

MultiStaR

E-Learning-Module zur Multivariaten Statistik in der Ökologie mit R

Michael Rudner

Michael.Rudner@biologie.uni-freiburg.de

Abteilung Geobotanik, Institut für Biologie II, Albert-Ludwigs Universität Freiburg
Schänzlestr. 1
79104 Freiburg

Hiltrud Brose

hiltrud.brose@wsl.ch

Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL Räumliche Ökologie
Zürcherstrasse 111
CH-8903 Birmensdorf

urn:nbn:de:0009-5-22419

Zusammenfassung

Multivariate Methoden stellen ein wesentliches Instrumentarium zur Datenanalyse in der Ökologie dar. Sie werden in der Ökologie häufig eingesetzt und sind seit langem Gegenstand der Lehre in der Abteilung Geobotanik der Universität Freiburg. In den letzten Jahren wurde als Werkzeug das Programm R eingeführt. R ist ein frei verfügbares, kommandozeilenorientiertes Statistikprogramm, das für eine Reihe von Betriebssystemen angeboten wird (R-Development Core-Team 2007). Das Programm befindet sich in rascher Entwicklung (derzeit Version 2.10) und wird zunehmend auch von Ökologen eingesetzt. Bislang existiert kein deutschsprachiges Lehrbuch zur Anwendung multivariater Methoden mit R. Mit MultiStaR wird versucht, diese Lücke zu schließen und den Studierenden Lernmaterialien an die Hand zu geben, die Übungen mit dem eigentlichen Analysewerkzeug mit einschließen.

Stichwörter: e-learning, statistics, mathematical application, multivariate methods, ecology, R-Statistical computing, Statistik, mathematische Anwendung, multivariate Methoden, Ökologie, Statistikprogramm R

Abstract

Multivariate methods form an essential tool for data analysis in ecology, where they are frequently employed. At the Geobotany Department of the University of Freiburg they have been taught for many years. In recent years the program R was introduced as an analysis tool. The statistics program R is command line oriented. It is freely available for several operating systems. The program is developing rapidly (currently version 2.10), being employed increasingly by ecologists. Up to now, no German textbook was available on the application of multivariate statistics using R. With MultiStaR we are seeking to close this gap and provide learning materials to the students, including exercises using the actual analysis tool.

Keywords: e-learning, statistics, mathematical application, multivariate methods, ecology, R-Statistical computing

Thematische Ausrichtung

Das Projekt wurde auf Anregung von Studierenden entwickelt, um die Lehre in der Geobotanik in einem Bereich zu verbessern, in dem sehr komplexe Lehrinhalte vermittelt werden und in dem es weitere Materialien zur Vertiefung bedarf. Die fachspezifische Ausbildung in speziellen statistischen Analyseverfahren mit engem Bezug zur Ökologie soll so gestärkt werden. MultiStaR umfasst einen Kanon von zwölf E-Learning-Modulen zur Verwendung von multivariaten statistischen Analyseverfahren zur Bearbeitung geobotanischer bzw. ökologischer Fragestellungen (Abb.1). Neben wichtigen Klassifikations- und Ordinationsverfahren werden auch wesentliche Schritte in der Vorbereitung der Analysen, wie die Berechnung von Distanzmaßen, behandelt. Eine kurze Einführung in das Statistikprogramm R ist ebenfalls enthalten, da die Analysen von den Lernenden parallel zur E-Learning-Einheit in R durchgeführt werden sollen. Das Projekt ist mit einem geplanten Zyklus von E-Learning-Modulen zur univariaten Statistik in der Ökologie (UniStaR) verzahnt.

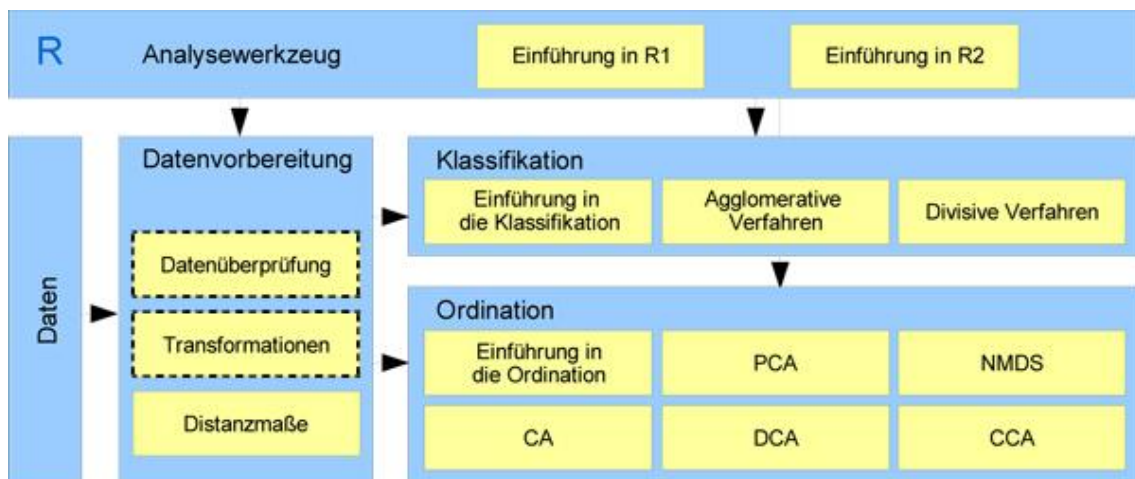


Abb. 1: Übersicht zu den Modulen in MultiStaR. Blaue Kästen: Themenbereiche; gelbe Kästen: E-Learning-Module; gerissener Rand: im Rahmen einer Modulreihe zur univariaten Statistik (UniStaR) geplant. (PCA: Hauptkomponentenanalyse, NMDS: Nichtmetrisches mehrdimensionales Skalieren, CA: Korrespondenzanalyse, DCA: Entzerrte Korrespondenzanalyse, CCA: Kanonische Korrespondenzanalyse)

Didaktischer Ansatz

Didaktisches Ziel der E-Learning-Module ist es, die Studierenden in die Lage zu versetzen, selbstständig ökologische Daten mit multivariaten Verfahren zu analysieren, um diese Fertigkeit bei der Erstellung von Bachelor- und Master-Arbeiten anzuwenden.

Dabei werden die folgenden Lehr-/Lernziele verfolgt:

- Kenntnisse und Fertigkeiten in der Anwendung von R (technische Voraussetzung)

- Kenntnisse der Methoden (v.a. Klassifikation und Ordination), deren Eignung und Grenzen
- Fertigkeiten in der Datenanalyse in allen Schritten vom Dateneinlesen, -überprüfen, -transformieren über die Analysen bis hin zur grafischen Aufbereitung und Interpretation der Ergebnisse.
- Fähigkeit zur kritischen Beurteilung der Anwendung dieser Methoden in der Fachliteratur.

Methodische Umsetzung:

Um ein umfangreiches Verständnis von komplexen Rechenmodellen zu erreichen, muss neben der Darstellung der Modellstruktur auch die Möglichkeit gegeben werden, einzelne Berechnungsschritte selbst im Detail nachzuvollziehen. Dies erfordert ein hohes Maß an Interaktivität, die Verwendung realitätsnaher Datensätze und praxisnaher Auswertungswerkzeuge (Mitesser et al. 2004).

Die Konzeption der Module erfolgte daher anhand der folgenden Gesichtspunkte:

- Ein Modul bildet eine kompakte thematische Einheit (Bearbeitungsdauer ca. 30 Minuten)
- Darstellung der wesentlichen Inhalte in einem einführenden Text
- Veranschaulichung der Sachverhalte anhand von Beispielen
- Kleine Übungen im Modul, um sich Verfahrensschritte zu erarbeiten (Selbstüberprüfung)
- Übungen in R (Aufgabenstellung, Datensätze und R-Code werden im Modul bereitgestellt)
- Aufbau der Module nach dem Muster:
 - Einführung: Titel, Lernziel, Bearbeitungszeit, Voraussetzungen
 - Inhalt: Theorie und Übungen, Anleitung für Übungen in R
 - Zusammenfassung: Aufgreifen des Lernziels, kurzes Rekapitulieren der Modulinhalte
 - Anhang: alle verwendeten R-Skripte, Quellenverzeichnis, Anleitung zum Modul
- Das Layout ist so gestaltet, dass sich Theorieteil und Umsetzung in R klar unterscheiden lassen (Abb. 2). Dadurch ist es möglich, eigene Schwerpunkte in der Bearbeitung des Moduls zu setzen und sich z.B. nur auf die Umsetzung in R zu konzentrieren oder umgekehrt.

NMDS - Nonmetric MultiDimensional Scaling
 Berechnung
 Erstellt von: Michael Rudner & Hiltrud Brose, Abteilung Geobotanik, Institut für Biologie II, Albert-Ludwigs-Universität Freiburg

Modulrahmen Methode **Berechnung** Grafik Abschluss Anhang

Mehrfachberechnung
 Die Eigenschaften der NMDS, Sicherheit, beim ersten Lauf eine optimale Konfiguration zu finden, für eine Analyse mehrere Programmläufe mit verschiedenen Startkonfigurationen durchzurechnen. Dafür wurde in R die Funktion `metaMDS` im Paket `vegan` entwickelt. Diese Funktion greift intern auf die Funktion `isoMDS` zu, weshalb das Paket `MASS` ebenfalls eingebunden werden muss.

Parameter von <code>metaMDS</code>	Bezeichnung	Standard
Datensatz	<code>comm</code>	-
Distanzmaß	<code>distance</code>	"bray"
Dimensionen	<code>k</code>	2
Max. Anzahl an Läufen	<code>trymax</code>	20

R-Anweisungen

```
library(MASS) # isoMDS
library(vegan) # dune, vegdist
data(dune) # Daten einladen
# Mehrfachberechnung der NMDS
dune.N <- metaMDS(dune, distance="bray")
dune.N # Aufruf des Ergebnisses
plot(dune.N) # Ordinationsdiagramm
```

Beispiel-Code für die Mehrfachberechnung einer NMDS und das Zeichnen eines Ordinationsdiagramms.
 (Sie benötigen die Pakete `vegan` und `MASS`.)

In dieser Übung können Sie maximal 0 Punkte erreichen. Nach dem 3. Versuch können Sie die Lösung ansehen. Dies ist Ihr 1. Versuch.

Eigenschaften der Funktion `metaMDS`

Welche Parameter müssen Sie beim Aufruf setzen?

- Distanzmaß
- Datensatz
- Dimensionen
- Anzahl von Läufen

Welche Startkonfigurationen werden in `metaMDS` verwendet?

- Immer die gleiche Konfiguration
- Ergebnis einer Ordination und zufällige Konfigurationen
- Nur zufällige Konfigurationen

Funktionsaufruf

Sie wollen den Dünen Datensatz mit einer 3-dimensionalen NMDS analysieren. Als Distanzmaß verwenden Sie die Bray-Curtis-Distanz.

Welcher Aufruf stimmt?

`x <- metaMDS(dune,k=3)`

Sie wollen den Stresswert ablesen. Wie geht das?

Abb. 2: Ansicht einer Modulseite in MultiStaR. Die Modulseite enthält eine Beschreibung des Themas, die zugehörigen R-Anweisungen und eine Übung, mit der die Nutzer den Lernerfolg überprüfen können. Die Navigation im E-Learning-Modul erfolgt über das Menü, das hier aufgeschlagen ist. Blättern ist über die Pfeiltasten oben rechts möglich.

Integrationsszenario in die Lehre:

Hauptziel: Einsatz begleitend zur Präsenzlehre in einem fortlaufenden Kurs mit 2 SWS (Abb. 3). Die Module sollen zur Vorbereitung der Kursinhalte oder Festigung des Erlernten zwischen zwei Kursterminen selbständig bearbeitet werden. Dies soll dazu führen, dass im Kurs mehr Zeit für Übungen auch an großen Datensätzen bleibt, da die Einführung in die Theorie und die Umsetzung in R bislang viel Zeit in Anspruch nehmen. Eine Besprechung der Arbeit mit den Modulen soll immer zu Beginn des folgenden Kurstermins stattfinden.

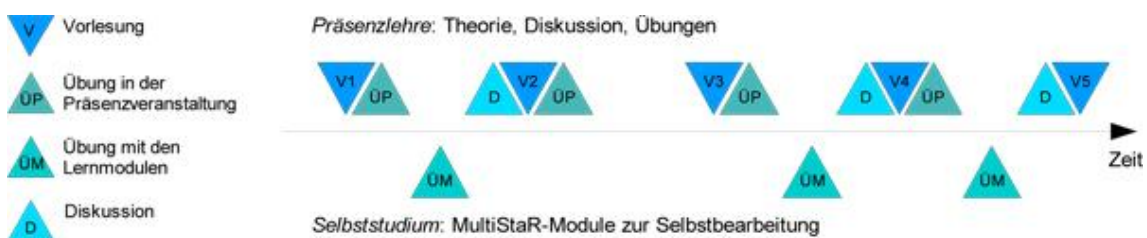


Abb. 3: Szenario zur Integration der E-Learning-Module in eine Lehrveranstaltung. Die zwei bis drei Komponenten einer Präsenzveranstaltung finden jeweils an einem Termin statt. (Grafik in Anlehnung an Petschenka & Kerres 2004)

Weiteres Ziel: Material bereit zu stellen zur selbständigen Auffrischung des Gelernten, wenn der Einsatz in einer Studienabschlussarbeit (z.B. Masterarbeit) bevorsteht. Die Gesamtheit der bereit gestellten Unterlagen umfasst neben den E-Learning-Modulen ein R-Skript zu multivariaten Methoden und eine Sammlung von Methoden-Steckbriefen.

Technischer Ansatz

Die Umsetzung der MultiStaR-Module erfolgt mit Flash in WebKit, dem Angebot der Servicestelle E-Learning der Albert-Ludwigs-Universität Freiburg (Gossmann et al. 2007). Den Nutzer/inne/n wird paralleles Arbeiten in R empfohlen. Auf den Computern in den genutzten Kursräumen muss daher neben einem Internetbrowser mit Flash-Plugin auch R installiert sein. Die Nutzung ist für alle Studierenden der Universität Freiburg über ein Angebot in der elektronischen Lernplattform CampusOnline möglich. Neben dem Einsatz in Lehrveranstaltungen der Biologie werden die Module auch permanent zur eigenständigen Nutzung angeboten. Dieses E-Learning-Angebot wird online tutoriert. Für die Nutzung durch Ökologen in anderen deutschsprachigen Universitäten werden die Module auf einem Server der Universität Freiburg angeboten (vgl. <http://www.biologie.uni-freiburg.de/geobotanik/Multistar/>).

Ausblick

Die folgenden Erweiterungen des Angebots werden angestrebt:

- Erweiterung des behandelten Themenspektrums wie z.B. partielle Ordination, Varianzaufteilung
- Modul zur Analyse eines komplexeren Datensatzes (große Übungsaufgabe)
- R-Skripten-Sammlung für verschiedene Probleme und Einbau in Anhang der Module
- Aufarbeitung geeigneter Datensätze aus verschiedenen ökologischen Fachgebieten

Um das Angebot zu statistischen Anwendungen in der Ökologie abzurunden, steht die Entwicklung von E-Learning-Modulen zu univariaten statistischen Methoden an, die insbesondere bei den Themen „Datenüberprüfung“ und „Transformation“ mit multivariaten Methoden verzahnt sind.

Dank

Das Projekt erhielt eine Förderung durch Mittel aus Studiengebühren von der Fakultät für Biologie.

Literatur

Gossmann, H.; Fuest, R.; Degenhardt, D.: Von WEBGEO zum WebKit Freiburg – ein Baukasten für interaktive Lernmodule mit Breitenwirkung. In: Schneider, G.; Couné, B.; Gayer, C.; Vögele, E.; Weber, C. (Hrsg.): Neue Medien als strategische Schrittmacher an

der Universität Freiburg - Wie Informations- und Kommunikationstechnologien Studium, Verwaltung und Forschung verändern. 2007, Kap. 7: Entwicklungen. S. 297-305. URL <http://www.freidok.uni-freiburg.de/volltexte/3228/> (last check 2010-01-04)

Mitesser, O.; Rudner, M.; Schmidt, F.: Modellierungs- und Auswertungswerkzeuge. In: Müller, M.; Kaule, G. (Hrsg.): E-Learning mit GIS- und Modellanwendungen. Wichmann, Heidelberg. 2004, S. 107-129.

Petschenka, A.; Kerres, M.: Mediendidaktische Konzeption und Implementierung von Lernmodulen in die Hochschullehre. In: Müller, M.; Kaule, G. (Hrsg.): E-Learning mit GIS- und Modellanwendungen. Wichmann, Heidelberg, 2004, S. 53-66.

R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Wien, 2007. URL: <http://www.r-project.org> (last check 2010-01-04)