

Bibliografía sobre estadística, bioestadística, data science y programación en R

Joaquín Amat Rodrigo j.amatrodrigo@gmail.com

2017

Versión PDF: <https://github.com/JoaquinAmatRodrigo/Estadistica-con-R>

El siguiente listado contiene las fuentes de información sobre estadística, *data science* y programación en R que he ido encontrando y que me han resultado útiles. Están ordenadas según, a mi parecer, facilitan el aprendizaje.

A la hora de adentrarse en estas disciplinas existen diferentes aproximaciones, en mi caso, el objetivo inicial fue aprender las principales herramientas estadísticas empleadas en el análisis de datos del ámbito biomédico. La programación (en este caso R) era simplemente la herramienta auxiliar de la que me ayudaba para realizar los cálculos necesarios. La experiencia me he convencido de que adquirir ambos conocimientos en paralelo mejora mucho el aprendizaje. ¡Nada mejor para asimilar un método estadístico que poder simular, representar y aplicar fácilmente los cálculos necesarios!

Siguiendo esta idea, *utilizar la programación como herramienta para el análisis de datos*, recomiendo el siguiente orden:

1. **Open Intro Statistics:** Una buena introducción a la estadística. Bien explicado y con ejemplos interesantes. Libro gratuito con un curso online también gratuito.
2. **Swirl:** aprender R desde el terminal de forma dinámica. Visitar página web para ver instrucciones de instalación.
3. **R for Data Science, Garrett Golemund and Hadley Wickham:** Introducción al uso de R aplicado al análisis de datos, con particular énfasis en el manejo, limpieza y visualización de la información.
4. **Handbook of Biological Statistics:** Un libro que explica los test estadísticos de forma muy clara sin profundizar en la matemática, identifica las ventajas y limitaciones de cada uno. Muy recomendable.
5. **Bioestadística Francisca Rius Díaz:** Libro en español de introducción a la bioestadística. Con ejemplos para realizar de forma manual, no por ordenador.
6. **Introduction to Statistical Learning with Applications in R:** Libro en el que se introducen y se describen los principales modelos de regresión así como su implementación en R. Muy recomendable.

7. **Practical Guide To Cluster Analysis in R, Alboukadel Kassambara:** Introducción a las técnicas de *clustering* y su implementación en R.
8. **Mastering Software Development in R Roger D. Peng, Sean Kross, and Brooke Anderson:** Libro dedicado a profundizar sobre el lenguaje de programación R.

Encuentro muy interesante leer la reflexión que hace *David Robinson* en su [blog](#) sobre las diferentes formas de introducir el lenguaje de programación R a principiantes.

Otras fuentes de información que me han resultado útiles:

- **Points of Significance, Nature:** La revista científica *Nature Methods* publica unas columnas en las que se explica de forma muy clara los principales métodos estadísticos empleados en ciencia.
- **Statistics Using R with Biological Examples:** En este libro se introduce de forma simultánea el uso básico de R y conceptos de estadística.
- **TheRBook Michael J Crawley:** Es un manual de uso de R. Recomendable leer los primeros capítulos para aprender los conceptos básicos. Los últimos capítulos son más útiles como consulta de temas concretos.
- **Métodos estadísticos en ingeniería Rafael Romero Villafranca, Luisa Rosa Zúnica Ramajo**
- **Statistical Bioinformatics with R, Sunil K. Mathur**
- **R Tutorials by William B. King, Ph.D** <http://ww2.coastal.edu/kingw/statistics/R-tutorials/>. Describe los test estadísticos más comunes de forma práctica, con sus ventajas y desventajas. Incluye ejemplos en R.
- <http://www.um.es/ae/FEIR/20/> apuntes de la Universidad de Murcia. Repaso de los diferentes test estadísticos, ventajas, limitaciones y ejemplos en R.
- **Bootstrap Methods and Permutation Tests by Tim Hesterberg:** Introducción muy intuitiva a los conceptos de *bootstrapping* y *resampling*.
- **Comparing groups Randomization and Bootstrap Methods using R Andrew S. Zieffler:** Descripción detallada de los métodos de *resampling* y *bootstrapping*.
- **Linear models with R, Faraway:** Libro en el que se introducen y se describen los principales modelos de regresión así como su implementación en R. Profundiza un poco más en el aspecto matemático que el libro *Introduction to Statistical Learning with Applications in R*, por lo que recomiendo leerlo a continuación de este último.