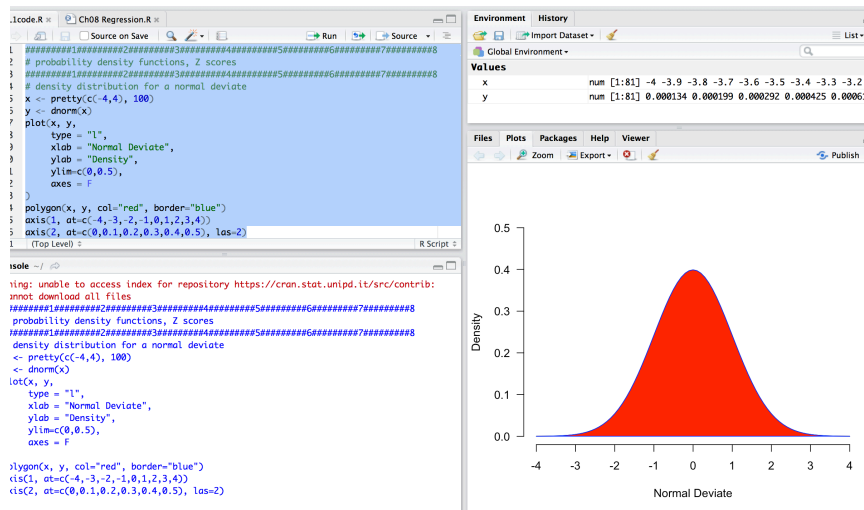




Statistical computing with R for applied biology.

1. Basic and intermediate methods



Prof. E. Parente

Dipartimento di Scienze – Università degli Studi della Basilicata

March 27 - July 3, 2019

Venue: room B6, Library building, Campus di Macchia

Romana

Course objective:

to provide an introduction to use of the R environment for graphical and statistical analysis in biology, biotechnology, medicine and food science and nutrition

Learning goals:

- **knowledge and understanding:** an introductory knowledge of principles of statistical computing for applied biology; working knowledge of basic methods for data wrangling, exploratory data analysis, statistical and graphical analysis
- **applying knowledge and understanding:** ability to develop code in R and use it for graphical and statistical analysis
- **making judgements:** ability to choose the graphical and statistical methods which are more appropriate in a given situation
- **communication skills:** ability to produce reports for the statistical and graphical analysis of experimental data in a variety of formats
- **learning skills:** ability to access and peruse literature and technical information on statistical computing

Prerequisites:

A BSc in Agriculture, Food Science, Chemistry, Biology, Biotechnology. At least 5 ECTS credits in Mathematics. Some statistics and some experience in coding may help. Ability to use spreadsheet software under Windows, MacOS or Unix/Linux operating systems. A knowledge of technical English language (B1 or B2 level is suggested).

Attendance. Only 10 highly motivated students can attend the course (and get their exercises graded). Further students can be accepted but their exercises and reports will not be graded.

Grading. To obtain full credits (4-5 ECTS) the students must turn in a report (in Word, pdf or .html format, generated using R markdown) within 1 month from the end of the course. The report shall describe in full (including code) the descriptive and inferential statistical analysis and the graphical analysis of one of their own experiments. A suitable dataset from a R package can be used.

Course content:

Lectures (24-32 h). 1. An introduction to statistical analysis and data science (2 h). 2. The R environment (1 h). 3. Importing data, data structures in R (3 h). 3. Data visualisation with base functions and ggplot2 (3 h). 4. Data wrangling, tidying and reshaping (2 h). 6. Numerical and graphical summaries of data. Generating reports with R markdown and knitr (3 h). 7. Group comparisons with t-tests and non-parametric tests; one-way ANOVA and multiple mean comparisons; tests of independence and association for contingency tables; power analysis (3 h). 8. Experimental design; ANOVA and ANCOVA (4 h). 9. Covariance, correlation and linear regression. (3 h) **Bonus lectures** 10. Factorial designs and empirical model building (4 h). 11. Non-linear regression (4 h)

Practicals: 16 h. Writing and running code, generating reports using datasets from R or case studies

Suggested readings.

Gacula, M., Singh, J., Bi, J., Altan, S. 2008. Statistical methods in food and consumer research. Academic Press.

Kabacoff, R.I. 2015. R in action. 2nd edition. Manning.

Grolemund, G., Wickham, H.. 2017. R for Data Science. <http://r4ds.had.co.nz>

McDonald J.H. 2014 Handbook of Biological Statistics 3rd ed. <http://www.biostathandbook.com>