adv-r.had.co.nz/>
- Williams G. (2011) Data Mining with Rattle and R.
- Yanchang Zhao (2012). R and Data Mining. Online book available, link: <http://www.rdatamining.com/books/rdm>