FISH 560: Applied Multivariate Statistics for Ecologists

Autumn 2018

COURSE SYLLABUS
INSTRUCTOR: Julian D. Olden

Office Location: Fisheries Science Blg., Room 318A
Contact information: olden@uw.edu, 616-3112
Class location: Fisheries Science Blg. Room 136
Office hours: Wed 12:30-1:30 pm
Class hours: Wed 1:30-3:20 pm, Fri 12:30-2:20 pm
Web page: https://canvas.uw.edu/courses/1220987
Prerequisite(s): QSCI 482 or equivalent or permission from instructor

COURSE DESCRIPTION

With recent advances in data collection technology and ambitious field research, ecologists are increasingly calling upon multivariate statistics to explore and test for patterns in their data. The goal of this course is to introduce graduate students in the ecological sciences to the multivariate statistical techniques necessary to carry out sophisticated analyses and to critically evaluate scientific papers using these approaches. This is a practical and hands-on course emphasizing the analysis and interpretation of multivariate analysis, and covers the majority of approaches in common use by ecologists. The focus of the course is on the conceptual understanding and practical use of the methods (not the matrix algebra), with the singular hope of de-mystifying the "alphabet soup" of multivariate analysis.

We will cover the three main categories of multivariate analysis that are common in ecology: (i) clustering, (ii) ordination and (iii) statistical tests of hypotheses. The intent of this course is to provide you with the following: (1) an introduction to the use of multivariate statistics in ecological research; (2) a conceptual organization of the various multivariate techniques, with respect to the types of research questions and data sets appropriate for each technique; and (3) a working understanding of how to use and interpret the results of each technique, including a conceptual overview, list of assumptions, diagnostics for assessing the assumptions, mechanics of performing the analysis using the R package, and how to interpret the statistical output of the analysis.

METHOD OF INSTRUCTION

Lectures/labs – Lectures will integrate both theoretical aspects of multivariate statistics and provide solutions and interpretations from the R package. For each topic there will be a formal lecture followed by a computer-based lab where R will be used to analyze ecological data using the particular multivariate technique.

Pop-quiz – A portion of your grade is based on a pop-quiz that will be administered at some point during the quarter. This quiz is used to test your understanding of the material, and promote self-evaluation of your progress.

Final report and peer review – A significant portion of your grade is based on a final written paper and peer review of other class members' papers. The final paper will consist of a multivariate analysis of a data set. The nature of the question, the source of the data, and the kinds of analysis employed are flexible. The requirement is that the data and analysis must address one or more specific biological hypotheses, which are to be tested using an appropriate method(s) of multivariate analysis. The primary goal is a coherent scientific paper, not excessive number crunching.

DATASETS

Personal dataset – A primary goal of this course is to provide you the opportunity to get better acquainted with your own data. The data set may be your own, one obtained from the literature or one provided by the Instructor. Ideally you should use data that you have collected or are otherwise somewhat familiar with. The data set should be one or more matrices of entities × attributes (e.g., samples × species, species × characteristics of species, sites ×

environmental factors, etc.). The only data requirements are that it be adequate to test the hypotheses addressed in your final report. If you do not have access to a multivariate dataset, then I would be please to provide one.

Class dataset – Even if you do have a multivariate dataset, it is unlikely to be suitable for all the techniques covered in class. To address this issue, a common dataset is provided to all students at the beginning of the quarter. This dataset is in addition to your own personal dataset that forms the basis for your final report. Using the class data you will be able to conduct all the statistical approaches covered in the class. Moreover, this dataset will serve as the basis for the R exercises. You will be expected to work with both your own dataset and the class dataset.

TEXTBOOK(S) AND REQUIRED TOOLS OR SUPPLIES

There is no required text for this course, however I recommend:
McGarigal, K., S. Cushman, and S. Stafford. 2000. Multivariate Statistics for Wildlife and Ecology Research. Springer.

Other statistical texts that are likely to be helpful (in order of value based on my personal experience) include:
Legendre, P., and L. Legendre. 2012. Numerical Ecology. 3rd edn. Elsevier Scientific.
Gauch, H.G. 1982. Multivariate Analysis in Community Ecology. Cambridge University Press.
Manly, B.F.J. 2004. Multivariate Statistical Methods: a primer. Chapman and Hall.
Borcard, D., F. Gillet, and P. Legendre. 2011. Numerical Ecology with R. Springer.
Digby, P.G.N. and R.A. Kempton. 1987. Multivariate Analysis of Ecological Communities. Chapman & Hall.
Jongman, R.H.G., C.J.F. ter Braak, and O.F.R. van Tongeren. 1995. Data analysis in Community and Landscape Ecology. Cambridge University Press.
Pielou, E.C. 1984. The interpretation of ecological data: a primer on classification and ordination. Wiley-Interscience.

GRADING PLAN

The Standard UW Numerical Grading System will be used according to the breakdown provided below. See: http://www.washington.edu/students/gencat/front/Grading_Sys.html for university description.

Task	Due date	% of grade
Participation in lecture and lab	Never-ending	10%
One-page proposal	October 17th	10%
Pop-quiz	Who knows	10%
Final paper	December 7th	50%
Peer-review reports	December 14th	20%

CLASS SCHEDULE (tentative – see Canvas website)

26 Sept — Course overview – The beast we call "multivariate statistics"
28 Sept — Data screening
3 Oct — Multivariate resemblance
- Modes of analysis, analytical spaces
- Similarity coefficients (binary, categorical, quantitative)
- Distance coefficients
- Coefficients of dependence
- Choice of coefficients

5 Oct — Cluster analysis
- Introduction, diversity of approaches
- Hierarchical agglomerative clustering (e.g., linkage, UPGMA)
- Hierarchical divisive clustering (e.g., TWINSPAN, K-means)

10 Oct — Cluster analysis
- Cluster diagnostics, limitations, and recommendations

	Presenting results from cluster analyses: The dos and don'ts!
12 Oct	Direct Ordination
	Principal component analysis (PCA)
	Introduction, purpose, Shepard diagrams
	Computing eigenvalues, principal components
	Covariance vs. correlation, meaningful components, misuses
17 Oct	Direct Ordination
	Principal component analysis (PCA), *continued*
19 Oct	Direct Ordination
	Principal coordinate analysis (PCoA)
	Non-metric multidimensional scaling (NMDS)
24 Oct	Direct Ordination
	Correspondence analysis (CA)
	Detrended correspondence analysis (DCA)
	Presenting results from ordination analyses: The dos and don'ts!
26 Oct	A Critical Review of Clustering and Ordination
31 Oct	Indirect Ordination
	Redundancy analysis (RDA)
	Canonical correspondence analysis (CCA)
	Canonical correlation analysis (CCorA)
	Canonical analysis on principal coordinates (CAP)
2 Nov	Indirect Ordination work period
7 Nov	Indirect Ordination
	Partial RDA and CCA
	Hierarchical RDA and CCA
9 Nov	Principal Response Curves (PRC)
14 Nov	Fourth-corner and RLQ methods
	Fourth-corner solution
	RLQ methods
	CATS (community assembly by trait selection) regression
	Model-based methods
16 Nov	TBA – Potentially Threshold Indicator Taxa ANalysis (TITAN)
21 Nov	Classification of Groups
	Discriminant Function Analysis
	Classification and Regression Trees
23 Nov	NO CLASS: Thanksgiving Holiday
28 Nov	Testing for Similarities and Differences among Groups
	Analysis of similarity (ANOSIM)
	Multi-response Permutation Procedure (MRPP)
	Permutational MANOVA (perMANOVA)
	Permutation test of multivariate dispersion
30 Nov	Testing for Associations among Matrices
	Mantel tests
	Procrustes Analysis
5 Dec	Emerging multivariate approaches
	Moran's eigenvector maps (MEM)
	Principal Coordinate Analysis of Neighbour Matrices (PCNM)
	Asymmetric eigenvector maps (AEM)
	Multi-scale ordination (MSO)
	Multiscale codependence analysis (MCA)
7 Dec	TBA

FISH 560: Final Report and Peer Review

Autumn 2018

PURPOSE

1. Improve your ability to select appropriate multivariate methods for answering ecological questions.
2. Improve your understanding of how the choice of method may influence your results.
3. Advance your graduate research projects.
4. Broaden your experience with multivariate software.

DATASET

The data set may be your own, one obtained from the literature or one provided by the Instructor. Ideally you should use data that you have collected or are otherwise somewhat familiar with. The data set should be one or more matrices of entities × attributes (e.g., samples × species, species × characteristics of species, sites × environmental factors, etc.). The only data requirements are that it be adequate to test the hypotheses addressed in your final report. If you do not have access to a multivariate dataset, then I would be please to provide one.

ANALYSIS

Methods of analysis should be chosen to be compatible with hypotheses and data. Procedures can be either exploratory or inferential.

ONE-PAGE PROPOSAL

Your proposal should include the following items:

- Title
- Name
- Date
- Brief background
- Questions or objectives for the analysis
- Source of the data
- Data Structure
    - What is the sample unit?
    - What is each data matrix?
    - Describe any hierarchical structure in the data (e.g., subsampling with plots which are then averaged to a single plot value) and a statement of what level(s) in this structure you plan to do your analyses.
    - If aggregation of the raw data is required to construct a data matrix, describe how you will do that.

FINAL PAPER

The final paper should be written in standard scientific format, and it will be considered for web publication in the virtual journal, the *Electronic Journal of Applied Multivariate Statistics* (EJAMS). Dr. Julian Olden is the Chief-in-Editor of EJAMS and will post all accepted paper on the web for future class use, with your permission. The journal will be available only to students registered in this class. See http://hdl.handle.net/1773/19723.

Format:

- **Title**
- **Author**
- **Institutional affiliation** (Department or Program, University)
- **Email address**
- **Abstract**: 200 word summary (separate page)

- **Introduction**: Background to the problem with enough information to allow me to understand what question you're asking, why you're asking it, and why you find it interesting. Explicit statements of research hypotheses/objectives.
- **Methods**: Description and justification of *methods* of analysis chosen.
- **Results:** Tabular and text summaries of results, including descriptive statistics, test statistics, degrees of freedom, probabilities, and assessments of statistical significance.
- **Discussion**: Biological interpretation of the results, contribution to science and implications.
- **Acknowledgements**
- **References** (follow *EJAMS* guidelines)
- **Tables**
- **Figures:** Please include all figures at the end of the document. Select your figures sparingly. Excess figures should be included in an appendix ONLY when necessary.

Other details:

- Upload your final paper (Word format) on the Canvas website by the deadline. Name the Word document by your last name and first two initials, e.g. oldenjd.doc
- Follow the conventions of Standard Written English, which should include correct spelling, grammar, punctuation, capitalization, paragraph structure, and sentence construction
- 15-page maximum (not including figures), 10-page minimum
- Times Roman 10 point font
- Headings in 12 point bold font
- Single spacing
- 1 inch margins on all sides

PEER-REVIEW OF FINAL PAPERS

Each student will be randomly assigned to review two submitted papers. Reviewers are required to write a 1-page report and must decide to either accept (as is) or reject the paper. Please qualify your decision. For example, you might point out if a paper has a few misleading flaws that you don't think should be published for others to follow, but the paper is otherwise in good shape. Note that the average acceptance rate for EJAMS is 20-30%. Please keep this in mind when you are rendering your decision. Be courteous and respectful in your reviews. A good practice for reviewing papers is to pretend that you are writing the review for a colleague who is a friend. You should provide constructive criticism, but it should be done in a friendly, gentle, non-hostile way. The Chief-in-Editor will make his final decision based on the reviewers' recommendations, and his own review of the paper. All peer-reviews must be uploaded to the Canvas website by the deadline.