

# Ciencia de datos con R: Fundamentos Estadísticos

Diciembre 2017

La Estadística ha adquirido un notable espacio en la ciencia moderna. Por una lado, brinda una vasta gama de herramientas que posibilitan entender el comportamiento de diversos fenómenos, cuantificando el rol de las diferentes variables involucradas en los mismos. Por otra parte, el desarrollo de métodos sumamente flexibles con alta capacidad predictiva revoluciona la disciplina, impactando incluso en nuestra vida cotidiana. Esta versatilidad permite que la disciplina ocupe un lugar cada vez más destacado en diversas áreas del conocimiento.

Por otro lado, R es un lenguaje de alto nivel y un entorno muy rico para el análisis, modelización y visualización de datos, que ha ganado un gran número de adeptos en todo el mundo. Este software tiene un enorme potencial en aplicaciones y como es de uso libre, miles de usuarios alrededor del mundo comparten su información. Sus funcionalidades están en permanente mejoramiento y extensión a través de librerías cuya documentación es de circulación libre y gratuita. Estas características convierten a R en una opción imbatible a la hora de elegir un paquete computacional, tanto por la cobertura que en la actualidad ofrece para el cómputo de distintos procedimientos estadísticos, como por su flexibilidad, que le permite incorporar paquetes para calcular nuevos métodos en forma permanente. Muestra de ello es que la mayor parte de la metodología estadística que se ha desarrollado en los últimos años es acompañada por un código R que facilita su transferencia a posibles usuarios de múltiples disciplinas.

Por estas razones, recorrer conceptos y técnicas estadísticas siguiendo el camino que ofrece el paquete computacional R merece un destacado lugar en la formación de nuestros egresados.

**Objetivos:** Esta materia propone revisar y extender diversos tópicos fundamentales de Estadística mediante la aplicación y el análisis de datos con el paquete computacional estadístico R. El curso tiene por principal objetivo transmitir los fundamentos del razonamiento estadístico; haciendo énfasis en el uso de la simulación computacional como herramienta clave para asimilar las nociones que se presentan a lo largo del mismo. Se estudian diferentes

modelos, especificando en cada uno de ellos los parámetros de interés y diferentes formas de estimarlos. Presentados los estimadores, se enfatiza en la necesidad de conocer o aproximar la distribución de los mismos a la hora de realizar inferencia (cálculo de p-valores, intervalos de confianza, tests de hipótesis).

A lo largo del curso se presentan diferentes aplicaciones de casos de estudio que han surgido a partir de colaboraciones con Investigadores de la Facultad en diferentes disciplinas.

**Modalidad:** Las clases serán teórico-prácticas y se desarrollarán en laboratorio de computación o espacio adecuado para que cada participante disponga de una computadora. Las clases serán de 3 horas cada una, dos veces por semana, a lo largo de todo el cuatrimestre.

**Contenidos:**

*Unidad 1.* Introducción al programa R. Entorno R Studio. Manejo de Datos en R. Visualización de Datos: Gráficos en R. Rudimentos de programación en R.

*Unidad 2.* Revisión de algunos conceptos básicos de Probabilidad. Simulación de Monte Carlo.

*Unidad 3.* Métodos de estimación. Simulación en R.

*Unidad 4.* Estimación no paramétrica de densidad. Selección de ventana: validación cruzada. Selección de modelos. Implementación en R y caso de estudio.

*Unidad 5.* Bootstrap: métodos de remuestreo. Aplicaciones en R.

*Unidad 6.* Aprendizaje estadístico: introducción al problema de la predicción.

*Unidad 7.* Regresión lineal múltiple. Selección de Modelos. Predicción. Implementación en R y caso de estudio.

**Prerrequisitos:** Algún curso introductorio de Probabilidad y/o Estadística.

**Sistema de evaluación:** Un exámen parcial y un exámen final.

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



### Bibliografía:

- Dalgaard, P. (2008). Introductory statistics with R. Springer Science & Business Media.
- Hothorn, T., & Everitt, B. S. (2014). A handbook of statistical analyses using R. CRC press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York: springer.
- Lock, R. H., Lock, P. F., & Morgan, K. L. (2012). Statistics: Unlocking the power of data. Wiley Global Education.
- Rice, J. (2006). Mathematical statistics and data analysis. Nelson Education.
- Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference. Springer Science & Business Media.
- Zuur, A., Ieno, E. N., & Meesters, E. (2009). A Beginner's Guide to R. Springer Science & Business Media.



Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 506.902/16

Buenos Aires, 12 MAR 2018

VISTO:

la nota a foja 17 presentada por la Dirección del Instituto de Cálculo, mediante la cual eleva la información del curso de posgrado **Ciencia de Datos R: Fundamentos Estadísticos**, para el año 2018.

CONSIDERANDO:

lo actuado por la Comisión de Doctorado,

lo actuado por la Comisión de Posgrado,

lo actuado por la Comisión de Presupuesto y Administración,

lo actuado por este cuerpo en la sesión realizada en el día de la fecha,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD DE CIENCIAS EXACTAS Y  
NATURALES  
RESUELVE:**

Artículo 1°.- Autorizar el dictado del curso de posgrado **Ciencia de Datos R: Fundamentos Estadísticos** de 96 hs de duración, que será dictado por las Dras. Daniela Rodríguez y Mariela Sued.

Artículo 2°.- Aprobar el programa del curso de posgrado **Ciencia de Datos R: Fundamentos Estadísticos** obrante a fs 18 y 19 del expediente de referencia, que será dictado durante el primer cuatrimestre de 2018.

Artículo 3°.- Aprobar un puntaje máximo de cuatro (4) puntos para la Carrera del Doctorado.

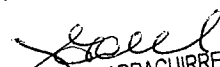
Artículo 4°.- Aprobar el siguiente arancel:

- 1200 módulos.

Disponer que los fondos recaudados ingresen en la cuenta presupuestaria habilitada para tal fin y sean utilizados de acuerdo a la Resolución 072/03.

Artículo 5°.- Comuníquese a la Dirección del Instituto de Cálculo, la Dirección de Alumnos, la Biblioteca de la FCEyN, la Dirección de Movimiento de Fondos, la Dirección de Presupuesto y Contabilidad, y la Secretaría de Posgrado con fotocopia del programa incluida. Cumplido archívese.

Resolución CD N° 0386  
ga 28/12/2017

  
Dr. JOSÉ GLABE PARRAGUIRRE  
SECRETARIO DE POSGRADO  
FCEN - UBA

  
Dr. JUAN CARLOS REBORES  
DECANO