Commented literature list
Module "Tools for complex data analysis"
Ralf B. Schäfer, University Koblenz Landau 2018/19

Note that this list represents a more comprehensive list than the literature used in the course. Literature that is heavily used in the course is indicated with an asterisk (*).

1. Mainly univariate data analysis and R

Crawley, M. J. 2015. Statistics: An Introduction Using R. Wiley & Sons, Chichester.

Good and comprehensive introduction to univariate statistics and R.

Dormann, C. F. 2017. Parametrische Statistik Verteilungen, maximum likelihood und GLM in R. Springer: Berlin; Heidelberg.

In german. Excellent introduction into ecologically-relevant methods of univariate statistics.

Efron B. & Hastie T. (2016) Computer age statistical inference: algorithms, evidence, and data science. Cambridge University Press, New York, NY. Free to download: https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf

Quite some overlap with the other books of these authors (see below) and rather brief on most topics, but provides a nice overview of the past, present and future development of statistical and machine learning techniques. Recommended if you want to obtain an overview of the major topics in data science and their history.

Faraway, J. J. 2002: Practical Regression and ANOVA using R.

Good and simple introduction to linear models (regression analysis and ANOVA)! Outdated, but can be downloaded for free:

cran.r-project.org/doc/contrib/Faraway-PRA.pdf

http://www.maths.bath.ac.uk/~jjf23/book/

An updated version is available as:

*Faraway, J.J. 2015. Linear models with R. CRC Press: Boca Raton, Fl.

Field, A.; Miles, J.; Field, Z. 2013. Discovering statistics using R. Repr. ed.; Sage: Los Angeles, California.

A very readable book for beginners that introduces to all relevant aspects of R and univariate statistics. Written in a rather casual style.

*Fox, J. 2015. Applied Regression Analysis and Generalized Linear Models.

3rd edition. Sage Publications, Thousand Oaks, California.

The book has over 800 pages and covers all subjects in the field of linear models in sufficient detail. A good standard reference.

*Fox, J. & Weisberg, H.S. 2019. An R Companion to Applied Regression. 3rd edition. Sage Publications, Thousand Oaks, California.

Related to Fox (2015), describes the implementation of statistical methods in R.

*Harrell F.E. 2015 Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. 2nd edition. Springer, New York.
A very profound treatment, very up to date – highly recommended if you have to apply statistical modelling.

Hastie, T.; Tibshirani, R.; Friedman, J. 2017. The elements of statistical learning: data mining, inference, and prediction. 12th printing with corrections; Springer: New York, NY. Free to download:

https://web.stanford.edu/~hastie/ElemStatLearn/

A thorough treatment of a range of topics including multivariate approaches. Contains many methods of modern data analysis, but rather targets advanced readers.

Hastie, T.; Tibshirani, R.; Wainwright, M. 2015. Statistical learning with sparsity: the lasso and generalizations. CRC Press: Boca Raton. Free to download:

https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS.pdf

Focuses on the application of the lasso and related techniques in univariate and multivariate data analysis.

Hothorn, T. and B. Everitt 2014. A Handbook of Statistical Analyses Using R. 3rd edition. Chapman & Hall/CRC, Boca Raton, FL.

A good, but on some subjects short introduction to statistical methods. The book is written in a very accessible style.

*James, G., Witten, D., Hastie, T., and Tibshirani, R. 2017. An introduction to statistical learning: with applications in R ; Springer: New York. Free to download: http://www-bcf.usc.edu/~gareth/ISL/

Concentrates on the application of the topics tackled in Hastie et al. 2017 and describes their implementation in R.

Logan, M. 2010. Biostatistical Design and Analysis Using R. Wiley-Blackwell, Hoboken NJ.

Well structured book for all fields of biostatistics and their implementation in R.

*Maindonald, J. and J. Braun 2010. Data Analysis and Graphics Using R. 3rd edition. Cambridge University Press, Cambridge.

An excellent book regarding the introduction to data analysis as well as to R with 560 pages. Most notable is its applied perspective.

*Matloff, N. S. 2017. Statistical regression and classification: from linear models to machine learning. CRC Press: Boca Raton, 2017.

An excellent book on linear models, their extensions and classification. Covers most of the topics (linear model, GLM, PCA, CART) discussed in this course as well as their implementation in R. Gives advice what to use in practice and provides extensive case studies. Comes with the right amount of mathematics to understand the background.

Sachs, L., and J. Hedderich. 2015. Angewandte Statistik: Methodensammlung mit R, 12th edition. Springer, Berlin.

In german! The former standard text book, now with added R-code. A thorough treatment of statistics - but not strongly focused on R. With exercises, over 900 pages. Represents an alternative to Crawley (2015) for those who struggle with English.

*Sheather, S. 2009. A modern approach to regression with R, 1st edition. Springer, New York.

An excellent book on linear models in R with a very thorough treatment of model development, diagnostics etc. Does not require much prior knowledge and fulfils the expectations set out with the title.

2. Mainly multivariate data analysis and R

*Borcard, D.; Gillet, F.; Legendre, P. 2018. Numerical ecology with R. 2nd edition. Springer: New York, NY.

R implementation of most of the topics of the classical textbook (see below). Highly recommended for ecological data analysis.

Everitt, B. S. and T. Hothorn 2011. An Introduction to Applied Multivariate Analysis with R. Springer, Berlin.

Covers all important topics of multivariate statistics, though with a limited

depth (less than 300 pages). Works best for people who know which method they want to use and who need information about how to implement it in R.

*Handl, A & Kuhlenkasper, T. 2017. Multivariate Analysemethoden. Theorie und Praxis mit R, 3rd edition. Springer, Berlin.

In german. Generally a very understandable book on multivariate data analysis with R. However, very mathematical and not targeted at environmental data analysis. A recommendation for those who want to understand the mathematics behind some multivariate methods.

Varmuza, K. & Filzmoser, P. 2009 Introduction to multivariate statistical analysis in chemometrics. CRC Press/Taylor & Francis: Boca Raton, Fla., p 321.

Includes many of the key subjects of multivariate data analysis in R - deals primarily with the application in chemistry.

3. Multivariate statistics in general

Hartung, J. 1999. Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik, 6th edition. Oldenbourg, München; Wien.

In german. The style is old-fashioned, but it is the best german textbook for multivariate statistics in general. Very mathematical, not suitable for beginners. In 2007 a new but unchanged edition was published, probably the layout was improved.

Johnson, R. A., and D. W. Wichern. 2007. Applied Multivariate Statistical Analysis, 6th edition. Prentice Hall.

Together with Hartung (1999) almost all fields are covered. Also very mathematical, not suitable for beginners.

4. Environmental, Ecological and Biostatistics

Fox, G.A., Negrete-Yankelevich, S. & Sosa, V.J. eds 2015 Ecological statistics: contemporary theory and application. Oxford University Press, Oxford.

Deals with different topics relevant to statistical analysis of ecological data, largely univariate. Covers topics such as how to analyse phylogenetically correlated data, to conduct meta-analyses and to analyse spatial variation.

*Legendre, P., and L. Legendre. 2012. Numerical Ecology. Elsevier, Amsterdam.

Certainly the standard textbook for ecologists. A beginner may need some time to get used to the style. Nevertheless the most recommendable book in the field of primarily multivariate data analysis in ecology.

Leyer, I.; Wesche, K., Multivariate Statistik in der Ökologie: Eine Einführung. Springer: Berlin, 2008.

In german. Introduction to multivariate statistics for ecologists that largely omits the mathematical basics. Suited for beginners, but lacks depth.

Newman, M.C. 2012. Quantitative Ecotoxicology; CRC Press: Boca Raton, FL.

The standard text book for quantitative approaches in ecotoxicology. Some examples have been implemented into R code by Eduard Szöcs. Ecotoxicologically complete, but misses out on important statistical topics such as GLMs.

Kindt, R.; Coe, R. 2005. Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. Nairobi, Kenya: World Agroforestry Centre.

Freely available book on data analysis in biodiversity research. Covers most of multivariate methods (but not PERMANOVA), but is outdated. Freely available:

http://www.worldagroforestry.org/downloads/publications/PDFs/B13695.pdf

Piegorsch, W.W. And Bailer, J.A. 1997: Statistics for Toxicology and Environmental Biology. Chapman & Hall, Cornwall.

Comprehensive overview of univariate biostatistics with a special emphasis on dose-response-relationships. Quite mathematical.

Quinn, G.P. and Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.

Excellent book on biostatistics that comprises univariate as well as multivariate statistics and experimental planning.

Sparks, T. (ed.) 2002: Statistics in Ecotoxicology. John Wiley & Sons, Chichester.

Treats all statistical aspects of ecotoxicology, but in parts relatively short and rather outdated.

Wildi, Otto (2013): Data analysis in vegetation ecology. 2nd edition. Chichester: Wiley-Blackwell.

Introduction to many relevant analyses for vegetation data, a bit short though. Covers ordination, grouping (cluster analysis) among other topics that are beyond the scope of this course.

Zar, J. H. 2013. Biostatistical Analysis. Prentice-Hall, Englewood Cliffs, New York. Fifth edition.

With 750 pages the classic and the most comprehensive textbook in the field of biostatistics. It touches subjects of multivariate statistics (MANOVA) and takes the fear out of math.

*Zuur, A. F., Ieno, E. N. and G. M. Smith 2007. Analysing Ecological Data. Series: Statistics for Biology and Health.

Specifically tailored to the analysis of ecological data. All important methods are presented – for this reason some topics are not treated in detail. There is a special section on case studies, demonstrating strategies of data analysis. Most analyses are done in R and the code is available. Would benefit from an update though.

5. R in general

Adler, J. 2011. R in a nutshell: A desktop quick reference. O'Reilly: Beijing.

A fantastic reference book for all aspects of the R language, which tackles many issues or refers to different sources otherwise.

Crawley M. J. 2012. The R Book. Second Edition. Wiley & Sons, Chichester.

This book represents an excellent handbook for R. It covers almost all fields of data analysis, but no ordination and only touches cluster analysis. Of course not all topics are treated exhaustively, but you will find some introductory information for many R-related data analysis topics in this book.

Murrell, P 2011. R Graphics. 2nd edition. Chapman & Hall/CRC, Boca Raton, FL.

Excellent book that gives an overview on the possibilities of generating and modifying graphs and maps.

Rahlf, T., 2017. Data visualisation with R: 100 examples. Springer, Cham.

Excellent book providing examples with R scripts for a wide range of plots. A Also available in german (Rahlf, T., 2018. Datenvisualisierung mit R: 111 Beispiele. Springer, München).

Venables, W. N., and B. D. Ripley. 2003. Modern Applied Statistics with S, 4th

edition. Springer, New York.

Probably the most cited R-book. High-level book, for beginners only suitable to a limited extent. Nevertheless all fields of data analysis are covered.

Wickham, H. 2016. Ggplot2: elegant graphics for data analysis, 2nd edition. Springer: New York: Springer.

Introduction to the ggplot2 graphics package, which is arguably the most flexible and elegant graphics tool in R, though not as intuitive as base graphics.

6. Web sources

The web is full of tutorials, scripts and forums. Several materials can be found on the pages of the R-project:
www.r-project.org
Moreover, you can find videos for beginners, for example:
http://dist.stat.tamu.edu/pub/rvideos/
http://www.learnviaweb.com/videos/r-programming/

A very good website that demonstrates all kind of statistical methods in R:
http://www.statmethods.net/
The success of the site has lead to a book that can be highly recommended:
*Kabacoff, R. 2015. R in Action. 2nd edition. Data Analysis and Graphics with R. Manning Publications.

Several publications on novel methods of data analysis in R can be found in the Open access „Journal of Statistical Software“:

http://www.jstatsoft.org/

Many blogs contain information and news on ecological data analysis (e.g. from the bottom of the heap). They can be accessed through R-bloggers:
http://www.r-bloggers.com/

The following web forum deals with advanced problems with R programming:
https://stackoverflow.com/tags/r/info
But many questions have already been answered, check also the mailing list archives:
http://tolstoy.newcastle.edu.au/~rking/R/
http://dir.gmane.org/gmane.comp.lang.r.general

Finally, there are information and introductions to multivariate analysis methods on the following webpages:
http://ordination.okstate.edu/
http://ecology.msu.montana.edu/labdsv/R/
http://pbil.univ-lyon1.fr/ADE-4/home.php?lang=eng