# Topic 2: Data Analysis

To deepen one's knowledge
Data Analysis learning resources

Statistics and data analysis is a hot topic, and a pain in the neck for Biological Sciences students. As a result, there is an unreachable number of resources available. Although the basic statistics remain as they were fifty years ago, new tools are today available, and many learning resources combine the explanation of classical data analysis science with new tools and new methodologies. Here we only provide a small selection of them, most of them following an open-access philosophy.

Textbook
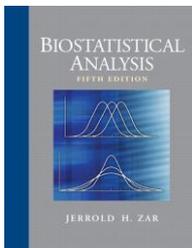
Biostatistical analysis: This book was britten by Jerrold H. Zar, associated professor in the Department of Biological Sciences in the Northern Illinois Univesity, in 1974. Since then it has gone through several editions, the latest in 2013, and remains a classic reference for Biostatistics. It reviews the basic statistical techniques required for biological investigation, from descriptive statistics to hypotheses tests or linear regressions. This is a highly recomended book as an introductory textbook and as a reference book for researchers consultance.
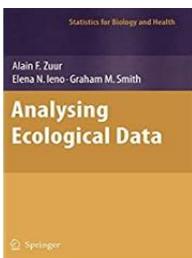
Analysing Ecological Data: Alain Zuur is a senior statistician in Scotland, and, together with Elena Leon, a senior marine biologist, both work in statistical consultancy and teaching. Graham M. Smith is a Senior Lecturer at Bath Spa University in the UK.

This book is a practical introduction in ecological data analysis, which provides the analysis path of real datasets. The case studies include from terrestrial ecology to marine biology, and the book covers data exploration, univariate methods, multivariate analysis, time series analysis and spatial statistical.
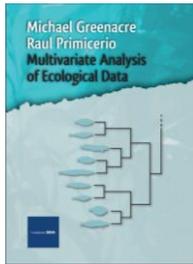
Multivariate Analysis of Ecological Data: This is a specific book to conduct multivariate analysis and ordination. The collaboration between Michael Greenacre, a Data Science professor in the Pomeu Fabra University in Barcelona, and Raul Primicerio, an ecologist in the Norwegian College of Fisheries Science, resulted in a clear and a rather non-mathematical but modern book. Fortunately, it is freely available at: www.multivariatestatistics.org

Open access online documentation

MIT Statistics for Applications: The Massachusetts institute of technology offers open courses on many topics, and this course is one of them. It is given by prof. Phillippe Rigollet, and it is composed of video lectures, lecture notes and assignments. All the material was uploaded in 2016 and is available at: https://ocw.mit.edu

Harvard Online Data Analysis Courses: Harvard University also offers online courses, most of them for free. Here, we will find general courses on data science, or more specifically focused on R basics, Linear Regression or Data Visualization. https://online-learning.harvard.edu/subject/data-analysis.

edX Online Courses: Harvard and MIT founded this platform for education and learning purposes in 2012. All the courses are MOOC (Massive Open Online courses). Here you will find courses from both universities but also from Berkeley University of California and The University of Texas System.

Coursera Online courses: Many universities and organizations offer their courses online through Coursera. Although most of them are completely free, they often require a donative to obtain a certificate. Courses such as **"Statistics with R", "Statistics with Python" or a more general approach such as "Biological Diversity (theories, Measures and Data sampling techniques)" and "Mathematical Biostatistics Boot Camp 1"** will certainly bring the user to a deep understand of the biological applications of statistics and data analysis.

Scientific publications

- Purvis, A., Hector, A. (2000). Getting the measure of biodiversity. *Nature* **405,** 212–219

- Izsák, J., Papp, L. (2000). A link between ecological diversity indices and measures of biodiversity *Ecol. Model.,* 130, pp. 151-156

- Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data*. Ecological Modelling, 240, 113–122.*

- Murtaugh, P. A. (2007). Simplicity and complexity in Ecological Data Analysis*. Ecology, 88(1), 56–62.*

- De'ath, G., & Fabricius, K. E (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology, 81(11), 3178–3192.*