

# Programa Introducción al análisis de datos con R

Asociación Argentina de Bioinformática y Biología Computacional

Fundación Instituto Leloir

Marzo 2021

## Objetivos:

El curso propone una introducción a la programación con el lenguaje R orientada al análisis de datos en general, con especial énfasis en el análisis de datos biológicos a partir de ejemplos.

## Audiencia:

El curso está pensado para investigadores y estudiantes de grado y posgrado que tengan interés en hacer sus primeros pasos en la programación y en el análisis de datos. No se requieren conocimientos previos de programación aunque es aconsejable tener conocimientos mínimos de computación, de estadística y de biología para comprender mejor los ejemplos (sin embargo, esto último no es excluyente).

## Duración y modalidad:

20 horas en total, con 15 horas de teoría y 5 horas de práctica. Las clases se dictarán de forma virtual.

## Certificados:

Se tomará un examen no obligatorio al finalizar el curso y se emitirá certificado de aprobación con nota. Se emitirán certificados de participación para aquellas personas que no rindan el examen.

## Docentes:

Lic. Andres Rabinovich

Lic. Maximiliano Beckel

Mag. Bruno Kaufman

Lic. Ariel Berardino

## Programa:

Clase 0) Preparación del entorno de trabajo antes de asistir al curso:

- Instalación de R, RStudio y paquetes que serán indicados.

#### Clase 1) Introducción a R y RStudio.

- Entorno Rstudio.
- Variables y tipos de datos.
- Estructuras de control.
- Funciones.
- Ayuda.

#### Clase 2) Introducción a R, continuación.

- Subsetting.
- Vectorización.
- Importación, exportación y limpieza de datos.

#### Clase 3) Estadística descriptiva.

- Elementos básicos de estadística.
- Visualización de datos: scatterplot, histograma, boxplot, etc.

#### Clase 4) Inferencia estadística.

- Tipos de errores.
- Intervalos de confianza.
- Pruebas de hipótesis.

#### Clase 5) Regresión lineal

- Modelos lineales.
- Regresión.

#### **Bibliografía recomendada (manuales, tutoriales y libros):**

- An introduction to R. [\[online\]](#)
- Peng R. (2015), R Programming for Data Science, Lean Publishing
- Everitt B. (2010), A Handbook of Statistical Analyses Using R, CRC Press.
- Verzani J. (2002), simpleR: Using R for Introductory Statistics. [\[pdf\]](#)
- Dalgaard P. (2008), Introductory Statistics with R, Springer
- Mangiafico S. (2015), AN R COMPANION FOR THE HANDBOOK OF BIOLOGICAL STATISTICS, New Brunswick, Rutgers University [\[pdf\]](#)
- McDonald J. (2014). HANDBOOK OF BIOLOGICAL STATISTICS, SPARKY HOUSE PUBLISHING. [\[pdf\]](#)
- Zeileis A. (2013), Extended Model Formulas in R: Multiple Parts and Multiple Responses. [\[pdf\]](#)
- Seefeld K. (2007), *Statistics Using R with Biological Examples*, Durham, NH, University of New Hampshire, Department of Mathematics & Statistics. [\[pdf\]](#)

- Mangiafico, S. (2016). *Summary and analysis of extension program evaluation in R*. New Brunswick, Rutgers University [\[pdf\]](#)
- Wickham h. (2009), *ggplot2: Elegant graphics for data analysis*, Springer
- Barabasi A. Network Science. [\[online\]](#)
- Tan, Steinbach & Kumar "Introduction to Data Mining". [\[online\]](#)
- Hastie, Tibshirani & Friedman, "The Elements of Statistical Learning", 2nd ed, Springer, 2009. [\[online\]](#)
- James, Witten, Hastie & Tibshirani, "An Introduction to Statistical Learning with Applications in R", 2<sup>nd</sup> ed, Springer, 2017 [\[online\]](#)