

Programa Introducción al análisis de datos con R

Asociación Argentina de Bioinformática y Biología Computacional

octubre 2020

# **Duración:** 17.5 horas.

**Objetivos:**

El curso introduce conceptos relacionados con el análisis de datos, haciendo uso de herramientas del lenguaje de programación R.

**Audiencia:**

El curso está pensado para investigadores y estudiantes de grado, que tengan interés en el análisis de datos en general.

**Docentes:**

Lic. Andres Rabinovich

Lic. Maximiliamo Beckel

Dr. Ariel Chernomoretz

**Programa:**

Clase 0) (ANTES DE COMENZAR EL CURSO)

**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) Puebas de hipótesis.

* Intervalos de confianza
* Tipos de errores
* Pruebas de hipótesis

Clase 5) Regresión lineal

* Modelos lineales
* Regresión

Bibliografía recomendada (manuales, tutoriales y libros):

* An introduction to R. [[online](https://cran.r-project.org/doc/manuals/r-release/R-intro.html)]
* 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](https://cran.r-project.org/doc/contrib/Verzani-SimpleR.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](http://rcompanion.org/documents/RCompanionBioStatistics.pdf)]
* McDonald J. (2014). HANDBOOK OF BIOLOGICAL STATISTICS, SPARKY HOUSE PUBLISHING. [[pdf](http://www.biostathandbook.com/HandbookBioStatThird.pdf)]
* Zeileis A. (2013), Extended Model Formulas in R: Multiple Parts and Multiple Responses. [[pdf](https://cran.r-project.org/web/packages/Formula/vignettes/Formula.pdf)]
* Seefeld K. (2007), *Statistics Using R with Biological Examples*, Durham, NH, University of New Hampshire, Department of Mathematics & Statistics. [[pdf](https://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf)]
* Mangiafico, S. (2016). *Summary and analysis of extension program evaluation in R*. New Brunswick, Rutgers University [[pdf](http://rcompanion.org/handbook/)]
* Wickham h. (2009), *ggplot2: Elegant graphics for data analysis*, Springer
* Barabasi A. Network Science. [[online](http://networksciencebook.com/chapter/0" \l "introduction0)]
* Tan, Steinbach & Kumar "Introduction to Data Mining". [[online](https://www-users.cs.umn.edu/~kumar001/dmbook/index.php" \l "chapters)]
* Hastie, Tibshirani & Friedman, "The Elements of Statistical Learning", 2nd ed, Springer, 2009. [[online](https://web.stanford.edu/~hastie/ElemStatLearn/)]
* James, Witten, Hastie & Tibshirani, "An Introduction to Statistical Learning with Applications in R", 6th ed, Springer, 2015 [[online](http://faculty.marshall.usc.edu/gareth-james/ISL/)]